Package 'soilKey'

November 12, 2026

Type Package

Title Automated Soil Profile Classification per WRB 2022, 'SiBCS' 5 and USDA Soil Taxonomy 13

Version 0.9.403

Date 2026-11-12

Description Implements deterministic classification keys for the World Reference Base for Soil Resources 2022 (4th edition) and the Brazilian System of Soil Classification ('SiBCS', 5th edition). Provides a unified profile representation with explicit per-attribute provenance, multimodal extraction from field reports and photos via vision-language models, spatial priors from 'SoilGrids' and national soil maps, and gap-filling of soil attributes from Vis-NIR or MIR spectra via the Open Soil Spectral Library ('OSSL'). The taxonomic key itself is never delegated to a language model; LLMs are restricted to schema-validated extraction. Each classification result reports a key trace, a provenance-aware evidence grade, and ambiguities that further measurement would resolve.

License MIT + file LICENSE

URL https://github.com/HugoMachadoRodrigues/soilKey, https://hugomachadorodrigues.github.io/soilKey/

BugReports https://github.com/HugoMachadoRodrigues/soilKey/issues

Encoding UTF-8

LazyData true

LazyDataCompression xz

Depends R (>= 4.1)

Imports R6, data.table, yaml, cli, rlang

Suggests aqp, SoilTaxonomy (>= 0.2.8), mpspline2, terra, foreign, sf, chromote, munsellinterpol, pls, prospectr, resemble, ellmer (>= 0.3.2), httr, jsonlite, jsonvalidate, pdftools, magick, shiny (>= 1.8.1), promises, later, mirai, DT, bslib (>= 0.7.0), shinyWidgets, plotly, leaflet, htmltools, withr, DBI, RSQLite, base64enc, maps, parallel, testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

VignetteBuilder knitr

Config/roxygen2/version 8.1.1

NeedsCompilation no

Author Hugo Rodrigues [aut, cre, cph] (ORCID:
<https://orcid.org/0000-0002-8070-8126>)

Maintainer Hugo Rodrigues <rodrigues.machado.hugo@gmail.com>

Repository CRAN

Date/Publication 2026-11-12 22:30:02 UTC

Contents

abrupt_textural_difference . . . 10
acrisol . . . 11
aeolic_material . . . 12
albeluvic_glossae . . . 12
albic . . . 13
alisol . . . 14
andic_properties . . . 14
andosol . . . 16
annotate_wrb_from_usda . . . 17
anthraquic . . . 18
anthric_horizons . . . 18
apply_soilgrids_depth_prior . . . 19
apply_spectral_preprocessing . . . 20
arenic_texture . . . 22
argic . . . 23
argic_aqp . . . 24
argic_with_strong_clay_films . . . 26
argillic_clay_films_test . . . 26
artefacts . . . 27
as_aqp . . . 28
attach_lucas_spectra . . . 29
audit_argic_strong_films . . . 30
auto_set_proj_env . . . 31
available_esdb_attributes . . . 32
batch_robustness . . . 32
benchmark_afsp . . . 33
benchmark_bdsolos . . . 34
benchmark_isis . . . 36
benchmark_lucas_2018 . . . 36
benchmark_natsoil . . . 39
benchmark_npdb . . . 40
benchmark_nsd . . . 41
benchmark_performance . . . 41

benchmark_redape . . . 43
benchmark_run_classification . . . 44
benchmark_sali . . . 45
benchmark_spectral_fill . . . 45
benchmark_unified . . . 47
benchmark_wrb_vs_usda . . . 49
build_taxon_profiles . . . 50
B_espodico . . . 50
B_incipiente . . . 51
B_latossolico . . . 52
B_nitico . . . 53
B_planico . . . 54
B_textural . . . 55
calcaric_material . . . 57
calcic . . . 57
cambic . . . 58
cambic_aqp . . . 60
canonicalise_kst13ed_gg . . . 61
canonical_reference . . . 62
cerosidade . . . 62
chernic . . . 63
chernozem . . . 64
chernozem_strict . . . 64
claric_material . . . 65
ClassificationResult . . . 66
classification_robustness . . . 69
classify_all . . . 70
classify_asc3 . . . 72
classify_asc3_family . . . 74
classify_by_spectral_neighbours . . . 75
classify_cpcs1967 . . . 77
classify_cscj2011 . . . 79
classify_cssc3 . . . 80
classify_cst2001 . . . 82
classify_csv . . . 83
classify_fao . . . 84
classify_fao1974 . . . 86
classify_fao1988 . . . 87
classify_from_documents . . . 88
classify_from_photos . . . 90
classify_hscs2024 . . . 92
classify_ka5 . . . 94
classify_kidpr2008 . . . 96
classify_nzsc4 . . . 98
classify_profiles . . . 100
classify_rp2008 . . . 102
classify_scsj2017 . . . 104
classify_sibcs . . . 105

classify_sibcs_familia . . . 106
classify_usda . . . 108
classify_usda_family . . . 109
classify_via_smartsolos_api . . . 110
classify_with_engine_heuristic . . . 112
classify_with_uncertainty . . . 112
classify_wrb2022 . . . 114
classify_za1991 . . . 115
clear_kst13_cache . . . 117
clear_ossl_cache . . . 117
cohesic . . . 118
combine_priors . . . 119
compare_engines . . . 119
compare_smartsolos . . . 120
compute_ki . . . 121
compute_kr . . . 121
compute_per_attribute_evidence_grade . . . 122
continuous_rock . . . 123
coverage_report . . . 123
cryic_conditions . . . 124
cryic_horizon . . . 125
DiagnosticResult . . . 126
distrofico . . . 128
dolomitic_material . . . 128
download_bdsolos . . . 129
download_extdata_cache . . . 130
download_ossl_subset . . . 131
download_ossl_subset_with_labels . . . 133
download_redape_dataset . . . 135
duric_horizon . . . 136
duripa . . . 136
ensure_horizon_schema . . . 137
eutrofico . . . 138
evaluate_rsg_tests . . . 138
extract_horizons_from_pdf . . . 139
extract_munsell_from_photo . . . 140
extract_site_from_fieldsheet . . . 141
familia_mineralogia_argila_geral . . . 142
FamilyAttribute . . . 143
fao_crosswalk . . . 144
febr_index_munsell . . . 145
ferralic . . . 146
ferralsol . . . 148
ferric . . . 149
fetch_aafc_normals . . . 149
fetch_natsoil . . . 150
fetch_sali . . . 152
fibrico . . . 153

fill_from_spectra . . . 154
fill_munsell_from_spectra . . . 155
fluvic_material . . . 156
folic . . . 157
format_wrb_name . . . 158
fragic . . . 158
fragipa . . . 159
from_aqp . . . 159
gapfill_by_predicted_taxon . . . 160
gapfill_derive_horizon . . . 161
gapfill_within_pedon . . . 162
get_perturbation_scale . . . 163
gleyic_properties . . . 164
gleysol . . . 165
grapes-or-or-grapes . . . 166
GSM_DEPTHS . . . 167
gypsic . . . 167
gypsiric_material . . . 168
harmonize_to_gsm . . . 168
hemico . . . 170
histic_horizon . . . 170
horizon_attribute_spec . . . 171
horizon_column_spec . . . 172
hortic . . . 172
humus_form_rp . . . 173
hydragric . . . 174
hypersulfidic_material . . . 175
hyposulfidic_material . . . 175
inspect_bdsolos_csv . . . 176
irragric . . . 177
isis_citation . . . 178
isis_profile_to_pedon . . . 178
kastanozem . . . 179
kastanozem_strict . . . 180
key_trace_table . . . 180
kst13_canonical . . . 181
kst13_codes . . . 182
kst13_criteria . . . 183
leptic_features . . . 183
limnic_material . . . 184
limonic . . . 185
lithic_discontinuity . . . 186
lixisol . . . 186
load_afsp_pedons . . . 187
load_afsp_sample . . . 188
load_bdsolos_csv . . . 189
load_embrapa_pedons . . . 190
load_febr_pedons . . . 191

load_kssl_nasis_sample . . . 192
load_kssl_pedons . . . 193
load_kssl_pedons_gpkg . . . 194
load_kssl_pedons_with_nasis . . . 194
load_kssl_sample . . . 195
load_lucas_pedons . . . 196
load_lucas_soil_2018 . . . 197
load_natsoil_pedons . . . 199
load_npdb_pedons . . . 201
load_redape_pedons . . . 202
load_rules . . . 203
load_wosis_sample . . . 204
load_wosis_stratified_sample . . . 205
lookup_esdb . . . 206
lookup_mapbiomas_solos . . . 207
lookup_soilgrids . . . 208
luvisol . . . 210
make_acrisol_canonical . . . 210
make_alisol_canonical . . . 211
make_andosol_canonical . . . 211
make_anthrosol_canonical . . . 212
make_arenosol_canonical . . . 212
make_argissolo_canonical . . . 213
make_asc3_example . . . 213
make_calcisol_canonical . . . 214
make_cambisol_canonical . . . 214
make_cambissolo_canonical . . . 215
make_chernossolo_canonical . . . 215
make_chernozem_canonical . . . 216
make_cpcs1967_example . . . 216
make_cryosol_canonical . . . 217
make_cscj2011_example . . . 217
make_cssc3_example . . . 218
make_cst2001_example . . . 219
make_durisol_canonical . . . 220
make_empty_horizons . . . 220
make_espodossolo_canonical . . . 221
make_fao_example . . . 221
make_ferralsol_canonical . . . 222
make_fluvisol_canonical . . . 223
make_gleissolo_canonical . . . 223
make_gleysol_canonical . . . 224
make_gypsisol_canonical . . . 224
make_histosol_canonical . . . 225
make_hscs2024_example . . . 225
make_ka5_example . . . 226
make_kastanozem_canonical . . . 227
make_kidpr2008_example . . . 228

make_latossolo_canonical . . . 228
make_leptosol_canonical . . . 229
make_lixisol_canonical . . . 229
make_luvisol_canonical . . . 230
make_luvissolo_canonical . . . 230
make_neossolo_canonical . . . 231
make_nitisol_canonical . . . 231
make_nitossolo_canonical . . . 232
make_nzsc4_example . . . 232
make_organossolo_canonical . . . 233
make_phaeozem_canonical . . . 233
make_planosol_canonical . . . 234
make_planossolo_canonical . . . 234
make_plinthosol_canonical . . . 235
make_plintossolo_canonical . . . 235
make_podzol_canonical . . . 236
make_retisol_canonical . . . 236
make_rp2008_example . . . 237
make_scsj2017_example . . . 239
make_solonchak_canonical . . . 239
make_solonetz_canonical . . . 240
make_stagnosol_canonical . . . 240
make_synthetic_pedon_with_spectra . . . 241
make_technosol_canonical . . . 241
make_umbrisol_canonical . . . 242
make_vertisol_canonical . . . 242
make_vertissolo_canonical . . . 243
make_za1991_example . . . 243
mineral_material . . . 244
mollic . . . 245
mudanca_textural_abrupta . . . 246
mulmic_material . . . 247
natric_horizon . . . 247
nitic_horizon . . . 248
nitisol . . . 250
normalise_febr_sibcs . . . 250
normalise_febr_usda . . . 251
normalise_febr_wrb . . . 252
normalise_kssl_subgroup . . . 252
npdb_dry_colour_model . . . 253
npdb_scenario . . . 254
npdb_soil_climate . . . 255
nsd_citation . . . 256
ollama_is_running . . . 256
organic_material . . . 257
organotechnic_material . . . 257
ornithogenic_material . . . 258
ossl_demo_sa . . . 258

ossl_library_template . . . 260
oxic_usda . . . 261
panpaic . . . 261
PedonRecord . . . 262
pedons_from_spectral_table . . . 265
pedon_json_schema . . . 267
pedon_to_spc . . . 268
petrocalcic . . . 268
petroduric . . . 269
petrogypsic . . . 270
petroplinthic . . . 270
phaeozem . . . 271
pick_engine . . . 271
pick_engine_batch . . . 272
pisoplinthic . . . 273
pi_to_confidence . . . 274
plaggic . . . 274
planic_features . . . 275
planosol . . . 276
plinthic . . . 277
posterior_classify . . . 278
predict.soilKey_pls_model . . . 279
predict_from_spectra . . . 279
predict_lab_from_spectra . . . 280
predict_munsell_from_spectra . . . 281
predict_ossl_mbl . . . 282
predict_ossl_plsr_local . . . 283
predict_ossl_pretrained . . . 284
predict_xyz_from_spectra . . . 285
preprocess_spectra . . . 285
pretic . . . 287
print.soilKey_pls_model . . . 287
prior_consistency_check . . . 288
protocalcic_properties . . . 288
protogypsic_properties . . . 289
protovertic . . . 290
read_febr_pedons . . . 291
read_isis . . . 292
read_nsd_pedons . . . 293
read_pedon_csv . . . 295
read_sali_pedons . . . 296
read_spectral_library . . . 298
read_wosis_layers_graphql . . . 300
read_wosis_profiles_graphql . . . 301
reducing_conditions . . . 303
report . . . 303
report_html . . . 305
report_pdf . . . 306

report_to_qgis . . . 307
resolve_wrb_qualifiers . . . 308
retic_properties . . . 309
rp2008_crosswalk . . . 310
rp2008_from_label . . . 311
run_all_benchmarks . . . 312
run_classify_app . . . 313
run_demo . . . 314
run_sibcs_grande_grupo . . . 315
run_sibcs_key . . . 316
run_sibcs_subgrupo . . . 316
run_sibcs_subordem . . . 317
run_taxa_list . . . 317
run_taxonomic_key . . . 318
run_usda_great_group . . . 319
run_usda_key . . . 319
run_usda_subgroup . . . 320
run_usda_suborder . . . 320
run_wrb_key . . . 321
salic . . . 321
sali_citation . . . 322
saprico . . . 323
save_ossl_models . . . 324
shrink_swell_cracks . . . 324
sideralic_properties . . . 325
site_attribute_spec . . . 325
soilgrids_usda_lut . . . 326
soilgrids_wrb_lut . . . 326
soilkey_systems . . . 327
soil_classes_at_location . . . 327
soil_organic_carbon . . . 329
solimovic_material . . . 330
sombric . . . 330
spatial_prior . . . 331
spatial_prior_embrapa . . . 332
spatial_prior_soilgrids . . . 333
spodic . . . 334
stagnic_properties . . . 335
st_features_canonical . . . 336
subgrupo_planossolo_espessos . . . 337
subgrupo_planossolo_mesicos . . . 338
subgrupo_plintossolo_endico_concrecionario . . . 338
subgrupo_plintossolo_endico_litoplintico . . . 339
subgrupo_plintossolo_espessos . . . 340
takyric_properties . . . 340
technic_features . . . 341
technic_hard_material . . . 342
tephric_material . . . 342

terric . . . 343
texture_class_from_pct . . . 343
thionic . . . 344
train_pls_from_ossl . . . 345
tsitelic . . . 346
umbric_horizon . . . 347
unrecognised_attributes . . . 348
usda_to_wrb_rsg . . . 349
validate_horizon_geometry . . . 350
validate_pedon_json . . . 351
vertic_horizon . . . 351
vertic_properties . . . 353
vertisol . . . 354
vitric_properties . . . 355
vlm_pick_provider . . . 356
vlm_provider . . . 356
within_tolerance . . . 357
wosis_citation . . . 358
wosis_profile_to_pedon . . . 359
wrb06_code_to_rsg . . . 360
wrb2022_canonical . . . 360
yermic_properties . . . 361

Index **362**

`abrupt_textural_difference`

Abrupt textural difference (WRB 2022 Ch 3.2.1)

Description

Sharp clay-content increase between two superimposed mineral layers meeting all of (WRB 2022 3.2.1, p. 72):

- underlying clay \>= 15% AND thickness \>= 7.5 cm;
- underlying starts \>= 10 cm below the mineral soil surface;
- underlying has, vs overlying: 2x clay if overlying < 20%, OR \>= 20pp (absolute) more clay if overlying \>= 20%;
- transitional layer, if any, \<= 2 cm.

Since v0.9.312 criterion 5 is read from the distinctness of the lower boundary of the overlying layer (`boundary_distinctness`): abrupt (2 cm or less) or very abrupt passes, clear, gradual or diffuse fails, and a boundary not recorded leaves the pair undecided (`NA`). Criterion 2 is counted from the mineral soil surface (below organic layers).

Usage

```
abrupt_textural_difference(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

acrisol *Acrisol RSG diagnostic (WRB 2022)*

Description

The Acrisols of the WRB 2022 key (Ch 4, p. 119): an argic horizon starting $\leq$ 100 cm from the mineral soil surface; a CEC (1 M NH4OAc, pH 7) of < 24 cmolc/kg clay in some subhorizon of the argic horizon within 150 cm of the mineral soil surface; and exchangeable Al > exchangeable (Ca+Mg+K+Na) in half or more of 50-100 cm from the mineral soil surface, or of the lower half of the mineral soil above a limiting layer starting $\leq$ 100 cm, whichever is shallower. Where the cations are not available, the pH is read by WRB 2022 Annex 2, Table 9.3.

Usage

```
acrisol(pedon, max_cec = 24, max_bs = 50)
```

Arguments

pedon	A PedonRecord.
max_cec	CEC per kg clay below which the clay is of low activity (default 24).
max_bs	Unused since v0.9.310 (WRB 2022 does not use the base saturation here), kept so that existing calls still run.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 4, Acrisols (p. 119).

aeolic_material *Aeolic material (WRB 2022 Ch 3.3.1)*

Description

Wind-deposited material in the upper 20 cm: rounded matt-surfaced sand grains OR aeroturbation features, AND < 1% SOC in the upper 10 cm. v0.3.3 detects via `rock_origin == "aeolian"` OR `layer_origin == "aeolic"`.

Usage

```
aeolic_material(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

albeluvic_glossae *Albeluvic glossae (WRB 2022 Ch 3.2.2)*

Description

Tongues of bleached, coarser-textured material penetrating an argic horizon. v0.3.3 detects via designation pattern `glossic|albeluvic` on a layer that overlies an argic-horizon-passing layer.

Usage

```
albeluvic_glossae(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`albic` *Albic horizon (WRB 2022)*

Description

A bleached eluvial horizon – claric material that has lost iron oxides and/or organic matter due to clay migration, podzolization, or redox under stagnant water. Diagnostic for parts of Podzols, Retisols and Planosols qualifiers.

Usage

```
albic(pedon, min_thickness = 1)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness in cm (default 1, per WRB 2022). The albic horizon has no canonical thickness gate; we keep a token min so that fully-NA layers don't pass.

Details

Sub-tests:

- `test_claric_munsell` – Munsell criteria of claric material (Ch 3.3.4).

Designation pattern `E` or `Eg` also serves as positive evidence when Munsell columns are missing (proxy path).

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Ch 3.1 – Albic horizon.

alisol *Alisol RSG diagnostic (WRB 2022)*

Description

The Alisols of the WRB 2022 key (Ch 4, p. 121): an argic horizon starting $\leq$ 100 cm from the mineral soil surface and exchangeable Al > exchangeable (Ca+Mg+K+Na) in half or more of 50-100 cm (or of the lower half of the mineral soil above a limiting layer starting $\leq$ 100 cm). The key reaches them after the Acrisols and Lixisols.

Usage

```
alisol(pedon, min_cec = 24, min_al_sat = 50)
```

Arguments

pedon	A PedonRecord.
min_cec, min_al_sat	Unused since v0.9.310, kept so that existing calls still run.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 4, Alisols (p. 121).

andic_properties *Andic properties (WRB 2022)*

Description

Tests for the andic property complex – volcanic-ash-derived allophanic / imogolitic / Al-humus material. Diagnostic of Andosols. Two alternative qualifying paths per WRB 2022 Ch 3.2:

1. **Al-Fe oxalate + low BD**: (Al_ox + 0.5*Fe_ox) >= min_alfe (default 2.0%) AND bulk_density <= max_bd (default 0.9 g/cm^3) on the same layer.
2. **Phosphate retention**: phosphate_retention_pct >= min_p_retention (default 70%).

Either path qualifies. The volcanic-glass criterion is the separate vitric_properties diagnostic; Andosols key on (andic OR vitric) at the RSG-gate level (andosol).

Usage

```
andic_properties(
  pedon,
  min_alfe = 2,
  max_bd = 0.9,
  min_p_retention = 70,
  min_oc_proxy = 4,
  max_bd_proxy = 0.9
)
```

Arguments

pedon	A PedonRecord.
min_alfe	Minimum (Al_ox + 0.5*Fe_ox) percent for the Al-Fe path (default 2.0).
max_bd	Maximum bulk density g/cm^3 for the Al-Fe path (default 0.9).
min_p_retention	Minimum phosphate retention % for the P path (default 70).
min_oc_proxy	Minimum SOC % for the v0.9.80 OC+BD proxy path (default 4.0). Only consulted when the proxy is enabled via options(soilKey.andic_oc_bd_proxy = TRUE).
max_bd_proxy	Maximum bulk density g/cm^3 for the v0.9.80 OC+BD proxy path (default 0.9). Only consulted when the proxy is enabled.

Value

A DiagnosticResult.

v0.9.313

The answer is read layer by layer and is NA when a value it needs is not recorded (until v0.9.312 an undecided profile returned FALSE); evidence$layer_status holds the answer per layer. A layer fails when Al_ox + 0.5*Fe_ox < min_alfe or the bulk density > max_bd is measured (phosphate retention, an optional path, can still pass it). Where bulk density is not measured, Al_ox + 0.5*Fe_ox >= min_alfe together with a pH in NaF (ph_naf) >= 9.5 in a carbonate-free layer passes, read from the field test of WRB 2022 Ch 3.2.3; this is a proxy, named in notes, reference and evidence$naf_field_test.

v0.9.80 OC + BD proxy (opt-in)

Field-described volcanic-ash soils (e.g.\ AfSP, KSSL/NASIS, SOTER) routinely lack oxalate Al/Fe and phosphate retention measurements, so the canonical paths return NA and Andosols cascade to other RSGs. The genetic signature is still detectable from coarser data: very high SOC (>= 4-5%) plus low bulk density (<= 0.9 g/cm^3) typical of allophanic / Al-humus complexation.

With options(soilKey.andic_oc_bd_proxy = TRUE) the function adds a third path that fires when both canonical paths fail and the surface horizon shows oc_pct >= min_oc_proxy AND bulk_density_g_cm3 <= max_bd_proxy (or OC alone >= 5% when BD is missing). Default is FALSE (canonical behaviour preserved).

v0.9.85 proxy contiguous-layer extension (opt-in)

When `options(soilKey.andic_oc_bd_proxy_extend = TRUE)` (only meaningful with `soilKey.andic_oc_bd_proxy = TRUE`), iteratively extend the proxy layers to include contiguous deeper layers whose `oc_pct >= min_oc_proxy / 2` AND whose `bulk_density_g_cm3` is missing OR `<= max_bd_proxy + 0.15`. The extension stops at the first horizon failing either constraint, so a ferralic / argic subsoil cannot accidentally inflate the andic thickness. Default is `FALSE` – canonical proxy behaviour preserved.

References

IUSS Working Group WRB (2022), Chapter 3, Andic properties.

---

andosol *Andosol RSG gate (WRB 2022 Ch 4, p 104)*

Description

WRB-canonical: layer(s) with *andic* OR *vitric* properties, combined thickness \>= 30 cm within 100 cm starting \<= 25 cm; OR \>= 60% of the entire soil thickness when a limiting layer starts 25-50 cm. Plus: no argic, ferralic, petroplinthic, pisoplinthic, plinthic or spodic horizon \<= 100 cm (unless buried below 50 cm).

Usage

```
andosol(
  pedon,
  min_thickness = 30,
  max_top_cm = 25,
  buried_below_cm = 50,
  strict = NULL
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Numeric threshold or option (see Details).
`max_top_cm`	Numeric threshold or option (see Details).
`buried_below_cm`	Numeric: layers of the exclusion diagnostics whose top_cm \>= this depth are treated as buried and do NOT exclude the Andosol (default 50, per WRB 2022 Ch 4 p 104).
`strict`	Logical or `NULL`. When `NULL` (default) it resolves via `getOption("soilKey.rsg_strict", FALSE)`. `TRUE` disables the buried-exclusion tolerance.

Details

v0.3.4 enforces (1) andic OR vitric AND (2) combined thickness \>= 30 cm starting in the upper 25 cm AND (3) the negative-list exclusions on argic / ferralic / plinthic / spodic.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

v0.9.313 (WRB 2022 Ch 4, p 104, as worded)

The combined thickness is that of every andic or vitric layer within 100 cm of the soil surface (clipped at 100 cm), provided the uppermost starts \<= 25 cm; until v0.9.312 only layers starting \<= 25 cm were added up. Item 1b applies: where a limiting layer starts \> 25 and \<= 50 cm, andic or vitric layers covering \>= 60% of the soil above it qualify. Petroplinthic and pisoplinthic horizons join the exclusions. An argic horizon read from a clay increase alone (no clay films recorded) whose layers all have andic properties does not exclude (Hugo, N4: the clay of andic material does not disperse fully). Layers whose andic or vitric status is not recorded leave the gate undecided (NA) where they could change it.

v0.9.85 buried-exclusion fix

WRB 2022 Ch 4 p 104 specifies the Andosol exclusion list (argic / ferralic / petroplinthic / pisoplinthic / plinthic / spodic) as "<= 100 cm *unless buried below 50 cm*". The earlier implementation excluded an Andosol whenever any of those diagnostics passed anywhere in the profile, including on layers starting deeper than 50 cm – which mis-fires on AfSP Andosol references like CM W3_0047, where an argic layer at 56-72 cm wrongly excluded the andic surface stack. v0.9.85 restricts the exclusion check to layers starting <= 50 cm: a buried argic / ferralic / plinthic / spodic at deeper levels no longer disqualifies the surface andic stack from Andosol.

Tier-3 strict mode (v0.9.98)

With strict = TRUE the v0.9.85 buried-exclusion tolerance is switched off: *any* argic / ferralic / plinthic / spodic horizon anywhere in the profile excludes the Andosol, regardless of depth.

---

annotate_wrb_from_usda

Annotate KSSL/NASIS pedons with a derived WRB Reference Soil Group

---

Description

Applies usda_to_wrb_rsg to each pedon's USDA classification (preserved as site$reference_usda + site$reference_usda_suborder by load_kssl_pedons_gpkg) and writes the result to site$reference_wrb_from_usc – a "best-guess" expected WRB label for benchmark comparison.

Usage

```
annotate_wrb_from_usda(pedons)
```

Arguments

pedons	List of PedonRecord objects.

Details

Pedons that already have `site$reference_wrb` populated (e.g.\ from external sources) are left untouched.

Value

The same list, with `site$reference_wrb_from_usda` populated where USDA classification is present.

`anthraquic` *Anthraquic horizon (WRB 2022): puddled-rice / paddy plough layer. v0.3.3 detects via designation pattern* `Apl|Ap|Hh`.

Description

Anthraquic horizon (WRB 2022): puddled-rice / paddy plough layer. v0.3.3 detects via designation pattern `Apl|Ap|Hh`.

Usage

```
anthraquic(pedon, min_thickness = 20, max_top_cm = 50)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Numeric threshold or option (see Details).
`max_top_cm`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`anthric_horizons` *Anthric horizons (WRB 2022)*

Description

Tests for any of five anthropogenic surface horizons recognised by WRB 2022 (hortic, irragric, plaggic, pretic, terric). Diagnostic of Anthrosols. Two alternative paths qualify:

1. **Designation**: any layer's designation contains one of `hortic|irragric|plaggic|pretic|terric`.
2. **Property-based**: a surface layer (top_cm <= 5) at least `min_thickness_cm` cm thick (default 20) with elevated dark colour (Munsell value moist <= `max_munsell_value`, default 4) AND elevated plant-available P (`p_mehlich3_mg_kg` >= `min_p_mg_kg`, default 50).

Either path qualifies.

Usage

```
anthric_horizons(
  pedon,
  min_thickness_cm = 20,
  min_p_mg_kg = 50,
  max_munsell_value = 4
)
```

Arguments

pedon	A PedonRecord.
min_thickness_cm	Minimum thickness for the property-based path (default 20).
min_p_mg_kg	Minimum plant-available P (Mehlich 3, mg/kg) for the property-based path (default 50).
max_munsell_value	Maximum Munsell value moist for the property-based path (default 4).

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 5, Anthrosols.

---

apply_soilgrids_depth_prior

Fill missing horizon attributes from a SoilGrids depth prior

Description

For each horizon and each requested attribute, interpolates the value at the horizon's mid-depth from the six standard SoilGrids 2.0 depth slices (0-5, 5-15, 15-30, 30-60, 60-100, 100-200 cm) and writes it into the pedon with `source = "inferred_prior"`. Existing values are preserved unless `overwrite = TRUE`; the PedonRecord authority order means a SoilGrids prior can never silently displace a measured, spectra-predicted or VLM-extracted value.

Usage

```
apply_soilgrids_depth_prior(
  pedon,
  attrs = NULL,
  depth_profiles = NULL,
  overwrite = FALSE
)
```

Arguments

pedon	A PedonRecord with at least one horizon. For the live fetch it must also carry site$lat and site$lon.
attrs	Character vector of horizon columns to fill. Defaults to all SoilGrids-backed attributes: clay_pct, sand_pct, silt_pct, ph_h2o, oc_pct, cec_cmol.
depth_profiles	Optional named list mapping an attribute to a numeric vector of six slice values (0-5 ... 100-200 cm). When supplied the SoilGrids network call is skipped entirely – this is the path the test suite and offline users take.
overwrite	If FALSE (default) only NA cells are filled. If TRUE, every requested cell is overwritten (subject to the provenance authority order).

Details

This is the depth-resolved companion to spatial_prior_soilgrids (which returns a site-level RSG probability vector, not horizon attributes), and the attribute-fill stage of classify_from_photos.

Value

Invisibly, the mutated pedon. An attribute "soilgrids_depth_fill" on the return value records how many cells were filled.

Examples

```
## Not run:
p <- make_cambisol_canonical()
p$horizons$clay_pct <- NA_real_
# Offline: supply the six-slice profiles directly.
apply_soilgrids_depth_prior(
  p, attrs = "clay_pct",
  depth_profiles = list(clay_pct = c(18, 20, 24, 28, 30, 30)))

## End(Not run)
```

---

apply_spectral_preprocessing

Apply a step-by-step Vis-NIR / MIR preprocessing pipeline

Description

Composes the canonical soil-spectroscopy sequence, each step optional and applied in this fixed scientific order: reflectance → (absorbance) → (Savitzky-Golay smoothing) → (Savitzky-Golay 1st or 2nd derivative). Each Savitzky-Golay pass trims (window - 1) / 2 columns from each edge; the wavelength axis is trimmed to match and returned so callers can plot the treated spectrum on the correct axis.

Usage

```
apply_spectral_preprocessing(
  X,
  wavelengths = NULL,
  absorbance = FALSE,
  sg_smooth = FALSE,
  sg_derivative = 0L,
  window = 11L,
  poly = 2L
)
```

Arguments

X	Numeric matrix (rows = samples/horizons, columns = wavelengths) or a numeric vector (treated as one sample).
wavelengths	Optional numeric wavelength axis. Defaults to the numeric part of colnames(X), else 1:ncol(X).
absorbance	Logical; apply $A = \log_{10}(1/R)$.
sg_smooth	Logical; apply Savitzky-Golay smoothing (m = 0).
sg_derivative	Integer 0, 1 or 2; Savitzky-Golay derivative order (0 = none).
window	Odd Savitzky-Golay window (default 11); coerced to a valid odd value in [3, ncol).
poly	Savitzky-Golay polynomial order (default 2); clamped to [1, window - 1].

Details

The transform is robust: reflectance that looks like a percentage (maximum > 1.5) is rescaled to a 0–1 fraction before the absorbance log, values are clamped away from zero to avoid log(0), and a Savitzky-Golay step that cannot fit the requested window into the available wavelengths is skipped (recorded in steps) rather than erroring.

Value

A list with X (the treated numeric matrix, wavelength column names trimmed to match), wavelengths (numeric) and steps (an ordered character vector describing the transforms actually applied, starting with "Reflectance").

See Also

preprocess_spectra

Examples

```
X <- matrix(seq(0.1, 0.5, length.out = 3 * 60), nrow = 3, byrow = TRUE)
colnames(X) <- seq(400, 2400, length.out = 60)
res <- apply_spectral_preprocessing(X, absorbance = TRUE,
                                    sg_smooth = TRUE, sg_derivative = 1L)
res$steps          # ordered treatment labels
```

```
dim(res$X)         # columns trimmed by the two SG passes
```

`arenic_texture` *Arenic texture (WRB 2022)*

Description

Tests the texture of the Arenosols (WRB 2022 Ch 4, p. 124): within 100 cm of the mineral soil surface, a weighted average texture of loamy sand or sand (`silt + 2 * clay < 30`), layers of finer texture and layers with $\geq$ 40% coarse fragments each with a combined thickness of < 15 cm (v0.9.314; until then every layer had to be coarser than sandy loam). Diagnostic of Arenosols.

Usage

```
arenic_texture(pedon, max_top_cm = 100, engine = NULL)
```

Arguments

pedon	A PedonRecord.
max_top_cm	Maximum top depth (cm) of layers to be tested (default 100, per WRB 2022).
engine	One of "soilkey" (default; the WRB 2022 test, test_arenosol_texture) or "aqp" (LUCAS-friendly fallback: passes when sand >= 70\ reads getOption("soilKey.diagnostic_en

Details

Sub-test: `test_arenosol_texture`. A part of the 100 cm whose texture or coarse fragments are not recorded, or that is not described, leaves the result undecided (NA) when it could change it.

v0.3 limitations: WRB 2022 Arenosol also requires that no other diagnostic horizon (argic, ferralic, etc.) is present, but those exclusions happen at the key level via canonical RSG order.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 5, Arenosols.

argic *Argic horizon (WRB 2022)*

Description

Tests whether any horizon meets the argic horizon criteria per Chapter 3 of the WRB 2022 (4th edition). Argic is a subsurface horizon with distinctly higher clay content than the overlying horizon, qualified by three depth-conditional clay-increase rules; it must also have texture of sandy loam or finer, satisfy a minimum thickness, and not exhibit albeluvic glossic features (which would direct the profile to the Retisol path).

Usage

```
argic(
  pedon,
  min_thickness = 7.5,
  system = c("wrb2022", "usda"),
  engine = NULL,
  require_t = NULL,
  lithic_discontinuity = NULL
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness in cm (default 7.5).
system	One of "wrb2022" (default) or "usda". Selects the clay-increase threshold set: WRB uses 6/1.4/20 pp/ratio/pp; KST 13ed uses 3/1.2/8 (looser). See test_clay_increase_argic for the table.
engine	v0.9.63+. One of "soilkey" (the hand-coded path, default for back-compat) or "aqp" (canonical NRCS dispatch via aqp::getArgillicBounds). When NULL (the new default) the function reads getOption("soilKey.diagnostic_engine", "soilkey") so a global options(soilKey.diagnostic_engine = "aqp") flips every argic() call without modifying call sites. See argic_aqp.
require_t	v0.9.63+. Forwarded to aqp::getArgillicBounds when engine = "aqp": TRUE requires a "t" suffix in the horizon designation (the strict KST 13ed text); FALSE accepts argic by clay-increase alone (more permissive on data-sparse profiles). NULL (default) auto-picks: TRUE for system = "usda", FALSE for system = "wrb2022". Ignored when engine = "soilkey".
lithic_discontinuity	v0.9.220. Apply WRB 2022 criterion 2.a.i: a clay increase over a coarser layer of another material (Ap over 2Bt, read from the designations' material numbers) counts only with clay films of "common" or more (illuvial clay, criterion 2.b). NULL (default): TRUE for system = "wrb2022". The SiBCS B textural, whose item (j) accepts such an increase by its textural ratio, and the USDA keys pass FALSE.

Details

Sub-tests called (each a list with `passed`, `layers`, `missing`, `details`, `notes`):

- `test_clay_increase_argic` – the three-pronged WRB 2022 clay-increase rule.
- `test_minimum_thickness` – thickness >= 7.5 cm (configurable via `min_thickness`).
- `test_texture_argic` – texture of sandy loam or finer (`silt + 2 * clay >= 30`).

Until v0.9.219 a glossic, albeluvic or retic designation anywhere voided the argic horizon (the WRB 2014 Albeluvisol logic); in WRB 2022 retic properties belong to an argic or natric horizon.

v0.9.310, for `system = "wrb2022"` (WRB 2022 Ch 3.1.3, p. 35-36):

- criterion 1: the layers have $\geq$ 8% clay;
- criterion 2.a.i (v0.9.317): the coarser-textured layer is not separated from the layer by a lithic discontinuity, read from the material numbers of the designations and from the data of each pair of layers in between (Ch 3.2.7, p. 76-77: the ratio sand to silt, criterion 2.a.iv; coarse fragments, criterion 6);
- criterion 2.a.ii (v0.9.317): no clay increase counts below a plough layer (Ap) that directly overlies the argic horizon;
- criterion 2.a.iii: a layer is compared with a coarser-textured layer above it only across a transitional horizon of $\leq$ 15 cm; layers already accepted and layers whose designation has the suffix t are part of the argic horizon, not transitional;
- criterion 2.b.ii: clay films of "common" or more (clay coatings on $\geq$ 15% of the aggregate surfaces) make a layer argic without a clay increase;
- a subsurface horizon: the layer at the mineral soil surface and the A horizons are not part of it.

The USDA argillic and order keys and the SiBCS B textural keep the earlier reading.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition. International Union of Soil Sciences, Vienna. Chapter 3 – Argic horizon.

`argic_aqp` *Argic / argillic horizon via aqp::getArgillicBounds()*

Description

Wraps `aqp::getArgillicBounds()` (Beaudette et al.) in soilKey's `DiagnosticResult` contract. The aqp implementation is the canonical NRCS R port and uses the tiered USDA-NRCS clay-increase thresholds:

- Eluvial clay < 15\
- Eluvial clay 15-40\
- Eluvial clay \>= 40\

(vs. soilKey's hand-coded `argic` which uses the WRB 6/1.4/20 thresholds). For BDsolos / FEBR / KSSL profiles the aqp rule is closer to KST 13ed and BDsolos field practice.

Usage

```
argic_aqp(pedon, require_t = FALSE, ...)
```

Arguments

`pedon`	A `PedonRecord`.
`require_t`	Whether to require an explicit "t" suffix in the horizon designation (default `FALSE` for BDsolos / FEBR; `TRUE` matches the strict KST 13ed text).
`...`	Reserved for future arguments.

Details

By default aqp requires a "t" suffix in the horizon designation (`require_t = TRUE`); we expose this so callers can be permissive on datasets where designation is missing or non-conforming (BDsolos exports often drop the "t").

Value

A `DiagnosticResult` with `name = "argic_aqp"`. `$layers` are the row indices of horizons in the argillic / argic depth interval. `$evidence` carries the raw aqp `c(ubound, lbound)` bounds for traceability.

See Also

`argic` (soilKey hand-coded; WRB 6/1.4/20), `aqp::getArgillicBounds`.

argic_with_strong_clay_films

Test whether a pedon's argic horizon has strong clay films

Description

Wraps `argic()` and inspects the `clay_films_amount` field at the argic-passing layers. Returns a structured result that `B_latossolico()` used, from v0.9.61 to v0.9.240, to decide whether a B textural excluded it.

Usage

```
argic_with_strong_clay_films(pedon)
```

Arguments

pedon	A `PedonRecord`.

Details

Since v0.9.241 the SiBCS key does not use this rule: the B textural is tested by the items of the book and excludes the B latossolico, whose cerosidade is tested layer by layer (see `B_textural` and `B_latossolico`). The function is kept to inspect a dataset.

Value

A list with:

- `passed` – logical, `TRUE` only when argic passes AND at least one argic-passing layer has a strong (*comum* / *abundante*) film qualifier.
- `layers` – integer vector of argic-passing layer indices (empty when `passed` is `FALSE`).
- `argic` – the underlying `DiagnosticResult` from `argic()`.
- `films` – character vector of the `clay_films_amount` values at the argic-passing layers.

argillic_clay_films_test

Test for clay-illuviation evidence (KST 13ed Ch 3 p 4)

Description

KST 13ed argillic horizon requires "evidence of illuvial accumulation of clay" alongside the clay-increase rule. Acceptable evidence:

- oriented clays bridging sand grains in >= 1% of the horizon;
- clay films lining pores or coating ped faces;
- lamellae more than 5 mm thick.

Usage

```
argillic_clay_films_test(pedon)
```

Arguments

pedon	A PedonRecord.

Details

This test reads three complementary slots, in order of evidence strength:

1. `pedon$site$nasis_diagnostic_features` – the NASIS `pediagfeatures.featkind` vector. The surveyor's explicit "Argillic horizon" entry directly confirms clay-illuviation evidence (~13 500 entries in the 2021 NASIS snapshot). Strongest evidence.
2. `pedon$horizons$clay_films_amount` – per-horizon clay-film abundance derived from NASIS `phpvsf`. Values: `"few"`, `"common"`, `"many"`, `"continuous"`. Direct measurement.
3. `pedon$horizons$designation` containing a 't' master suffix (e.g. `Bt`, `Btk`, `Btx`, `Bt1`, `2Bt`). v0.9.28: the pedologist who wrote that designation explicitly identified the horizon as clay-illuvial – per KST 13ed Ch 18, the 't' suffix means "accumulation of silicate clay" – so it counts as positive evidence even when NASIS records are absent. This unlocks the KST 13ed argillic thresholds for the ~47 pediagfeatures and phpvsf records.

Any of the three sources counts as positive evidence (logical OR). `passed = NA` when none is populated AND no horizon designation field is present at all (lab-only loaders without horizon descriptions). `passed = FALSE` when designations exist but none has a 't' suffix and NASIS slots are empty.

Value

A DiagnosticResult.

References

Soil Survey Staff (2022), Keys to Soil Taxonomy 13th ed., Ch. 3, argillic horizon (clay-illuviation criteria, p. 4); Ch. 18, master horizon symbols (`t`: silicate-clay accumulation, p. 332).

`artefacts` *Artefacts (WRB 2022 Ch 3.3.2)*

Description

Per the canonical definition: human-made / human-altered / human- excavated material. v0.3.3 returns the layers where `artefacts_pct >= 1`.

Usage

```
artefacts(pedon, min_pct = 1)
```

Arguments

pedon	A PedonRecord.
min_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

as_aqp *Convert one or more PedonRecord objects to an aqp SoilProfileCollection*

Description

Builds a aqp::SoilProfileCollection from one PedonRecord or a list of them. Standard soilKey columns (top_cm, bottom_cm, designation, clay_pct, sand_pct, silt_pct) are renamed to aqp's canonical convention (top, bottom, name, clay, sand, silt). All other columns are passed through unchanged. Site-level slots (lat, lon, country, parent_material, reference_*, nasis_diagnostic_features, etc.) are attached to the SPC's site table.

Usage

```
as_aqp(x)
```

Arguments

x	A PedonRecord or a list of them.

Details

Requires the aqp package, listed in Suggests; the function raises a clear error if aqp is not installed.

Value

A aqp::SoilProfileCollection.

See Also

from_aqp, the inverse conversion.

Examples

```
## Not run:
library(soilKey)
library(aqp)

pedons <- list(make_ferralsol_canonical(), make_luvisol_canonical())
spc <- as_aqp(pedons)
length(spc)         # 2 profiles
aqp::horizons(spc)  # one row per horizon, aqp-named columns

## End(Not run)
```

`attach_lucas_spectra` *Attach LUCAS 2018 Vis-NIR spectra to a list of PedonRecord objects*

Description

Joins the LUCAS Soil 2018 Spectral Library (separate ESDAC release, ~83 GB) onto the pedons returned by `load_lucas_soil_2018`, by matching the LUCAS `POINT_ID` of the spectra against `pedon$site$id`. Each matched pedon gets `$spectra$vnir` populated as a numeric matrix (rows = horizons, cols = wavelengths).

Usage

```
attach_lucas_spectra(
  pedons,
  spectra,
  point_id_col = "POINT_ID",
  verbose = TRUE
)
```

Arguments

`pedons`	List of `PedonRecord` objects.
`spectra`	A wide or long `data.frame` as described above.
`point_id_col`	Name of the LUCAS point-id column in `spectra`. Default `"POINT_ID"`.
`verbose`	If `TRUE` (default), reports the join hit rate.

Details

Two input shapes are accepted:

- A wide `data.frame` keyed by an integer `POINT_ID` column with one column per wavelength (column names parseable as numeric nm). One row per LUCAS point.
- A long `data.frame` with columns `POINT_ID`, `wavelength_nm`, `reflectance`.

Spectra are attached only to the topsoil horizon (row 1); the subsoil horizon (if any) is left without spectra. After this call, `benchmark_lucas_2018(..., fill_topsoil_from = "spectra", ossl_models = ...)` feeds the spectra through `predict_from_spectra` (v0.9.46) to fill any chemistry / texture gap not already populated by SoilGrids.

Value

The list of pedons (mutated in place; returned invisibly).

See Also

`predict_from_spectra`, `predict_munsell_from_spectra`, `load_lucas_soil_2018`.

`audit_argic_strong_films` *Audit the strong-clay-films exclusion across a list of pedons*

Description

Applies `argic_with_strong_clay_films()` to every pedon in `pedons` and returns a per-pedon table summarising how the rule that `B_latossolico()` applied from v0.9.61 to v0.9.240 resolves on the benchmark sample: a B textural (the WRB argic horizon, then) excluded the B latossolico only when its clay films were common or more.

Usage

```
audit_argic_strong_films(pedons, reference_filter = NULL)
```

Arguments

`pedons`	List of `PedonRecord` objects.
`reference_filter`	Optional regex applied to `p$site$reference_sibcs` to keep only pedons whose reference matches (case-sensitive, ICU). Default `NULL` keeps every pedon.

Details

Since v0.9.241 the SiBCS key does not use that rule (see `B_latossolico`); the audit shows what it would have done on field-described datasets such as BDsolos and Redape, where clay-film qualifiers are recorded in mixed Portuguese / English tokenisation. The audit is read-only and never invokes `classify_sibcs()`.

Value

A `data.frame` with columns `id`, `reference_sibcs`, `argic_passed`, `has_films_at_argic`, `strong_films_at_argic`, and `would_exclude_from_latossolo`.

Examples

```
## Not run:
peds <- load_bdsolos_csv("RJ.csv")
a <- audit_argic_strong_films(peds, reference_filter = "LATOSSOLO")
table(a$would_exclude_from_latossolo)

## End(Not run)
```

`auto_set_proj_env` *Auto-detect PROJ_LIB and GDAL_DATA directories*

Description

Probes the common system locations for PROJ `proj.db` and GDAL data directories, on macOS Homebrew (Apple silicon and Intel), Linuxbrew, conda / mamba environments, and Debian / Ubuntu / Fedora apt or dnf installs. Sets the corresponding environment variables only when they are not already set, so a user-provided value always wins. Idempotent: safe to call repeatedly.

Usage

```
auto_set_proj_env(verbose = FALSE)
```

Arguments

`verbose`	If `TRUE`, emits a `cli` message confirming what was detected.

Details

Called automatically from `.onLoad`; call manually after installing PROJ / GDAL via Homebrew if you want to refresh the env without restarting R.

Value

Invisibly, a named list with `PROJ_LIB` and `GDAL_DATA` (the values that were set, or `NA_character_` if a value was already present or no candidate was found).

available_esdb_attributes

List ESDB Raster Library attributes available at a given root

Description

Walks ‘raster_root‘ and returns the folder names that contain a valid ‘<NAME>.tif‘ raster. Useful for discovery before calling `lookup_esdb`.

Usage

```
available_esdb_attributes(raster_root)
```

Arguments

`raster_root`	Path to the unpacked ESDB raster directory (typically ‘<some>/ESDB-Raster-Library-1k-GeoTIFF-...‘).

Value

A character vector of attribute names (sorted).

Examples

```
## Not run:
available_esdb_attributes("~/data/ESDB-Raster-Library-1k-GeoTIFF-20240507")
#> [1] "AGLI1NNI" "AGLI2NNI" "AGLIM1" "AGLIM2" "ALT" "ATC" "AWC_SUB" ...
#>     [continued: 71 attributes]

## End(Not run)
```

batch_robustness

Batch robustness across many pedons

Description

Runs `classification_robustness` on each pedon in a list and returns a tidy data.frame with one row per pedon. Useful for paper-grade claims like "85 to a 5

Usage

```
batch_robustness(pedons, ...)
```

Arguments

`pedons`	List of `PedonRecord` objects.
`...`	Passed to `classification_robustness`.

Value

A data.frame with columns `id`, `baseline`, `robustness`, `n_flipped`.

Examples

```
## Not run:
pedons <- list(make_ferralsol_canonical(),
                make_luvisol_canonical(),
                make_chernozem_canonical())
batch_robustness(pedons, system = "wrb2022", n = 50)
#>            id   baseline robustness n_flipped
#> 1 FR-canon-01 Ferralsols       0.96         2
#> 2 LV-canon-01   Luvisols       1.00         0
#> 3 CH-canon-01 Chernozems       0.94         3

## End(Not run)
```

---

benchmark_afsp *Benchmark soilKey WRB predictions against AfSP ground truth*

Description

Benchmark soilKey WRB predictions against AfSP ground truth

Usage

```
benchmark_afsp(pedons, verbose = TRUE)
```

Arguments

`pedons`	A list of `PedonRecord` objects, as `load_afsp_pedons` returns, or the list that `load_afsp_sample` returns (the profiles are then taken from its `$pedons`; until v0.9.222 that list stopped with "$ operator is invalid for atomic vectors").
`verbose`	Print progress.

Value

List with `accuracy`, `n_compared`, `confusion`, `per_class_recall`.

benchmark_bdsolos *Benchmark soilKey classifiers against BDsolos national reference labels*

Description

Runs classify_wrb2022, classify_sibcs, and classify_usda on each PedonRecord loaded from a BDsolos CSV via load_bdsolos_csv, then compares each predicted classification against the corresponding BDsolos reference label (reference_sibcs, reference_wrb, reference_st) and reports per-system accuracy, per-class recall, and a confusion matrix.

Usage

```
benchmark_bdsolos(
  pedons,
  systems = c("wrb2022", "sibcs", "usda"),
  sibcs_level = c("order", "subordem"),
  max_n = NULL,
  verbose = TRUE,
  sampling_type = "Perfil completo"
)
```

Arguments

pedons	A list of PedonRecord objects, typically produced by load_bdsolos_csv.
systems	Character vector. Any subset of c("wrb2022","sibcs", "usda"). Default runs all three.
sibcs_level	One of "order" (default) or "subordem". Forwarded to normalise_febr_sibcs.
max_n	Optional integer; cap classification at the first max_n pedons. NULL (default) classifies every pedon.
verbose	If TRUE (default), emits cli progress messages.
sampling_type	The kinds of sampling point the figures of each system are computed on, as site$sampling_type words them. Default "Perfil completo". NULL counts every record, as the function did until v0.9.241. Ignored when no pedon carries the field.

Value

A list with elements:

- per_system – named list (one entry per requested system) of list(accuracy, n_compared, n_correct, n_errors, confusion, per_class, by_sampling_type, sampling_type) (or list(accuracy = NA_real_, message) when no reference labels were present). by_sampling_type is a data frame with one row per kind of point and one for all the records (n_compared, n_correct, accuracy), or NULL when no pedon carries the field; sampling_type is the kinds the main figures were computed on (NULL for all).

- `coverage` – named list of `list(n_with_ref, n_total, pct)` per system.
- `config` – named list capturing `n_pedons, systems, sibcs_level, soilKey_version, timestamp`.

Reference label coverage

BDsolos densely populates `reference_sibcs` (~82 of the v0.9.59 audit) but sparsely populates `reference_wrb` and `reference_st` (UF-dependent; ~5 states). The function always reports the per-system label coverage (`$coverage`) so the caller can judge how representative each accuracy figure is.

Comparison level

SiBCS comparison is at `level = "order"` by default, which converts the BDsolos all-caps Portuguese label (e.g. `"ARGISSOLO VERMELHO Tb EUTROFICO ..."`) to the soilKey plural Title Case form (`"Argissolos"`) via `normalise_febr_sibcs`. Set `sibcs_level = "subordem"` to compare the first two SiBCS tokens (Ordem + Subordem).

WRB and USDA comparisons are at the Reference Soil Group / Order level: `normalise_febr_wrb()` strips qualifier parens and pluralises the bare RSG (`"Xanthic Ferralsol" -> "Ferralsols"`); `normalise_febr_usda()` maps the suffix of the last subgroup token to the USDA Order (`"Typic Haplorthox" -> "Oxisols"`).

Errors and missing-label handling

Pedons without a reference label for a given system are silently excluded from THAT system's comparison (but still classified for the other two systems). If a system has zero pedons with a reference label, the corresponding `$per_system` entry has `accuracy = NA_real_` and `message = "no_reference_labels"`. Classifier errors are caught per-pedon and recorded in `n_errors`; they do not abort the run.

Kinds of sampling point

The export marks each record as `"Perfil completo"`, `"Perfil parcial"` or `"Amostra extra"` (`site$sampling_type`, read by `load_bdsolos_csv` since v0.9.242). An extra sample is not a described profile: in most cases it is two sampled layers, and the thickness a diagnostic horizon is asked for cannot be read from it. The figures of each system (`accuracy`, `n_compared`, `n_correct`, `confusion`, `per_class`) are therefore those of the complete profiles by default, and `by_sampling_type` gives the same counts for every kind and for all the records together. Pedons that carry no `site$sampling_type` (another source, a hand-made list) are all counted, as before.

See Also

`load_bdsolos_csv`, `benchmark_lucas_2018`, `classify_all`, `normalise_febr_sibcs`, `normalise_febr_wrb`, `normalise_febr_usda`.

Examples

```
## Not run:
# Single UF -- typical SiBCS-dense slice
peds <- load_bdsolos_csv("RJ.csv")
```

```
bench <- benchmark_bdsolos(peds, systems = c("sibcs", "wrb2022", "usda"))
bench$coverage      # how many pedons had each reference label
bench$per_system$sibcs$accuracy
bench$per_system$sibcs$confusion

# Subordem level
bench2 <- benchmark_bdsolos(peds, systems = "sibcs",
                            sibcs_level = "subordem")

## End(Not run)
```

`benchmark_isis` *Compare soilKey with ISRIC's names for ISIS profiles*

Description

Classifies each pedon read by `read_isis` with the WRB 2022 and Soil Taxonomy keys and sets the result beside the names ISRIC gives the profile. The WRB comparison uses the Reference Soil Group, and only when ISRIC's name is from the 2022 edition; the USDA one uses the order (ISRIC's USDA names are from the 1975 to 1999 editions, so subgroups are shown but not scored).

Usage

```
benchmark_isis(pedons)
```

Arguments

`pedons`	A list of `PedonRecord`s from `read_isis`.

Value

A data frame, one row per profile, with ISRIC's and soilKey's names, `wrb_agree` and `usda_order_agree` (`NA` where ISRIC gives no comparable name) and the number of close calls of each key.

`benchmark_lucas_2018` *Run the LUCAS Soil 2018 / ESDB WRB benchmark*

Description

For each pedon in `pedons`, attaches the canonical Reference Soil Group at its coordinate via `lookup_esdb`, runs `classify_wrb2022` (or `classify_sibcs`), and tabulates predicted vs reference. Optionally fills missing texture from ISRIC SoilGrids 250m before classifying so that WRB diagnostic horizons that depend on clay (argic, ferralic, nitic) are reachable.

Usage

```
benchmark_lucas_2018(
  pedons,
  esdb_root,
  attribute = "WRBLV1",
  fill_texture_from = NULL,
  fill_topsoil_from = c("none", "soilgrids", "spectra"),
  fill_subsoil_from = c("none", "soilgrids"),
 fill_properties = c("clay", "sand", "silt", "phh2o", "soc", "cec", "bdod", "nitrogen",
    "cfvo"),
  ossl_models = NULL,
  classify_with = c("wrb2022", "sibcs"),
  max_n = NULL,
  soilgrids_lookup_fn = lookup_soilgrids,
  verbose = TRUE
)
```

Arguments

pedons	List of PedonRecord objects, typically from load_lucas_soil_2018.
esdb_root	Path to the unpacked ESDB raster directory (containing the WRBLV1/ sub-folder).
attribute	ESDB attribute to use as reference. Default "WRBLV1" (Reference Soil Group, 31 codes). Other sensible choices: "FAO90LV1" (legacy FAO 1990).
fill_texture_from	Deprecated alias for fill_topsoil_from (v0.9.49 signature). When "soilgrids", treated as fill_topsoil_from = "soilgrids" with fill_properties = c("clay", "sand", "silt") and fill_subsoil_from = "none".
fill_topsoil_from	One of "none" (default), "soilgrids" (fills topsoil 0-20 cm from SoilGrids 250m at 0-5 cm), or "spectra" (runs predict_from_spectra with the supplied ossl_models; pedons must have $spectra$vnir attached, e.g. via attach_lucas_spectra).
fill_subsoil_from	One of "none" (default) or "soilgrids" (synthesises a 30-60 cm B horizon from SoilGrids 250m). Unlocks WRB diagnostic horizons that depend on subsoil features (cambic, argic, mollic).
fill_properties	Character vector of SoilGrids properties to fill when fill_topsoil_from = "soilgrids" or fill_subsoil_from = "soilgrids". Default uses all 9 properties: clay, sand, silt, phh2o, soc, cec, bdod, nitrogen, cfvo. Set to c("clay", "sand", "silt") to recover the v0.9.49 behaviour. cfvo is mapped to coarse_fragments_pct, which drives the Leptosols diagnostic (>= 90 within 25 cm).
ossl_models	Required when fill_topsoil_from = "spectra". A list of soilKey_pls_model objects from train_pls_from_ossl (v0.9.46).
classify_with	One of "wrb2022" (default) or "sibcs".
max_n	Optional integer cap on the number of pedons benchmarked. Useful for quick development runs.

`soilgrids_lookup_fn`
: Internal: SoilGrids lookup function (defaults to `lookup_soilgrids`). Override for unit tests to inject a deterministic stub.

`verbose`
: If `TRUE` (default), prints progress.

Details

This closes Route B of the v0.9.27 EU-LUCAS roadmap end-to-end: v0.9.44 `lookup_esdb` provides the reference label; v0.9.49 (this) provides the loader and the comparison loop; v0.9.48 `lookup_soilgrids` fills texture; v0.9.46 `predict_from_spectra` and v0.9.47 `predict_munsell_from_spectra` can fill the chemistry / Munsell gaps when Vis-NIR is available.

Value

A list with elements:

`predictions` data.frame with one row per pedon: `point_id`, `lon`, `lat`, `country`, `predicted`, `reference_code`, `reference_name`, `agree`.

`confusion` Confusion table (predicted vs reference) over in-scope rows.

`accuracy` Overall fraction of correct classifications among in-scope rows.

`per_rsg` Per-RSG recall data.frame.

`n_in_scope` Number of pedons with both predicted and reference set.

`n_total` Total pedons benchmarked.

`n_errors` Number of pedons where the classifier errored out.

`errors` List of `(i, id, error)` tuples for classifier errors.

`config` Recap of arguments used.

See Also

`load_lucas_soil_2018`, `lookup_esdb`, `lookup_soilgrids`.

Examples

```
## Not run:
pedons <- load_lucas_soil_2018(
  "soil_data/eu_lucas/LUCAS-SOIL-2018-data-report-readme-v2/LUCAS-SOIL-2018-v2",
  countries = c("ES"), max_n = 50)
bench <- benchmark_lucas_2018(
  pedons,
  esdb_root = "soil_data/eu_lucas/ESDB-Raster-Library-1k-GeoTIFF-20240507",
  fill_texture_from = "soilgrids")
bench$accuracy
bench$per_rsg

## End(Not run)
```

benchmark_natsoil *Benchmark the Australian order key on NatSoil*

Description

Classifies NatSoil pedons with `classify_asc3` and compares the order with the one the surveyors gave (`site$reference_asc_order`). A pedon is decided when no order before the assigned one was left undecided for missing data; the agreement is given for all pedons and for the decided ones. The surveys predate the 3rd edition's Arenosols (2021), whose soils earlier editions placed in the Tenosols and Rudosols: the *edition* figures count an Arenosol given to a surveyed Tenosol or Rudosol as agreement.

Usage

```
benchmark_natsoil(pedons, cores = 1L, verbose = TRUE, family = FALSE)
```

Arguments

`pedons`	A list of `PedonRecord` from `load_natsoil_pedons`.
`cores`	Number of processes (`parallel::mclapply`; always 1 on Windows).
`verbose`	Print a summary.
`family`	Also derive the families (`classify_asc3(include_family = TRUE)`) and compare them (since 0.9.292).

Value

A list: `n`; `accuracy` and `accuracy_edition`; `decided` (count), `decided_accuracy` and `decided_accuracy_edition`; `by_confidence` (a data frame by the surveyors' confidence, 1 = all necessary analytical data); `recall` by surveyed order; `confusion` (rows the survey, columns soilKey); `first_undecided` (the first order left undecided, among provisional pedons); `suborders`, the same for the suborders where the key has them (overall, and over the pedons whose order agrees); `great_groups`, the same for the great groups (overall, and over the pedons whose order and suborder agree); `subgroups`, the same for the subgroups (over the pedons whose order, suborder and great group agree); `families` (with `family = TRUE`), each family criterion against the surveyors' family codes (`site$reference_asc_family`), over the pedons whose order (and, for the Hydrosols, suborder) agrees; and `results`, one row per pedon.

See Also

`load_natsoil_pedons`, `classify_asc3`.

benchmark_npdb *Benchmark the Canadian key against the National Pedon Database*

Description

Classifies NPDB pedons (load_npdb_pedons) with classify_cssc3 and compares the result with the survey's classification at the level asked.

Usage

```
benchmark_npdb(
  pedons,
  level = c("order", "great_group", "subgroup"),
  set_aside = "Chernozemic",
  cores = 1L,
  verbose = TRUE
)
```

Arguments

pedons	A list of PedonRecord from load_npdb_pedons.
level	"order", "great_group" or "subgroup".
set_aside	Reference orders reported apart (character(0) for none).
cores	Number of processes (parallel::mclapply; always 1 on Windows).
verbose	Print a summary.

Details

The records hold no soil climate and no dry colour, which the key reads for the Chernozemic order and its great groups (Ch 3, p. 34; Ch 5, p. 63) and for the Gray Brown Luvisols; the orders named in set_aside (by default the Chernozemic) are therefore reported apart, with where they land, and left out of accuracy. A result is provisional when an undecided class comes before the one assigned ($provisional_because of classify_cssc3); accuracy_decided counts only the results that are not provisional at the level asked or above.

Pedons from npdb_scenario carry estimates as ranges; each is classified at its central estimate and at every combination of the ends of its ranges (and of the moisture subclasses between its two estimates), and a class counts as decided only where every version gives it and none is provisional (otherwise the central class is reported as provisional). Run such pedons with set_aside = character(0) to include the Chernozems.

Value

A list: level; accuracy (share of agreement outside set_aside); accuracy_decided; n_compared; n_decided; accuracy_all (all orders); set_aside (a table of where those pedons land); confusion; per_class_recall; predictions (one row per pedon; scenario marks pedons classified at the ends of their ranges and scenario_split those whose versions disagree).

benchmark_nsd *Benchmark the New Zealand key against the National Soils Database*

Description

Classifies NSD pedons (read_nsd_pedons) with classify_nzsc4 and compares the result with the surveyor's NZSC class at the level asked: the order (the first letter of the code), the group (the first two letters) or the subgroup (the whole code). A pedon is compared at a level only when its recorded class reaches that level (a code of one letter names only the order). A result is provisional when a class before the assigned one at the level asked, or a level above it, could not be decided ($provisional_because of classify_nzsc4); accuracy_decided counts only the results that are not provisional.

Usage

```
benchmark_nsd(
  pedons,
  level = c("order", "group", "subgroup"),
  cores = 1L,
  verbose = TRUE
)
```

Arguments

pedons	A list of PedonRecords from read_nsd_pedons.
level	"order", "group" or "subgroup".
cores	Number of processes (parallel::mclapply; always 1 on Windows).
verbose	Print a summary.

Value

A list: level; accuracy; accuracy_decided; n_compared; n_decided; confusion (reference by prediction, codes); per_class_recall; predictions (one row per pedon with a reference: the recorded and predicted codes at every level, agree, provisional and provisional_because).

benchmark_performance *Run the soilKey performance benchmark*

Description

Generates n synthetic pedons (5 horizons each, with the chemistry / morphology populated for typical Argissolo / Latossolo / Cambissolo cases), calls each classifier on each pedon, and reports per-call latency + total throughput.

Usage

```
benchmark_performance(
  n = 100L,
  systems = c("wrb2022", "sibcs", "usda"),
  include_familia = FALSE,
  seed = 42L,
  verbose = TRUE
)
```

Arguments

n	Integer. Number of synthetic pedons to generate. Default 100; pass 1000 or higher for batch-level measurements.
systems	Character vector. Which classifiers to time. Default `c("wrb2022", "sibcs", "usda")` (all three).
include_familia	Pass-through to `classify_sibcs` when `"sibcs"` is in `systems`. Default `FALSE`.
seed	RNG seed for reproducibility. Default 42.
verbose	If `TRUE` (default), prints a per-system summary line.

Details

Designed to be a one-shot reproducible benchmark: the synthetic pedons use a fixed RNG seed so timings on the same machine are comparable across releases.

Value

A list with elements:

`summary` data.frame: `system`, `n_pedons`, `total_seconds`, `mean_seconds`, `median_seconds`, `pedons_per_minute`.

`per_pedon` data.frame with one row per (pedon, system) call: `i`, `system`, `seconds`, `status`.

`config` list with `n`, `seed`, `soilKey_version`, `R_version`, `platform`.

Examples

```
## Not run:
bench <- benchmark_performance(n = 100)
bench$summary
#>     system n_pedons total_seconds mean_seconds median_seconds pedons_per_minute
#> 1 wrb2022      100        ~ 5-12      0.05-0.12      ~          ~
#> 2   sibcs      100        ~ 5-15      0.05-0.15      ~          ~
#> 3    usda      100        ~ 4-10      0.04-0.10      ~          ~

## End(Not run)
```

benchmark_redape *Benchmark soilKey SiBCS predictions against the Redape gold standard*

Description

Runs `classify_sibcs` on each pedon and compares against the curator-validated reference label (Order / Suborder / Great Group / Subgroup). Returns per-level accuracy and the confusion matrix at the requested granularity.

Usage

```
benchmark_redape(
  pedons,
  level = c("order", "subordem", "gde_grupo", "subgrupo"),
  verbose = TRUE
)
```

Arguments

`pedons`	List of `PedonRecord` objects (typically from `load_redape_pedons`).
`level`	One of `"order"` (default), `"subordem"`, `"gde_grupo"`, or `"subgrupo"`.
`verbose`	Print progress (default `TRUE`).

Value

A list with `accuracy`, `n_compared`, `confusion`, `per_class_recall`, and the per-pedon `predictions` table. `predictions` now also includes columns `ref_norm` and `pred_norm` – the canonical comparison keys – for downstream auditing.

v0.9.81 level-aware comparison

Earlier versions accepted the `level` argument but always used `rsg_or_order` for the prediction and the order field for the reference, so all four levels reported identical accuracy. v0.9.81 reads the level-specific slots from `res$trace` (subordem, grande_grupo, subgrupo) and concatenates the matching reference fields, applying SiBCS-aware Portuguese pluralisation so the comparison key matches the predictor's plural Title Case form.

benchmark_run_classification *Run a benchmark across one of the loaded pedon lists*

Description

Classifies each pedon in pedons against the named system, compares against the published reference (e.g. site$reference_wrb), and returns a confusion matrix + top-1 / top-3 accuracy + bootstrap CI on top-1.

Usage

```
benchmark_run_classification(
  pedons,
  system = c("wrb2022", "sibcs", "usda"),
  level = c("order", "subgroup", "subordem", "great_group", "suborder"),
  boot_n = 1000L
)
```

Arguments

pedons	List of PedonRecord objects (output of one of the load_* functions).
system	One of "wrb2022", "sibcs", "usda".
level	Granularity of the comparison: • "order" (default) – the top-level RSG / Ordem / Order, compared against cls$rsg_or_order; • "subgroup" – the full classified name (Subgroup in USDA, Subgrupo in SiBCS, RSG + qualifiers in WRB), compared against cls$name after case-insensitive token normalisation; • "subordem" – SiBCS-only, the 2nd-level "Ordem + Subordem" (e.g. "Latossolos Vermelhos"). Comparison via the first two normalised tokens of the predicted name vs the reference; • "great_group" (USDA, v0.9.24) – the LAST token of the subgroup name (e.g. "typic hapludalfs" -> "hapludalfs"). Isolates whether the Great Group machinery is correct independent of subgroup modifiers (Typic / Aquic / Vertic / Cumulic / Pachic / etc.). Reads site$reference_usda_grtgroup; • "suborder" (USDA, v0.9.24) – maps the Great Group prediction to its canonical Suborder suffix ("hapludalfs" -> "udalfs") using the KST 13ed Ch 4 ~70-Suborder list. Reads site$reference_usda_suborder.
boot_n	Bootstrap replicates for CI (default 1000).

Value

A list with elements accuracy_top1, accuracy_ci, confusion, and per_pedon (one row per pedon with predicted vs reference).

benchmark_sali *Benchmark the Australian key on Queensland SALI*

Description

Classifies SALI pedons (read_sali_pedons) with classify_asc3 and compares them with the surveyors' classification, with the metrics of benchmark_natsoil (which it calls): the order agreement over all pedons and over the decided ones (no order before the assigned one left undecided for missing data), the Arenosols of the 3rd edition counted as Tenosols or Rudosols in the *edition* figures, the suborders given the order, the great groups given the suborder and the subgroups given the great group. It adds a table by surveyed order and one by the ASC edition the report cites.

Usage

```
benchmark_sali(pedons, cores = 1L, verbose = TRUE, family = FALSE)
```

Arguments

pedons	A list of PedonRecord from read_sali_pedons.
cores	Number of processes (parallel::mclapply; always 1 on Windows).
verbose	Print a summary.
family	Also compare the families (see benchmark_natsoil; SALI labels carry no family codes, so this adds nothing yet).

Value

The list of benchmark_natsoil, plus by_order (by surveyed order: n, agree and decided_agree in per cent, decided, and soilKey's most frequent order, top_prediction) and by_edition (the same by the edition of the ASC reference, site$sali_asc_edition).

See Also

read_sali_pedons, benchmark_natsoil.

benchmark_spectral_fill *Benchmark the accuracy lift of spectral gap-fill (ON vs OFF), k-fold*

Description

The honest measurement that has been data-blocked until a spectra-bearing, labelled dataset exists. For each cross-validation fold it calibrates a spectral library on the training profiles, then classifies the held-out profiles twice – **OFF** (spectra-only pedon, no lab attributes) and **ON** (fill_from_spectra predicts the lab attributes from the scan first) – and scores both against the reference label. Non-circular: the calibration library never includes a test profile.

Usage

```
benchmark_spectral_fill(
  reflectance,
  metadata,
  id_col = "id",
  system = c("sibcs", "wrb2022", "usda"),
  profile_col = NULL,
  folds = 5L,
  properties = NULL,
  method = c("mbl", "plsr_local", "pretrained"),
  wavelengths = NULL,
  resample_to = NULL,
  property_map = NULL,
  label_map = NULL,
  normalize = c("auto", "none", "percent"),
  fold_id = NULL,
  verbose = TRUE
)
```

Arguments

`reflectance`	Reflectance data: a matrix / data.frame with rows = samples and columns named by wavelength (nm); OR a long data.frame with `id_col`, `wavelength_nm`, `reflectance`; OR a path to a CSV in either form.
`metadata`	A data.frame with one row per sample carrying `id_col` plus lab attributes and optional taxonomic labels and `lat`/ `lon`. Rows are aligned to `reflectance` by `id_col`.
`id_col`	Sample identifier column shared by both tables (default `"id"`).
`system`	One of `"sibcs"` (default), `"wrb2022"`, `"usda"`.
`profile_col`	Column grouping samples into profiles (default `id_col`).
`folds`	Number of CV folds (default 5).
`properties`	Attributes to predict from spectra (default the `fill_from_spectra` set).
`method`	Spectral model: `"mbl"`, `"plsr_local"` or `"pretrained"` (passed to `fill_from_spectra`).
`wavelengths`	Optional explicit wavelength vector (nm) when the reflectance columns are not wavelength-named.
`resample_to`	Optional target wavelength grid (nm) to linearly resample every spectrum onto (e.g. `350:2500`); default keeps the native grid.
`property_map, label_map`	Optional named lists overriding the alias auto-detection, e.g. `property_map = list(clay_pct = "ARGILA")`.
`normalize`	One of `"auto"` (divide by 100 when values look like percent), `"percent"`, or `"none"`.
`fold_id`	Optional integer vector (one per profile, in sorted-id order) to use fixed folds instead of the deterministic modulo split.
`verbose`	Print a one-line summary (default `TRUE`).

Value

A list with `accuracy_off`, `accuracy_on`, `delta`, `n`, per-fold rows, and the per-profile `predictions` frame.

See Also

`read_spectral_library`, `fill_from_spectra`

`benchmark_unified` *Unified cross-dataset benchmark across SiBCS / WRB / USDA*

Description

Runs a system's soilKey classifier on every dataset that has reference labels for that system, then pools the results into a single nation-/world-wide accuracy estimate.

Usage

```
benchmark_unified(
  systems = c("all", "wrb2022", "sibcs", "usda"),
  datasets = c("all", "bdsolos", "febr", "kssl", "lucas_esdb"),
  paths = NULL,
  max_n_per_dataset = NULL,
  engine = c("soilkey", "aqp", "both"),
  harmonize = FALSE,
  gapfill = FALSE,
  verbose = TRUE
)
```

Arguments

`systems`	Character vector. Any subset of `c("wrb2022","sibcs", "usda")`. Default `"all"` runs all three.
`datasets`	Character vector. Any subset of `c("bdsolos", "febr", "kssl", "lucas_esdb")`. Default `"all"` pools every dataset that has reference labels for the requested systems. Datasets without reference labels for a system are silently excluded from that system's pooled result.
`paths`	Named list of dataset paths. Element names should match those in `datasets`. If `NULL` (default), soilKey looks for canonical paths under `"~/soil_data/"`.
`max_n_per_dataset`	Optional integer to cap per-dataset sample size (useful for development / debugging). `NULL` (default) classifies every available pedon.
`engine`	Currently forwarded to Phase-1 aqp wiring. One of `"soilkey"` (default), `"aqp"`, `"both"`. When `"aqp"`, sets `options(soilKey.diagnostic_engine = "aqp")` for the duration of the benchmark, which routes `argic()` / `cambic()` through the canonical `aqp::getArgillicBounds` / `getCambicBounds`.

harmonize — If TRUE (default FALSE), applies harmonize_to_gsm to each dataset's pedons before classification, putting all chemistry/texture on the GSM depth grid (0-5 / 5-15 / 15-30 / 30-60 / 60-100 / 100-200 cm). Required for cross-dataset pooling integrity (Phase 2.3) but slow (~1-2 min for 1k pedons) and may degrade per-dataset accuracy slightly because the splined depths are approximations.

gapfill — If not FALSE (the default), applies gapfill_within_pedon to each dataset's pedons before classification, filling interior NA cells of the continuous depth-trending attributes by within-pedon linear interpolation. Accepts the same values as the gapfill argument of classify_all (TRUE, a character vector of attributes, or a named list). Lets you measure the ON/OFF accuracy lift of gap-fill reproducibly through the harness.

verbose — If TRUE (default), emits cli progress.

Value

A list with elements:

- per_system – per-system pooled list(accuracy, n_compared, n_correct, confusion, per_class).
- per_system_per_dataset – per-(system, dataset) same shape, for breakdown.
- coverage – per-(system, dataset) sample sizes and label coverage.
- config – captures systems, datasets, engine, soilKey_version, timestamp.

Datasets and their reference labels

Dataset	Systems with reference labels
BDsolos	SiBCS (dense), WRB (sparse), USDA (sparse)
FEBR superconjunto	SiBCS, WRB, USDA (most rows have all 3)
KSSL+NASIS	USDA only (samp_taxsubgrp universal)
LUCAS + ESDB raster	WRB (via lookup_esdb on coords)

For each (system, dataset) pair, this function:

1. Loads pedons via the appropriate load_* helper.
2. Filters to pedons with a populated reference label for the requested system.
3. Normalises both reference and predicted labels via normalise_febr_*() / KSSL canonicalisation helpers.
4. Calls the system's classifier and records pred-vs-ref.

Then pools per-system results across datasets.

Engine selection (Phase 1 wiring)

For datasets with morphological data (BDsolos / FEBR), the diagnostics that pivot Argissolos / Latossolos / Cambissolos classification can be run with two engines:

- `engine = "soilkey"` (default) – the hand-coded WRB 6/1.4/20 thresholds.
- `engine = "aqp"` – aqp::getArgillicBounds / getCambicBounds (KST 13ed 3/1.2/8 thresholds).

On the v0.9.62 RJ benchmark (722 perfis), aqp was 14.8 pp stricter on argic and 40.6 pp more permissive on cambic; the SiBCS Argissolos / Latossolos / Cambissolos boundary is sensitive to both. `engine` is currently forwarded to a future v0.9.63 wired `argic()` / `cambic()`; for now, `benchmark_unified()` reports separately per engine when `engine = "both"`.

See Also

`benchmark_bdsolos`, `benchmark_lucas_2018`, `benchmark_run_classification`, `harmonize_to_gsm`.

---

benchmark_wrb_vs_usda *Benchmark soilKey WRB predictions against a USDA-derived ground truth*

Description

Convenience wrapper: applies `annotate_wrb_from_usda` to attach derived WRB labels, runs `classify_wrb2022` on each pedon, and returns top-1 accuracy + per-RSG recall.

Usage

```
benchmark_wrb_vs_usda(pedons, verbose = TRUE)
```

Arguments

`pedons`	List of `PedonRecord` objects with `site$reference_usda` populated (typically from `load_kssl_pedons_gpkg`).
`verbose`	Print progress.

Value

A list with `accuracy`, `n_compared`, `confusion`, `per_class_recall`.

build_taxon_profiles *Build per-taxon mean depth profiles for predicted-taxon gap-fill*

Description

For each taxon (the first word of the reference label at the requested level), averages each attribute across the calibration pedons into the six standard depth slices (0-5 ... 100-200 cm). The result feeds gapfill_by_predicted_taxon. Calibrate on a set DISJOINT from the pedons you will fill (e.g. a train split) to keep the fill non-circular.

Usage

```
build_taxon_profiles(pedons, ref_field = "reference_sibcs", attrs = NULL)
```

Arguments

pedons	A list of PedonRecord with a reference label.
ref_field	Site field holding the reference label (default "reference_sibcs"; e.g. "reference_usda" / "reference_wrb").
attrs	Attributes to profile (default the continuous gap-fill set).

Value

A named list taxon -> attr -> numeric(6) (NA where a taxon has no measured value in a slice).

See Also

gapfill_by_predicted_taxon

B_espodico *Horizonte B espodico (SiBCS Cap 2, p 62-65; v0.7)*

Description

Subsuperficial com acumulo iluvial de Al + Fe + materia organica; espessura \>= 2.5 cm. Tipos: Bs, Bhs, Bh, ortstein. Reuso de spodic (WRB) que ja codifica criterios essencialmente identicos.

Usage

```
B_espodico(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Reserved for future arguments.

Details

In the morphological mode the horizon is also read from its designation. Since v0.9.247 that includes the "ir" of the older surveys (`Bir`, `B2ir`, `B22irx`), the illuvial iron that the present notation writes "s": the mode read "h" and "s" only.

Since v0.9.255 a horizon that the WRB test does not find is looked for by item (a) of the book (p. 65), in both modes. The B espodico has an "acumulacao iluvial de compostos organometalicos" and, by that item, "um horizonte E (albico ou nao) sobrejacente e cores umidas de acordo com um dos itens a seguir: 1) matiz 5YR ou mais vermelho; 2) matiz 7,5YR com valor 5 ou menor e croma 4 ou menor; 3) matiz 10YR, com valor e croma 3 ou menor; 4) cores neutras com valor 3 ou menor". A B layer of 2.5 cm or more passes when:

- an E lies above it in the profile;
- its moist colour is one of the four;
- its organic carbon is 0.5% or more and at least twice that of the nearest E above it; and
- no cerosidade is recorded in it.

The last two stand for the illuvial accumulation, which the sheets do not measure where there is no oxalate extraction. The first takes the form the book gives in item (b)(3) for aluminium and iron by oxalate, "valor 0,50 ou maior, sendo este valor pelo menos o dobro do encontrado no horizonte sobrejacente"; the second keeps out the B whose illuvial material is clay. It is a reading by analogy, decided on 2026-10-13, and not the letter of the book: the colours alone are those of the B textural of many an Argissolo below an E. Items (b) and (c), the cementation and the ortstein, are not tested.

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`B_incipiente` *Horizonte B incipiente (SiBCS Cap 2, p 59-61; v0.7)*

Description

Subsuperficial sob A/Ap/AB com alteracao fisica e quimica incipiente, NAO satisfazendo a B textural / latossolico / nitico / espodico / planico, com:

- espessura \>= 10 cm;
- textura francoarenosa ou mais fina;
- < 50% estrutura da rocha original;
- evidencias de pedogenese (cor mais viva OR remocao de carbonatos OR designation `Bw`/`Bi`);
- NAO satisfaz: argic, ferralic, espodic, planic, e nao tem duripa/petrocalcico/fragipa.

Usage

```
B_incipiente(pedon, min_thickness = 10)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

B_latossolico *Horizonte B latossolico (SiBCS Cap 2, p 57-59)*

Description

The requirements of the book that the data of a profile can show (v0.9.241):

- (d) texture francoarenosa or finer;
- (g) cation exchange capacity below 17 cmolc per kg of clay, with no correction for carbon;
- (e) a molecular ratio SiO2/Al2O3 (Ki) of 2.2 or less, where the sulfuric attack is recorded (sio2_sulfuric_pct, al2o3_sulfuric_pct): a layer with a Ki above 2.20 is no B latossolico (v0.9.243);
- (h) cerosidade, if present, at most few and weak: a layer with a larger amount or a higher grade is no B latossolico (a range counts by its upper end, the requirement being a maximum);
- (b) a thickness of 50 cm or more, of the horizon: of the adjoining sub-horizons that meet the other requirements, taken together;
- it "nao apresenta caracteristicas diagnosticas de horizontes glei, B textural, B nitico e plintico": the layers of a B_textural, of a horizonte glei, of a horizonte plintico and of a B_nitico are no B latossolico. Since v0.9.249 they take out their own layers only: until then the glei and the plinthic horizon were those of WRB (gleyic properties, plinthic horizon), and one of them, or a B nitico, anywhere in the profile voided the B latossolico of the whole profile.

Where the designations are recorded, only B horizons are candidates: the B latossolico is a subsurface horizon below the A. Items (a) structure, (c) rock structure and (f) weatherable minerals are not tested.

Usage

```
B_latossolico(
  pedon,
  min_thickness = 50,
  max_cec_per_clay = NULL,
  engine = NULL,
  ...
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness of the horizon, in cm (50).
max_cec_per_clay	CEC per kg clay the horizon must stay below. Defaults to NULL (engine-aware): 17 in soilkey engine (SiBCS Cap 2: "menor que 17 cmolc kg-1 de argila, sem correcao para carbono") or 20 in aqp engine (v0.9.68 regional tolerance for Embrapa lab methodology offset). Strict since v0.9.212: a value of exactly 17 is not latossolic, it is the B incipiente's "17 ou maior".
engine	One of "soilkey" (default) or "aqp"; NULL reads getOption("soilKey.diagnostic_engine").
...	Reserved for future arguments.

Details

Until v0.9.240 the 50 cm were asked of each layer on its own, so that a B latossolico described as B1, B21 and B22 of 35 cm each was none; and since v0.9.61 a B textural did not exclude it unless its clay films were common or more, on a rule the comments attributed to Cap 18 of the book, which deals with families and series.

In the morphological mode a "Bw" or "Bo" of a tropical profile stands for the B latossolico when the data do not show it.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

B_nitico *Horizonte B nitico (SiBCS Cap 2, p 61-62; v0.7)*

Description

Subsuperficial nao hidromorfico, textura argilosa/muito argilosa (clay \>= 35% desde a superficie), com pequeno incremento de argila (B/A \<= 1.5), estrutura em blocos sub/angulares ou prismatica grau moderado/forte, cerosidade no minimo comum + moderada, espessura \>= 30 cm. Argila ativ baixa OR ativ alta + carater aluminico.

Usage

```
B_nitico(
  pedon,
  min_thickness = 30,
  min_clay_pct = 35,
  max_b_a_ratio = 1.5,
  min_cerosidade = c("common", "many", "abundant", "strong")
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_clay_pct	Numeric threshold or option (see Details).
max_b_a_ratio	Numeric threshold or option (see Details).
min_cerosidade	Terms of the amount of cerosidade that satisfy criterion (c). The lowest of them on the scale none < few < common < many sets the threshold (the default is "common", the "comum" of SiBCS), and an amount of that class or above passes, in English or Portuguese. A term the scale does not know passes only if it is literally listed.

Details

The amount of cerosidade (clay_films_amount) and the type of structure (structure_type) are read as the field sheet words them, in English or in Portuguese, whatever the case and the accents: "Comum", "Abundante" and "Prismatica que se desfaz" (written with its accent) of the Brazilian surveys count as "common", "many" and "prismatic" do. Until v0.9.226 the amount was compared with the literal English terms of min_cerosidade, so no profile described in Portuguese could have a B nitico.

The criteria of the B nitico are minima ("cerosidade em quantidade no minimo comum e com grau forte ou moderado", "estrutura ... de grau de desenvolvimento moderado ou forte"), and a range meets a minimum only if all of it does: an amount or a grade written as a range is read by its lower end. "Pouca a comum" is below "comum" and, since v0.9.231, "Fraca a moderada" is below "moderada" for the cerosidade and for the structure.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

B_planico *Horizonte B planico (SiBCS Cap 2, p 65-66)*

Description

"E um tipo especial de horizonte B textural, com ou sem carater sodico, subjacente a horizontes A ou E, apresentando mudanca textural abrupta ou transicao abrupta associada a relacao textural com valor dentro do especificado para o horizonte B textural (subitem h), porem calculado entre o primeiro sub-horizonte B e o horizonte imediatamente acima (A ou E)".

Usage

```
B_planico(pedon)
```

Arguments

pedon	A PedonRecord.

Details

Since v0.9.244 the horizon is looked for where the book puts it, in the first B sub-horizon, below an A or an E, by one of the two:

- a mudanca_textural_abrupta from the A or E to it; or
- a transition recorded as abrupt (boundary_distinctness) with a ratio of its clay to that of the horizon above over the limit of item (h) of the B textural (1.50, 1.70 or 1.80).

As a type of B textural, that sub-horizon has a texture francoarenosa or finer (p. 54): one known to be coarser is no B planico, whatever its clay over that of the E. The structure and the colours below are asked of that first sub-horizon. Until v0.9.243 any layer with a jump of clay over the layer above it could be the B planico, the second way was not tested, and the structure and the colours were asked of the layer with the jump. A profile with no horizon designation at all is still tested that way.

The structure (p. 65): "Apresenta estrutura prismatica, colunar ou em blocos angulares e subangulares grandes ou medios e, as vezes, estrutura macica". It is tested where the sheet records it (structure_type, structure_size): a prismatic, columnar or massive structure passes; blocks pass when medium or larger, or when their size is not recorded; any other recorded structure does not pass.

The colours "devem atender a pelo menos um dos seguintes requisitos" (p. 65-66):

- (a) the matrix colour, with or without mottles: "Matiz 10YR ou mais amarelo, cromas <= 3" or "Matizes 7,5YR ou 5YR, cromas <= 2";
- (b) "Coloracao variegada com pelo menos uma cor apresentando matiz e croma conforme especificado no item (a)" (variegated_colors);
- (c) "matiz 10YR ou mais amarelo, cromas >= 4, combinado com mosqueado de croma conforme especificado no item (a)": a mottle colour (mottle_colors) that satisfies item (a).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

B_textural *Horizonte B textural (SiBCS Cap 2, p 54-57)*

Description

The B textural as the book states it: a subsurface horizon of texture francoarenosa or finer, at least 7.5 cm thick (thickness items a to e), that meets one or more of the items below (v0.9.241).

- (f) an E horizon above the B, the B having more clay than the horizons above it and being no B espodico, plintico or planico;
- (g) a mudanca textural abrupta from the A or E to the B (Cap 1, p 30-31): at least twice the clay when the A or E has less than 200 g/kg, 200 g/kg more otherwise, within 7.5 cm. The distance is read from the distinctness of the transition (`boundary_distinctness`: abrupt or clear); where it is not recorded the two layers are compared;
- (h) the relacao textural B/A of footnote 4, above 1.50, 1.70 or 1.80 according to the clay of the A (`test_ratio_textural_sibcs`);
- (i) where the increase is below item (h), the cerosidade of one or more sub-horizons of the upper part of the B (read as its upper half): (1) texture media with no macroaggregates and cerosidade of moderate grade; (2) texture media with prismatic or blocky structure of moderate or strong grade and cerosidade of at least moderate grade; (3) texture argilosa or muito argilosa with that structure and cerosidade at least common and weak or few and moderate; (4) relacao textural of 1.4 or more with a fragipa within 200 cm, the B being no B espodico;
- (j) with a discontinuity of parent material between the A or E and the B, or with only a ploughed layer above the B, items (h) or (i) are needed.

Amounts and grades written as ranges are read by their lower end, the requirements being minima. Lamellae (item d) are not tested.

Usage

```
B_textural(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Passed to argic for a profile with no horizon designation; ignored otherwise.

Details

Until v0.9.240 this function was the WRB argic horizon, whose clay increase (6 percentage points, a ratio of 1.4 or 20 points, between a layer and the least clayey layer anywhere above it) is not the one of the book, joined to item (h); items (f), (g) and (i) were not tested.

The layers returned are the B sub-horizons (BC excluded) of texture francoarenosa or finer; `evidence$upper_layers` are those that carry the evidence, from the top of the B down.

The items are stated for the A, E and B horizons. A profile with no horizon designation at all (depth slices, for instance) cannot be read by them and is tested as before, by the clay increase between layers.

In the morphological mode (`options(soilKey.morphological_inference = TRUE)`) a "t" in the designation of a B horizon stands for the B textural when the items do not show it. The diagnostic engine (`options(soilKey.diagnostic_engine = "aqp")`) does not change this test: it changed the argic horizon that this function used to be.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

calcaric_material *Calcaric material (WRB 2022 Ch 3.3.3): \>= 2% CaCO3 throughout the fine earth, primary carbonates from the parent material.*

Description

v0.9.315: WRB 2022 Ch 3.3.3 (p. 86), the material "contains >= 2 carbonate equivalent" and "shows visible effervescence with 1 M HCl throughout the fine earth". Each layer is read as the Calcaric qualifier reads it (since v0.9.311): the CaCO3 equivalent decides where it is recorded (caco3_pct >= min_caco3_pct), otherwise the field reaction (effervescence_class: moderate or stronger is visible effervescence, none and slight are not). Before v0.9.315 only caco3_pct was read, so a profile described with the HCl reaction alone was never calcaric. TRUE when some layer is calcaric material, FALSE when every layer is not, NA otherwise.

Usage

```
calcaric_material(pedon, min_caco3_pct = 2)
```

Arguments

pedon	A PedonRecord.
min_caco3_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

calcic *Calcic horizon (WRB 2022)*

Description

Tests whether any horizon meets the calcic horizon criteria. The calcic horizon is a horizon of secondary carbonate accumulation, diagnostic for Calcisols and qualifying many other RSGs.

Usage

```
calcic(pedon, min_thickness = 15, min_caco3_pct = 15)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness in cm (default 15).
min_caco3_pct	Minimum CaCO3 percent in fine earth (default 15).

Details

Sub-tests called:

- test_caco3_concentration – CaCO3 >= 15%.
- test_minimum_thickness – thickness >= 15 cm.

v0.2 limitations: the WRB criterion of "5% absolute or relative more CaCO3 than the underlying horizon" is not enforced; this captures true calcic horizons but may also mark uniformly carbonate-rich substrates that are not pedologically calcic. Cementation (petrocalcic) is not yet detected. Both refinements are scheduled for v0.3.

Value

A DiagnosticResult.

v0.9.314

The WRB calcic horizon follows WRB 2022 Ch 3.1.4 (p. 37-38) criterion by criterion: criterion 2 is protocalcic properties (secondary carbonates on $\geq$ 5% of the exposed area, secondary_carbonates_pct) or $\geq$ 5% more CaCO3 than an underlying layer with no lithic discontinuity between; a layer of a petrocalcic horizon is not calcic; the 15 cm are those of the adjoining layers. A layer whose criterion 2 cannot be decided (secondary carbonates not recorded) is undecided, not calcic. The USDA calcic horizon and the SiBCS horizonte calcico keep the earlier reading.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition. International Union of Soil Sciences, Vienna. Chapter 3 – Calcic horizon.

---

cambic *Cambic horizon (WRB 2022)*

Description

Tests whether any horizon meets the cambic horizon criteria. The cambic horizon is a subsurface horizon with evidence of pedological alteration that does not meet the criteria for any stronger diagnostic horizon. It is the diagnostic of Cambisols.

Usage

```
cambic(pedon, min_thickness = 15, min_top_cm = 5, engine = NULL)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness in cm (default 15).
min_top_cm	Minimum top depth (cm) for a horizon to be considered cambic-eligible (default 5). Anchors the candidate set to subsurface layers. Since v0.9.314 the WRB reading asks it only when it is given.
engine	v0.9.63+. One of "soilkey" (hand-coded path, default for back-compat) or "aqp" (canonical NRCS dispatch via aqp::getCambicBounds). When NULL (the new default) the function reads getOption("soilKey.diagnostic_engine", "soilkey"), so a global options(soilKey.diagnostic_engine = "aqp") flips every cambic() call without modifying call sites. The aqp engine fired 40.6 soilkey 0 v0.9.50 LUCAS WRB benchmark from 0 100 cambic_aqp.

Details

v0.2 implementation tests three conditions:

- thickness >= 15 cm (test_minimum_thickness)
- texture sandy loam or finer (test_texture_argic)
- NOT argic AND NOT ferralic

v0.2 limitations: WRB 2022 also excludes profiles with spodic, calcic, gypsic, plinthic, vertic, and several other diagnostic horizons. Those exclusions, plus the WRB criteria of "evidence of alteration" (color/structure differences from parent material, carbonate removal), are scheduled for v0.3.

Value

A DiagnosticResult.

v0.9.314

The WRB cambic horizon follows WRB 2022 Ch 3.1.5 (p. 39-40): criterion 4 is read layer by layer, so a layer that forms part of an argic, ferralic, calcic, mollic or other listed horizon (or of a layer with andic properties) is set aside, not the whole profile; the subsurface layers are those below the layer at the mineral soil surface that are not A horizons; the 15 cm are those of the adjoining layers; a layer whose texture or structure is not recorded leaves the result undecided. Criterion 3 is not tested. The USDA keys keep the earlier reading.

References

IUSS Working Group WRB (2022), Chapter 3, Cambic horizon.

cambic_aqp *Cambic horizon via aqp::getCambicBounds()*

Description

Wraps `aqp::getCambicBounds()` in soilKey's `DiagnosticResult` contract. The aqp test enforces the KST 13ed cambic criteria:

- Texture finer than loamy fine sand (i.e. NOT in the sandy-texture pattern).
- Soil structure or absence of rock structure.
- Evidence of pedogenic alteration (chroma / value / clay).
- NOT meeting argic / oxic / spodic / mollic criteria.

soilKey's `cambic` (and the SiBCS proxy `B_incipiente`) implements similar logic but with SiBCS / WRB-flavoured exclusions; the aqp engine here is an independent canonical reference.

Usage

```
cambic_aqp(pedon, argi_bounds = NULL, ...)
```

Arguments

`pedon`	A `PedonRecord`.
`argi_bounds`	Optional `c(ubound, lbound)` for argillic bounds (forwarded to aqp). `NULL` (default) means the aqp internals re-detect.
`...`	Reserved for future arguments.

Value

A `DiagnosticResult` with `name = "cambic_aqp"`.

See Also

`cambic` (soilKey hand-coded), `aqp::getCambicBounds`.

`canonicalise_kst13ed_gg` *Canonicalise a USDA Great Group label to a KST 13ed-compatible key*

Description

Maps both obsolete (pre-KST 13ed) and modern Great Group names to a single canonical key, so that direct equality between predicted and reference Great Group names ignores edition-driven renaming. Names that have no known mapping pass through unchanged.

Usage

```
canonicalise_kst13ed_gg(gg)
```

Arguments

`gg`	Character vector of Great Group names (lower case, no whitespace).

Details

Examples of the canonicalisation (each pair is rendered equivalent):

- `"haplaquolls"` (KST 8) === `"endoaquolls"` (KST 13ed)
- `"pellusterts"` (KST 8) === `"hapluderts"` (KST 13ed)
- `"camborthids"` (KST 8) === `"haplocambids"` (KST 13ed)
- `"vitrandepts"` (KST 8) === `"vitrudands"` (KST 13ed)

Value

Character vector of canonical keys. Unmapped names pass through. NA stays NA. Empty input returns empty vector.

References

Soil Survey Staff (2022), Keys to Soil Taxonomy 13ed, Ch 4 (Order keys); previous editions for the obsolete names.

canonical_reference *Load a canonical reference dataset from the SoilTaxonomy package*

Description

Reads one of the parsed WRB 2022 / Keys to Soil Taxonomy datasets from the installed **SoilTaxonomy** package (NCSS-tech, GPL-3). soilKey ships no copy of these data: install SoilTaxonomy from CRAN to use this function.

Usage

```
canonical_reference(
  name = c("WRB_4th_2022", "ST_criteria_13th", "ST_features"),
  prefer_pkg = TRUE
)
```

Arguments

name	One of "WRB_4th_2022", "ST_criteria_13th", "ST_features".
prefer_pkg	Retained for compatibility and no longer used. Until v0.9.203 FALSE selected a copy bundled with soilKey; there is no bundled copy any more, so the data always come from SoilTaxonomy.

Value

The dataset as the original R object (list or data.frame).

See Also

wrb2022_canonical, kst13_canonical, st_features_canonical.

cerosidade *Cerosidade quantitativa (SiBCS Cap 13, p 207; Cap 1)*

Description

Diagnostico parametrizado quantidade x intensidade de cerosidade (clay films / cutans). Consume as colunas v0.7.2 clay_films_amount (ordinal: few/pouca, common/comum, many/abundante, continuous/continua) e clay_films_strength (ordinal: weak/fraca, moderate/moderada, strong/forte; "shiny" mapeado a "strong"), introduzidas em substituicao ao legado clay_films.

Usage

```
cerosidade(pedon, min_amount = "common", min_strength = "moderate")
```

Arguments

pedon	A PedonRecord.
min_amount	Quantidade minima: "few", "common", "many", "continuous" (ou equivalentes em PT-BR). Default "common".
min_strength	Intensidade minima: "weak", "moderate", "strong". Default "moderate". Pass NULL para ignorar a dimensao de intensidade.

Details

Discriminante critico Nitossolos vs Argissolos no Cap 13: Nitossolos exigem cerosidade $\geq$ comum + $\geq$ moderada (defaults).

Value

DiagnosticResult; passed = TRUE se ao menos um horizonte B atende ambos os limiares.

References

Embrapa (2018), SiBCS 5a ed., Cap 13 (Nitossolos), p 207; Cap 1 (atributos diagnosticos).

---

chernic *Chernic horizon (WRB 2022 3.1.6)*

Description

A thick, well-structured, very dark surface horizon with a high base saturation (WRB 2022 3.1.6, pp. 40-41). Every layer of the horizon, from the mineral soil surface down, must have: (1) \>= 50% fine earth and no mulmic material; (2) granular structure, subangular blocky structure of average size \<= 2 cm (very fine, fine or medium), or cloddy / agricultural structure; (3) \>= 1% SOC; (4) moist value \<= 3, dry value \<= 5 and moist chroma \<= 2, or the carbonate-rich (15-40% CaCO3, \>= 1.5% SOC) or the \>= 40% CaCO3 / loamy-sand (value \<= 5, \>= 2.5% SOC) alternatives; (5) \>= 1% more SOC than a parent-material layer of moist value \<= 4, where one is recorded; (6) BS \>= 50%. The layers together must be (7) \>= 30 cm thick. Worm holes belong to the Vermic qualifier, not to this horizon.

Usage

```
chernic(pedon, min_thickness = 30)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness in cm (default 30).

Details

A missing dry value is taken as the moist value plus one, as in the mollic colour test. A missing BS is read from pH (Annex 2, Table 9.2).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

chernozem *Chernozem RSG diagnostic (WRB 2022)*

Description

Tests whether a profile satisfies the Chernozem RSG criteria: a mollic horizon plus secondary carbonates somewhere in the profile, plus chroma (moist) <= 2 in at least one layer of the upper 20 cm.

Usage

```
chernozem(pedon, max_chroma_upper = 2)
```

Arguments

pedon	A PedonRecord.
max_chroma_upper	Maximum moist chroma in the upper part (default 2, per WRB 2022).

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 5, Chernozems.

---

chernozem_strict *Chernozem RSG gate (strengthened, WRB 2022 Ch 4, p 111)*

Description

WRB-canonical: chernic horizon AND, starting \<= 50 cm below the lower limit of the mollic horizon and (if a petrocalcic horizon is present) above it, a layer with protocalcic properties \>= 5 cm thick OR a calcic horizon AND base saturation \>= 50% from the mineral soil surface to the protocalcic / calcic layer throughout. Protocalcic properties are visible secondary carbonates (protocalcic_properties); where BS is not recorded it is read from pH (WRB 2022 Annex 2, Table 9.2).

Usage

```
chernozem_strict(pedon, min_bs = 50, max_top_cm = 50, strict = NULL)
```

Arguments

pedon	A PedonRecord.
min_bs	Minimum base saturation % above the carbonate layer (default 50).
max_top_cm	Maximum depth of the carbonate layer below the lower limit of the mollic horizon, cm (default 50).
strict	Logical or NULL. When NULL (default) it resolves via getOption("soilKey.rsg_strict", FALSE). TRUE raises the base-saturation floor to 80%.

Details

v0.3.4 strengthens the previous v0.2 chernozem (which only required mollic + chernic_color) by adding the protocalcic / calcic gate and the BS \>= 50% requirement.

Note: the v0.2 chernozem() diagnostic remains available as a less-strict variant; chernozem_strict() is what the v0.3.4 key.yaml uses for the CH RSG.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

Tier-3 strict mode (v0.9.98)

With strict = TRUE the base-saturation floor above the carbonate-bearing layer is raised from 50% to 80%, in line with the very high base status expected of a textbook Chernozem.

---

claric_material *Claric material (WRB 2022 Ch 3.3.4): light-coloured fine earth with Munsell criteria.*

Description

Claric material (WRB 2022 Ch 3.3.4): light-coloured fine earth with Munsell criteria.

Usage

```
claric_material(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`ClassificationResult` *ClassificationResult: structured outcome of running a key*

Description

ClassificationResult: structured outcome of running a key

ClassificationResult: structured outcome of running a key

Details

Returned by `classify_wrb2022` (and the future `classify_sibcs`). Carries the full decision trace — which RSGs were tested, which passed, which failed, which were indeterminate because of missing data — plus the assigned class, qualifiers, ambiguities (RSGs that nearly satisfied), missing data that would refine the result, the provenance-aware evidence grade, and any biogeographical or prior-based warnings.

Public fields

`system` Character. `"WRB 2022"` or `"SiBCS 5"`.

`name` Character. Full classification name with qualifiers (e.g. `"Rhodic Ferralsol (Clayic, Dystric, Humic)"`).

`rsg_or_order` Character. Bare RSG (WRB) or order (SiBCS), e.g. `"Ferralsols"`.

`qualifiers` List. Principal and supplementary qualifiers in canonical order.

`trace` List. One element per RSG tested (in key order), each with `code`, `name`, `passed`, `evidence`, `missing`.

`ambiguities` List. RSGs that came close to passing — useful hints for follow-up measurements.

`missing_data` Character vector. Attributes whose measurement would refine or resolve the result.

`evidence_grade` Character. `"A"` (measured), `"B"` (spectra-predicted), `"C"` (prior-inferred), `"D"` (VLM-extracted), `"E"` (user-assumed), or `NA_character_` when no horizon carries a soil property (only depths and designations; since v0.9.213), so nothing in the class was verified.

`prior_check` List or NULL. Result of the spatial-prior sanity check (consistent / inconsistent / not run).

`warnings` Character vector. Free-form warnings.

`system_id` Character. Registry id of the system (see `soilkey_systems`), e.g. `"wrb2022"`; since v0.9.265.

`levels` `data.frame` with one row per level of the key that assigned a class (`level`, `code`, `name`), top level first; set by the generic classifier of ordered keys and `NULL` for WRB, SiBCS and USDA, whose levels are in `trace`.

`provisional_because` Character vector. In a first-match key a class that could not be decided for lack of data, tested before the one assigned, makes the assignment provisional; these are those classes, as `"<level>:<code>"`. Empty when the assignment is firm or the classifier does not report it.

`phases` Soil phases of the systems that have them (the SCSJ 2017: temperature, man-made and forest phases; the CSCJ 2011: temperature and man-made phases), a `data.frame` with `kind`, `code`, `name`, `name_native` and `basis`; `NULL` otherwise. Since v0.9.324. The FAO-Unesco Revised Legend (since v0.9.386) gives every phase three-valued, with `status`, `present` and `criterion`, and the classes of the map symbol (`classify_fao`).

`candidates` For the systems that attach a soil to several classes rather than key it to one (the French Referentiel pedologique 2008, since v0.9.346, `classify_rp2008`), every class that may attach it: a `data.frame` with `code`, `name`, `group`, `degree` (simple, simple_or_imperfect, imperfect, undecided; fail for the nearest classes when nothing attaches), `rank` (the order of precedence), `in_name`, `double_with`, `double_possible`, `n_failed`, `distance` (the size of the failures in tolerance units), `failed`, `imperfect_by`, `undecided_by`, `excluded_by` (the criteria, joined with "; ") and `nearest`. `NULL` for every other system, and then not printed. Since v0.9.340.

`humus_form` The humus form of the French Referentiel pedologique 2008 (Annex 1; `humus_form_rp`), a one-row `data.frame` with `form`, `name`, `type`, `label`, `group`, `decided`, `route`, `candidates`, `recorded`, `conflict` and `qualifiers`; the whole reading is in `trace$humus_form`. `NULL` for every other system. Since v0.9.353.

Methods

Public methods:

- `ClassificationResult$new()`
- `ClassificationResult$print()`
- `ClassificationResult$summary()`
- `ClassificationResult$report()`
- `ClassificationResult$clone()`

Method `new()`**:** Build a ClassificationResult.

Usage:

```
ClassificationResult$new(
  system,
  name,
  rsg_or_order = NA_character_,
  qualifiers = list(),
  trace = list(),
  ambiguities = list(),
  missing_data = character(0),
  evidence_grade = NA_character_,
  prior_check = NULL,
  warnings = character(0),
  system_id = NULL,
  levels = NULL,
  provisional_because = character(0),
  candidates = NULL
)
```

Arguments:

`system` System name.

`name` Classification name.

`rsg_or_order` RSG (WRB) or order (SiBCS).

`qualifiers` List of qualifier names.

`trace` List of per-RSG test entries.

`ambiguities` List of close-call RSGs.

`missing_data` Character vector.

`evidence_grade` Single character A/B/C/D or NA.

`prior_check` List or NULL.

`warnings` Character vector.

`system_id` Registry id of the system, or NULL.

`levels` Assigned class per level (data.frame), or NULL.

`provisional_because` Character vector of undecided classes tested before the assigned ones.

`candidates` The classes that may attach the soil (a data.frame), or NULL; since v0.9.340.

Method `print()`**:** Pretty-print the result with key trace, ambiguities, and warnings.

Usage:

```
ClassificationResult$print(...)
```

Arguments:

`...` Ignored (S3 print signature compatibility).

Method `summary()`**:** Compact summary list.

Usage:

```
ClassificationResult$summary(...)
```

Arguments:

`...` Ignored (S3 summary signature compatibility).

Method `report()`**:** Render this classification as a self-contained report (delegates to the package-level `report` generic). HTML output is dependency-free; PDF requires `rmarkdown` and a working LaTeX engine.

Usage:

```
ClassificationResult$report(
  file,
  format = c("auto", "html", "pdf"),
  pedon = NULL,
  ...
)
```

Arguments:

`file` Output path. Format is inferred from the extension.

`format` One of "html" or "pdf" (defaults to "auto", which infers from the extension).

`pedon` Optional `PedonRecord` whose horizons / provenance are added to the report.

`...` Forwarded to `report_html` or `report_pdf`.

Method `clone()`**:** The objects of this class are cloneable with this method.

Usage:

```
ClassificationResult$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

`classification_robustness` *Robustness of classification under input perturbation*

Description

For a given `PedonRecord`, perturb a chosen list of horizon attributes by a configured fractional amount, re-classify under the requested system, and report how often the classification `$rsg_or_order` (or full `$name`) matches the unperturbed baseline.

Usage

```
classification_robustness(
  pedon,
  system = c("wrb2022", "sibcs", "usda"),
  level = c("order", "name"),
  n = 50L,
  perturbations = NULL,
  provenance_aware = FALSE,
  seed = 42L
)
```

Arguments

`pedon`	A `PedonRecord`.
`system`	One of `"wrb2022"`, `"sibcs"`, `"usda"`.
`level`	Either `"order"` (compare `$rsg_or_order`) or `"name"` (compare full classification name).
`n`	Number of Monte-Carlo perturbed runs (default 50).
`perturbations`	Named list. Each name is a horizon column; each element is a function taking the original value and returning a perturbed value. NA-tolerant. Ignored when `provenance_aware = TRUE`.
`provenance_aware`	If `FALSE` (default) every cell is perturbed by the fixed `perturbations` panel – the exact v0.9.42 behaviour. If `TRUE`, each `(horizon, attribute)` cell is perturbed by an amount scaled to its provenance evidence grade, and `perturbations` is ignored. See `classify_with_uncertainty` for the full provenance-weighted posterior.
`seed`	Random seed for reproducibility.

Details

Default perturbation panel:

- clay_pct: ±5
- sand_pct: ±5
- silt_pct: ±5
- ph_h2o: ±0.2 absolute
- oc_pct: ±10

Value

A list with elements baseline (the unperturbed classification name), n (number of MC runs), robustness (fraction of perturbed runs matching baseline), flipped_to (table of alternative classifications when the perturbation flipped the result).

Examples

```
## Not run:
p <- make_ferralsol_canonical()
classification_robustness(p, system = "wrb2022", n = 50)
#> $baseline    : "Ferralsols"
#> $robustness  : 0.96  (48 / 50 perturbed runs landed on Ferralsols)
#> $flipped_to  : table(c("Cambisols" = 1, "Acrisols" = 1))

## End(Not run)
```

---

classify_all *Classify a pedon across all three taxonomic systems*

Description

Convenience wrapper that runs classify_wrb2022, classify_sibcs, and classify_usda on the same PedonRecord and returns a single named list with one entry per system (plus a summary table that's handy for reports).

Usage

```
classify_all(
  pedon,
  systems = "all",
  on_missing = c("warn", "silent", "error"),
  include_familia = TRUE,
  include_family = FALSE,
  specifiers = FALSE,
  gapfill = FALSE,
  ...
)
```

Arguments

pedon	A PedonRecord.
systems	Character vector. Any subset of the registered systems ("wrb2022", "sibcs", "usda"; ids and aliases as listed by soilkey_systems), or the literal "all" (default) to run every default system.
on_missing	One of "warn" (default), "silent", "error". Forwarded verbatim to each classifier.
include_familia	Forwarded to classify_sibcs (default TRUE). Has no effect on the other systems.
include_family	Forwarded to classify_usda (default FALSE) to derive the USDA 5th-level family, and to classify_asc3 when it is run (the Australian family, since 0.9.292). No effect on the other systems.
specifiers	Forwarded to classify_wrb2022 (default FALSE) to auto-attach WRB depth specifiers. No effect on the other systems.
gapfill	Forwarded to all three classifiers (default FALSE => byte-identical). Opt-in within-pedon depth gap-fill; see gapfill_within_pedon. Applied independently per system on a deep copy, so the caller's pedon is never mutated.
...	Additional named arguments are silently ignored.

Details

Each classifier still produces its own ClassificationResult with the full key trace and evidence grade – nothing is collapsed or homogenised. The wrapper exists for ergonomics, not abstraction.

Value

A named list with elements:

- wrb – ClassificationResult from classify_wrb2022() (or NULL if the system was skipped or errored).
- sibcs – as above, from classify_sibcs().
- usda – as above, from classify_usda().
- summary – a 1-row data.frame with one column per system, holding the resulting $name (or NA when the system was skipped / errored). Useful for tabulating many pedons in one shot.

Selecting a subset of systems

Pass systems = c("wrb2022", "sibcs") (or any other subset) to skip systems you don't need. Default systems = "all" runs all three. The systems and their aliases come from soilkey_systems.

Errors and partial results

If a single classifier raises an error, the corresponding slot of the returned list is set to NULL and a one-line warning is emitted (so you can rerun the offender on its own to see the full traceback). The other classifiers still run and their results are returned. This matches the spirit of on_missing = "warn" on the individual classifiers.

Side effects

None. The classifiers do not mutate pedon; the wrapper does not attach any side-channel state.

See Also

classify_wrb2022, classify_sibcs, classify_usda.

Examples

```
pr <- make_ferralsol_canonical()
all_three <- classify_all(pr)
all_three$summary

# WRB + USDA only (skip SiBCS):
classify_all(pr, systems = c("wrb2022", "usda"))$summary
```

---

classify_asc3 *Classify a pedon under the Australian Soil Classification*

Description

Keys a PedonRecord through the Australian Soil Classification, 3rd edition (Isbell and the National Committee on Soil and Terrain 2021), following the online version with its errata. **Experimental**: this release keys the 15 orders (the key to soil orders) and the suborders of all of them: the Kurosols, Sodosols, Chromosols, Ferrosols, Dermosols, Kandosols and Vertosols (colour classes, and the Aquic Vertosols), the Organosols (the dominant kind of peat), the Podosols (saturation of the B, sulfuric and hypersulfidic material), the Hydrosols (the setting, then mottling), the Calcarosols (gypsum and the carbonate classes), the Anthroposols (the kind of human-made material), the Arenosols and Tenosols (chernic horizons, bleaching, hard materials within 0.5 m, carbonate, nodules, then colour) and the Rudosols (salinity, gypsum, shells, carbic materials, hard materials, sand, gravel, layers), and the great groups of the Kurosols, Sodosols, Chromosols, Ferrosols, Dermosols and Kandosols (pans in or under the B, structure, plasticity, base status, exchangeable Ca/Mg, ESP and carbonate of the B2 or B2t), the Vertosols (the surface), Calcarosols (what the soil directly overlies, and the B), Podosols (the kinds of Podosol B), Organosols (drainage, acid sulfate materials, carbonate and acidity), Hydrosols (acid sulfate materials, the upper 0.5 m, and the order the soil would key to apart from wetness), Arenosols, Rudosols and Tenosols (what the soil directly overlies, nodular, gypsic and argic horizons, carbonate, texture, gravel); the Anthroposols have none in the text; and the subgroups of the Kurosols, Sodosols and Chromosols (chernic and vertic features, bleaching, ferric and manganic horizons, sodicity of the lower B, mottling of the B2t, base status, carbonate) and of the Dermosols, Kandosols and Ferrosols (the same features read on the B2, a strongly acid B2, argic and bauxitic horizons in the B, a snuffy A), of the Vertosols and Calcarosols (acid sulfate materials, pans, sodicity, acidity, carbonate and mottling of the upper 0.5 m of the solum and below it, subplasticity, surface carbonate), the Podosols (coherence of the B, pans), the Organosols (what the organic materials overlie), and the Hydrosols, Arenosols, Rudosols and Tenosols (peat kinds, sulfidic materials near the surface, acidity and carbonate of the upper 1.0 m, the B, the A2, the solum or the soil, horizons with all but the thickness; the Rudosolic and Tenosolic

great groups of the Redoxic and Oxyaquic Hydrosols take the list of the Rudosol or Tenosol suborder the soil keys to apart from wetness). The name is that of the deepest class assigned, in the book's form "Subgroup, Great Group, Suborder Order" ("Bleached, Eutrophic, Red Chromosol"); `rsg_or_order` keeps the order. With `include_family = TRUE` the family follows the name as the book writes it ("Bleached, Eutrophic, Red Chromosol; thick, non-gravelly, loamy/clayey, deep"): see `classify_asc3_family`.

Usage

```
classify_asc3(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE,
  include_family = FALSE
)
```

Arguments

pedon	A `PedonRecord`.
rules	Optional pre-loaded rule set (`load_rules("asc3")`).
on_missing	One of `"warn"` (default), `"silent"`, `"error"`.
gapfill	Opt-in within-pedon gap-fill, as in `classify_usda`.
include_family	Derive the family (`classify_asc3_family`) and write it after the name. Default `FALSE` (the result is as in earlier versions).

Details

The ASC is largely morphological. The diagnostics read field texture grades (`field_texture`, NCST Field Handbook codes; lab clay stands in and is undecided near the limits), the Australian horizon notation (A1, A2, B21t, Bhs, Bk ...; FAO/USDA designations are read by the role of each horizon), the boundary between the A and the B, pH (strongly acid: pHw below 5.5), ESP (sodic: 6 or more), subplasticity, free iron, grade of pedality, carbonates and the moisture of the horizons when described (`soil_water_status`: the Vertosols' cracks are not required in a solum described moist), mottles, soft and hard carbonate, gypsum, the kind of peat, acid sulfate tests, EC of a 1:5 extract, dry and moist strength, ped size, nodules, pans and exchangeable bases. From the site it reads `drainage_class` (Hydrosols, organic materials, Podosol suborders), `tidal_inundation` (Hydrosols, organic materials), `surface_condition` (a cracking surface, Vertosols; a saline one, Salic Hydrosols), `landform`, `landform_element`, `tidal_zone`, `subaqueous` and `salinised` (Hydrosol suborders) and `designation_system` (`"asc"` for the Australian notation). The choices where the text is silent are listed in `system.file("rules", "asc3", "readings.yaml", package = "soilKey")`.

Value

A `ClassificationResult` with `system_id` `"asc3"`, the deepest class in `name`, the order in `rsg_or_order`, the classes of each level in `levels`, the key trace, and `provisional_because` (classes before the assigned ones that could not be decided); with `include_family`, the family criteria in `trace$family`, their names in `trace$family_label` and their codes in `trace$family_code`.

References

Isbell, R.F. and the National Committee on Soil and Terrain (2021). The Australian Soil Classification, 3rd ed. CSIRO Publishing, Melbourne. Online version: www.soilscienceaustralia.org.au/asc.

See Also

make_asc3_example, classify_asc3_family, soilkey_systems.

Examples

```
res <- classify_asc3(make_asc3_example("chromosols"), on_missing = "silent")
res$name
classify_asc3(make_asc3_example("chromosols"), on_missing = "silent",
              include_family = TRUE)$name
```

---

classify_asc3_family *Family of a soil under the Australian Soil Classification*

Description

Places a soil in the family classes of its order (Australian Soil Classification, 3rd edition, the 'Family Criteria' of each order page). The family is not a level of the key: each criterion is a class of its own, given by a one- or two-letter code. Most orders use the thickness of the A1 horizon (of the whole A horizon in the Kurosols, Sodosols and Chromosols and in the Kurosolic, Sodosolic and Chromosolic Hydrosols, as the book's general rule has it), with any organic horizon over it; gravel (2-60 mm) of the surface and A1 horizon (the larger of the estimate of the ground surface, `site$surface_gravel_pct` or `site$surface_coarse_fragments_pct`, and that of the A1); the field texture of the A1 (its major part, peaty under organic surface horizons) and the most clayey field texture of the B; soil depth; and water repellence of the surface soil (`water_repellence`). The Vertosols use laboratory clay (upper 0.1 m and the B), the Calcarosols the depth of the first Bk horizon, the Arenosols and Rudosols the surface soil, and the Organosols the kind and thickness of their organic materials and the texture under them. Soil depth is the depth to rock, a cemented pan (not a thin ironpan), another cemented layer or saprolite; otherwise, in a soil with a B horizon, the depth to the C below it, and in a soil without one, the depth described. The Anthroposols have no family, nor do the Subtidal, Subaqueous and Intertidal Hydrosols. The readings where the text is silent are in `system.file("rules", "asc3", "readings.yaml",package = "soilKey")` (topic "Families").

Usage

```
classify_asc3_family(pedon, order = NULL, suborder = NULL, great_group = NULL)
```

Arguments

pedon	A PedonRecord.
order, suborder, great_group	Codes of the soil's classes (as in `levels` of a classify_asc3 result). When `order` is `NULL` the soil is classified first.

Value

A named list of FamilyAttribute, one per criterion of the order that applies, in the order of the book's family name: value is the class name (NULL when undecided or when no class fits), evidence$code the class code ("YY" undecided, "ZZ" no class of the order fits). An empty list when the soil has no family.

References

Isbell, R.F. and the National Committee on Soil and Terrain (2021). The Australian Soil Classification, 3rd ed. CSIRO Publishing, Melbourne. Online version: www.soilscienceaustralia.org.au/asc.

See Also

classify_asc3.

Examples

```
fam <- classify_asc3_family(make_asc3_example("chromosols"))
vapply(fam, function(a) a$evidence$code, "")
```

---

classify_by_spectral_neighbours

Classify a soil by spectral similarity to OSSL reference profiles

---

Description

Given a Vis-NIR (or MIR) spectrum and an OSSL reference library enriched with WRB / SiBCS / USDA labels, returns the K most spectrally similar profiles plus a probabilistic class prediction aggregated from their labels.

Usage

```
classify_by_spectral_neighbours(
  spectrum,
  ossl_library,
  system = c("wrb2022", "sibcs", "usda"),
  k = 25L,
  preprocess = "snv+sg1",
  region = NULL,
  verbose = TRUE
)
```

Arguments

spectrum	Numeric vector or 1-row matrix (the query spectrum). Must align (after preprocessing) with the column space of ossl_library$Xr.
ossl_library	A list with Xr (numeric matrix, rows = OSSL training profiles, cols = wavelengths) and Yr (data frame keyed by property; *must include* a column named wrb_rsg and / or sibcs_ordem / usda_order for the labels to aggregate over). ossl_library may also carry lat and lon columns in Yr for the regional filter.
system	One of "wrb2022" (default), "sibcs", "usda". Controls which label column of Yr is aggregated.
k	Number of nearest neighbours (default 25).
preprocess	Pre-processing pipeline; passed to preprocess_spectra. Default "snv+sg1".
region	Optional list(lat, lon, radius_km) for a regional filter on ossl_libraryYrlat / lon.
verbose	Emit a cli summary.

Details

This is the **spectral analogy** classifier. It does not replace the deterministic key in classify_wrb2022 / classify_sibcs / classify_usda; instead it provides a high-prior "expected class" before the user has lab data, reducing the search space when collecting confirming attributes.

Value

A list with three elements:

distribution A data.table with columns class, n_neighbours, probability (= n_neighbours / k), sorted by probability.

neighbours A data.table with one row per neighbour (top K), columns rank, distance, class, plus any other columns present in ossl_library$Yr.

query The query metadata (system, k, region filter, n_library_rows, n_filtered).

Distance metric

By default we compute distances on PLS scores (matching the resemble / OSSL recipe), with PLS components fit on the OSSL reference Yr matrix. When resemble is unavailable, we fall back to PCA scores from stats::prcomp on the preprocessed Xr – a defensible-but-simpler heuristic.

Region filter

Optional lat / lon / radius_km arguments filter the OSSL library to profiles within radius_km (great-circle) of the query location before computing distances. This implements the "biome-aware" use case the architecture document calls for: a Cerrado profile shouldn't have its class inferred from spectral neighbours in the Boreal taiga.

See Also

predict_ossl_mbl (predicts attributes), classify_wrb2022 (the deterministic key).

Examples

```
## Not run:
# Toy run against the bundled demo library (synthetic):
data(ossl_demo_sa)
# Inject a fake label column for the demo (real OSSL has it):
ossl_demo_sa$Yr$wrb_rsg <- sample(c("FR", "AC", "LX", "AL"),
                                  nrow(ossl_demo_sa$Yr),
                                  replace = TRUE)
query <- ossl_demo_sa$Xr[1, ]
res <- classify_by_spectral_neighbours(query, ossl_demo_sa,
                                       k = 10)
res$distribution    # ranked classes
res$neighbours      # the 10 most similar profiles

## End(Not run)
```

`classify_cpcs1967` *Classify a pedon under the French soil classification (CPCS 1967)*

Description

Keys a `PedonRecord` through the twelve classes of the French soil classification of 1967 (Commission de Pedologie et de Cartographie des Sols). **Experimental**: this release keys the classes (sols mineraux bruts, sols peu evolues, vertisols, andosols, sols calcimagnesiques, sols isohumiques, sols brunifies, sols podzolises, sols a sesquioxydes de fer et manganese, sols ferrallitiques, sols hydromorphes, sols sodiques), their 35 sub-classes, the climatic ones read from `site$koppen_climate`, else the USDA soil temperature and moisture regimes, else the latitude (with no climate, the non-climatic or temperate sub-class), their 101 groups and their 276 subgroups (a group or subgroup the data cannot decide leaves the name provisional; when no group of a sub-class, or no subgroup of a group, fits, the name stops at the level above, and none is invented). The names are French, as the text writes them, in `name` and in `levels$name_native`; a subgroup named by an adjective is shown after its group ("Rankers, a mor"), one the text names in full stands alone, and the other subgroups whose tests pass are noted in the trace as intergrades.

Usage

```
classify_cpcs1967(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	Optional pre-loaded rule set (`load_rules("cpcs1967")`).
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`.
`gapfill`	Opt-in within-pedon gap-fill, as in `classify_usda`.

Details

The CPCS gives no key. soilKey keys the classes in an order built from the text's own statements of precedence: raw mineral soils, vertisols, sodic soils, andosols, podzolised, ferrallitic and sesquioxide soils, hydromorphic soils, weakly developed soils, isohumic soils, calcimagnesic soils, and the brunified soils last; a peat is a hydromorphic soil, whatever lies under it. Every threshold of the text is in depths from the top of the first mineral horizon.

The key reads the old French horizon notation (Aoo, Ao, A1, A2, A3, (B), B, Bh, Bfe, Bt, C, Cca, G, g, R, a Roman numeral for a change of parent material), declared with `site$designation_system = "cpcs"` or recognised by its symbols, beside the FAO letters; where a profile has no designations, soilKey's diagnostic horizons stand in. From the horizons it reads organic carbon and nitrogen, texture, pH, the CEC and the exchangeable bases, carbonate and gypsum, the EC of the saturated paste, Ki (`sio2_sulfuric_pct`, `al2o3_sulfuric_pct`), the andic properties, moist and dry Munsell colours, mottles, slickensides, cracks, structure and consistence, clay films, coarse fragments, the sand minerals, the fibre content of peat, ice (`ice_pct`, `ice_form`), and the recorded origin of the layers; from the site the climate, the soil temperature regime, the parent material, recorded erosion, the water table, the elevation, patterned ground (`patterned_ground`) and the humus form (`humus_form`); for some subgroups also active carbonate (`caco3_active_pct`), the coarse and fine sand (`coarse_sand_pct`, `fine_sand_pct`) and gibbsite (`gibbsite_pct`). The choices where the text is silent are listed in `system.file("rules", "cpcs1967", "readings.yaml", package = "soilKey")`.

Value

A `ClassificationResult` with `system_id` `"cpcs1967"`, the deepest class assigned in `name`, the class in `rsg_or_order`, the classes of each level in `levels` (with `name_native`), the key trace, and `provisional_because` (classes before the assigned one that could not be decided).

References

Commission de Pedologie et de Cartographie des Sols (1967). Classification des sols. ENSA, Grignon. Digitised by the AFES (2011).

See Also

`make_cpcs1967_example`, `soilkey_systems`.

Examples

```
res <- classify_cpcs1967(make_cpcs1967_example("ferrallitic"), on_missing = "silent")
res$name
res$levels
```

`classify_cscj2011` *Classify a pedon under the Comprehensive Soil Classification System of Japan*

Description

Keys a `PedonRecord` through the Comprehensive Soil Classification System of Japan, First Approximation (Obara et al. 2011), following its key to great groups. **Experimental**: this release keys the ten great groups (Man-made soils, Organic soils, Podzols, Andosols, Dark Red soils, Lowland soils, Red-Yellow soils, Stagnic soils, Brown Forest soils, Regosols), their 27 groups (for instance Fulvic and Allophanic Andosols, Calcaric Dark Red soils, Brown Lowland soils) and their 116 subgroups (Haplic, Argic, Low-humic, Terrace ...) and their 381 soil series groups (lithic, skeletal, fine-textured, high-humic ..., read with the new attributes `texture_class_jp` and `hardness_yamanaka_mm`), and gives the soil phases of chapter VIII (temperature, estimated from the air temperature plus 2 C when the soil temperature is not recorded, and the man-made phases). Each level gives the English name of the text's appendix 2 and, in `levels$name_native`, the Japanese one.

Usage

```
classify_cscj2011(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	Optional pre-loaded rule set (`load_rules("cscj2011")`).
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`.
`gapfill`	Opt-in within-pedon gap-fill, as in `classify_usda`.

Details

The CSCJ shares most of its diagnostics with the Soil Classification System of Japan (2017), and soilKey reads them with the same code (see `classify_scsj2017` for the attributes read), by the CSCJ's own limits where it has them: the argic horizon (the WRB 2006 clay increase), the immature kuroboku feature (P retention as its only substitute), and the dark red colour of the Dark Red soils, which also read the origin from limestone or calcareous deposits (`rock_origin`, `layer_origin`, the site `parent_material`, `caco3_pct`). The choices where the text is silent are listed in `system.file("rules", "cscj2011", "readings.yaml", package = "soilKey")`.

Value

A ClassificationResult with system_id "cscj2011", the deepest class in name, the great group in rsg_or_order, the classes of each level in levels (with name_native), the key trace, provisional_because (classes before the assigned one that could not be decided), and in phases the soil phases of chapter VIII (temperature and man-made phases; the class name is left as it is).

References

Obara, H., Ohkura, T., Takata, Y., Kohyama, K., Maejima, Y. and Hamazaki, T. (2011). Comprehensive Soil Classification System of Japan, First Approximation. Bulletin of the National Institute for Agro-Environmental Sciences 29: 1-73.

See Also

make_cscj2011_example, classify_scsj2017, soilkey_systems.

Examples

```
res <- classify_cscj2011(make_cscj2011_example("dark_red"), on_missing = "silent")
res$name
res$levels
```

---

classify_cssc3 *Classify a pedon under the Canadian System of Soil Classification*

Description

Keys a PedonRecord through the Canadian System of Soil Classification, 3rd edition (Soil Classification Working Group 1998). **Experimental**: this release keys the ten orders (Key to Soil Orders, Ch 3, pp. 33-34), their 31 great groups (Key to Soil Great Groups, pp. 35-36) and their 231 subgroups (Key to Soil Subgroups, pp. 37-51). The name is the deepest class assigned: the subgroup (e.g. "Terric Mesisol"), or the great group (e.g. "Solonetz") when no subgroup can be given: the Solonetzic subgroups are named by the colour of the A, and the key has no other subgroup for a Solonetzic soil without an A (one eroded down to its solonetzic B); the order is in rsg_or_order and every level in levels.

Usage

```
classify_cssc3(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set (load_rules("cssc3")).
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Details

The Canadian keys are written largely in terms of horizon designations, which the book defines partly by measurable properties. Each diagnostic reads the measured properties first (podzolic B chemistry with pyrophosphate-extractable Fe and Al, the Bt clay increase, the solonetzic B Ca/Na ratio, the chernozemic A, the gley colours) and the surveyor's designations in the Canadian notation second (Bt, Bf, Bm, Btj, Bn, Bv, Bss, Bg, Cz...). When the assigned order rests on designations alone, a warning says so. A profile that declares the FAO notation in site$designation_system ("fao2006", as isis_profile_to_pedon sets it, "fao", "wrb" or "wrb2022") is read through a translation of its designations into the Canadian notation where the letters mean the same (the book's Table 6: E = Ae, Bw = Bm, Bs = Bf, Btn = Bnt, f = z, z = s ...); the letters with no Canadian equivalent are not read, and the evidence lists what each designation was read as. The choices where the book is silent are listed in system.file("rules", "cssc3", "readings.yaml", package = "soilKey").

The chernozemic A horizon includes a soil-climate criterion (mean annual soil temperature at least 0 C, moisture subclass drier than humid), read from site$mean_annual_soil_temp_C and site$soil_moisture_subclass, or approximated from the USDA regimes. Without it the Chernozemic order is undecided and any order keyed after it is reported as provisional ($provisional_because). The Gray and Dark Gray subgroups of the Solods and Solodized Solonetz also ask for a subhumid climate (site$soil_moisture_subclass).

Value

A ClassificationResult with system_id "cssc3", the deepest class assigned in name, the order in rsg_or_order, every level in levels, the key trace, and provisional_because.

References

Soil Classification Working Group (1998). The Canadian System of Soil Classification, 3rd ed. Agriculture and Agri-Food Canada Publication 1646. NRC Research Press, Ottawa.

See Also

make_cssc3_example, soilkey_systems.

Examples

```
res <- classify_cssc3(make_cssc3_example("podzolic"), on_missing = "silent")
res$name
```

`classify_cst2001` *Classify a pedon under the Chinese Soil Taxonomy (2001)*

Description

Keys a `PedonRecord` through the Chinese Soil Taxonomy (Cooperative Research Group on Chinese Soil Taxonomy, 2001, the English edition published by Science Press). **Experimental**: this release keys the 14 orders (key 5.1: Histosols, Anthrosols, Spodosols, Andosols, Ferralosols, Vertosols, Aridosols, Halosols, Gleyosols, Isohumosols, Ferrosols, Argosols, Cambosols and Primosols), their 39 suborders and their 138 groups (key 5.2), with soilKey's residuals `[E2] [Other Ferralosols]`, `[A1.4] [Other Permagelic Histosols]` and `[A2.5] [Other Orthic Histosols]` where the book prints none (the name stops there, with a note). Since v0.9.375 it keys the subgroups of orders A to G (176, with `[A1.3.x]` Lithic Hemi-Permagelic Histosols, listed in Appendix 1 only, at its Appendix position); their criteria are written as presets (`inst/rules/cst2001/presets.yaml`), expanded when the rules are loaded. The groups of orders H to N stop at the group, with a note, until their subgroups follow in a later release.

Usage

```
classify_cst2001(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	Optional pre-loaded rule set (`load_rules("cst2001")`).
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`.
`gapfill`	Opt-in within-pedon gap-fill, as in `classify_usda`.

Details

The key is read in the book's order, the first class whose criteria pass assigned, as the book prescribes (each class lacks the properties of the classes before it). The criteria read the 33 diagnostic horizons and the 25 diagnostic characteristics of chapter 3 on the recorded data (see `identify_horizons` with `"cst2001"`), with their weaker evidence grades where the book gives them, the soil moisture and temperature regimes (recorded on the site, or from the aridity index, the site's records of paddy, groundwater and permafrost, and the air temperature), and the few criteria the key defines where it uses them. Every test is three-valued: a measurement not recorded leaves it undecided, a field observation not recorded is absent. A class tested before the assigned one that could not be decided makes the name provisional (`provisional_because`); a class assigned on a reading the book leaves open (gley colours without a ferrous-iron test, cracks observed for their duration) carries a warning; a class the book itself marks provisional is noted as such (`levels$book_provisional`).

Names are the English of the book, in the plural of the key ("Permagelic Histosols"), with soilKey's Portuguese gloss in `levels$name_pt`. The approximate equivalents of the order in US Soil Taxonomy (1999) and the WRB (1998), Tables 4.2 and 4.3, are added as a note (`qualifiers$correlation_cst` and a line in `warnings`); they never change the classification. A residual suborder or group named for the udic regime (the Udic Andosols, Vertosols, Isohumosols, Ferrosols, Argosols and Cambosols, the Udi- groups) or for the aquic one (the Aqui-Orthic Halosols) reached by a soil whose regime read is another keeps its class, with a note. The diagnostic horizons and characteristics of each layer are in the trace (phase `horizons`). The readings where the text is silent are listed in `system.file("rules", "cst2001","readings.yaml", package = "soilKey")`.

Value

A `ClassificationResult` with `system_id` `"cst2001"`, the subgroup (or the deepest class assigned) in `name`, the order in `rsg_or_order`, the classes of each level in `levels` (with `name_pt` and `book_provisional`), the key trace with the horizon panel, `provisional_because`, and, when a class was assigned on a provisional reading, in `candidates` the class the key would reach without it (`kind` "alternative").

References

Cooperative Research Group on Chinese Soil Taxonomy (2001). Chinese Soil Taxonomy. Science Press, Beijing and New York.

See Also

`make_cst2001_example`, `identify_horizons`, `soilkey_systems`.

Examples

```
res <- classify_cst2001(make_cst2001_example("K3"), on_missing = "silent")
res$name
res$levels
```

---

classify_csv *Classify a horizon spreadsheet in all three systems - one file, one line*

Description

The shortest path from a spreadsheet to an answer. Reads `file` with `read_pedon_csv` and returns the WRB 2022 / SiBCS / USDA names as a one-row `data.frame`. Missing attributes are handled silently (`on_missing = "silent"`).

Usage

```
classify_csv(
  file,
  site = NULL,
  sep = "auto",
  systems = c("wrb", "sibcs", "usda")
)
```

Arguments

file	Path to a .csv (comma) or .tsv (tab) file, one row per horizon.
site	Optional named list of site metadata (see PedonRecord). Defaults to list(id = <file base name>).
sep	Field separator. "auto" (default) uses a tab for .tsv files and a comma otherwise.
systems	Character vector of systems to run; any of "wrb", "sibcs", "usda" (default: all three).

Details

The full ClassificationResult objects (with the key trace, evidence grade, qualifiers, ...) and the parsed PedonRecord are attached as attributes "results" and "pedon" for anyone who wants to dig deeper.

Value

A one-row data.frame with one column per system.

See Also

read_pedon_csv, classify_all.

Examples

```
f <- system.file("extdata", "perfil_exemplo.csv", package = "soilKey")
classify_csv(f)
```

---

classify_fao *Classify a pedon in the FAO-Unesco soil legend*

Description

Keys a PedonRecord through the FAO-Unesco legend of the Soil Map of the World. **Experimental**: this release keys the 28 major soil groupings of the Revised Legend (FAO 1988, World Soil Resources Report 60; the text of the ISRIC reprint of 1997, Technical Paper 20), chapter X, pp. 75-89, and their 153 soil units, in the key's order (Histosols, Anthrosols, Leptosols, Vertisols, Fluvisols, Solonchaks, Gleysols, Andosols, Arenosols, Regosols, Podzols, Plinthosols, Ferralsols, Planosols, Solonetz, Greyzems, Chernozems, Kastanozems, Phaeozems, Podzoluvisols, Gypsisols, Calcisols, Nitisols, Alisols, Acrisols, Luvisols, Lixisols, Cambisols), on the 16 diagnostic horizons and 27 diagnostic properties of chapters IV and V read on the data. Under its grouping a soil keys to the first soil unit whose gate passes, in the key's order; the last unit of each grouping is the key's "Other ..." residual (Haplic, Eutric, and the Terric Histosols and Urbic Anthrosols, whose descriptions are shown, not tested). Since v0.9.386 the result also gives the 16 phases of chapter VIII (pp. 67-71) and the texture and slope classes of the map symbol (Appendix 2, p. 107) in $phases, beside the unit, which they never change. Since v0.9.389, edition

= "1974" keys the Legend of 1974 (FAO-Unesco, Soil Map of the World, Vol. I, Unesco Paris), pp. 43-53: its 26 major soil groups in the key's order (Histosols, Lithosols, Vertisols, Fluvisols, Solonchaks, Gleysols, Andosols, Arenosols, Regosols, Rankers, Rendzinas, Podzols, Ferralsols, Planosols, Solonetz, Greyzems, Chernozems, Kastanozems, Phaeozems, Podzoluvisols, Xerosols, Yermosols, Nitosols, Acrisols, Luvisols, Cambisols) and their 106 soil units (the Lithosols, Rankers and Rendzinas are single units), on the diagnostic horizons and properties of 1974 read on the data; its registry id is fao1974 (classify_fao1974). Since v0.9.390 its result gives in $phases the 12 phases of 1974 (pp. 5-7: stony, lithic, petric, petrocalcic, petrogypsic, petroferric, phreatic, fragipan, duripan, saline, sodic, cerrado) and the texture and slope classes of its map symbol (pp. 4-5), in the same form.

Usage

```
classify_fao(
  pedon,
  edition = c("1988", "1974"),
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
edition	"1988" (the Revised Legend, default) or "1974" (the Legend of 1974).
rules	Optional pre-loaded rule set (load_rules("fao1988"), or load_rules("fao1974") with edition = "1974").
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Details

The tests are three-valued: a measurement not recorded leaves its criterion undecided, a field observation not recorded is absent. A grouping or unit tested before the assigned one that the data leave undecided makes the result provisional (provisional_because, "major_groups:<code>" or "units:<code>": the candidates). The key has no residual grouping after the Cambisols: a soil that no gate takes is "[Not classified in the FAO-Unesco Revised Legend (1988)]", in brackets because the name is soilKey's, with the gates that failed in the warnings; when no gate passes but some are undecided, the name lists them in brackets as the candidates. Depths are read from the soil surface, H and O horizons included. The readings where the text is silent are in system.file("rules", "fao1988", "readings.yaml", package = "soilKey").

Value

A ClassificationResult with system_id "fao1988" ("fao1974" for the Legend of 1974); name holds the soil unit as printed ("Haplic Luvisols") and rsg_or_order the major soil grouping ("Luvisols"); levels has one row per level (major_groups, units) with the code ("LV", "LVh"), name, page of the key and soilKey's Portuguese gloss (name_pt); the key trace of both levels and

provisional_because; and phases, a data.frame with one row per phase and one per class of the map symbol (kind "phase", "texture_class", "slope_class"), its status ("present", "absent", "undecided", or "not shown" for a unit the text excludes), present and criterion (TRUE, FALSE, NA), page, basis, flags and the attributes that would decide an undecided phase (missing).

References

FAO (1988). FAO/Unesco Soil Map of the World, Revised Legend, with corrections and updates. World Soil Resources Report 60, FAO, Rome. Reprinted with updates as Technical Paper 20, ISRIC, Wageningen, 1997.

See Also

classify_fao1988, classify_fao1974, make_fao_example, soilkey_systems.

Examples

```
res <- classify_fao(make_fao_example("LV"), on_missing = "silent")
res$name
res$rsg_or_order
res$levels
res$phases[res$phases$status != "absent", c("kind", "code", "name", "status")]
classify_fao(make_fao_example("Lc", edition = "1974"), edition = "1974",
             on_missing = "silent")$name
```

---

classify_fao1974 *Classify a pedon in the FAO-Unesco Legend of 1974*

Description

The registry's classifier of fao1974 (aliases fao74, fao_unesco_1974): classify_fao(pedon, edition = "1974"). Experimental and opt-in. See classify_fao.

Usage

```
classify_fao1974(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set (load_rules("fao1988"), or load_rules("fao1974") with edition = "1974").
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Value

A ClassificationResult, as classify_fao, with system_id "fao1974".

References

FAO-Unesco (1974). Soil Map of the World 1:5 000 000. Volume I, Legend. Unesco, Paris.

See Also

classify_fao, make_fao_example.

Examples

```
classify_fao1974(make_fao_example("L", edition = "1974"), on_missing = "silent")$name
```

---

classify_fao1988 *Classify a pedon in the FAO-Unesco Revised Legend (1988)*

Description

The registry's classifier of fao1988 (aliases fao, fao88, revised_legend): classify_fao(pedon, edition = "1988"). See classify_fao.

Usage

```
classify_fao1988(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set (load_rules("fao1988"), or load_rules("fao1974") with edition = "1974").
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Value

A ClassificationResult, as classify_fao.

See Also

classify_fao, make_fao_example.

Examples

```
classify_fao1988(make_fao_example("CM"), on_missing = "silent")$name
```

`classify_from_documents` *Build a fully-classified 'PedonRecord' from documents in one call*

Description

Highest-level entry point of the soilKey VLM pipeline. Given a soil-description PDF and / or a profile-wall photograph, this function:

Usage

```
classify_from_documents(
  pdf = NULL,
  image = NULL,
  fieldsheet = NULL,
  pedon = NULL,
  provider = "auto",
  model = NULL,
  systems = c("wrb", "sibcs", "usda"),
  report = NULL,
  overwrite = FALSE,
  verbose = TRUE
)
```

Arguments

pdf	Optional path to a soil-description PDF.
image	Optional path to a profile-wall image (JPG / PNG); if supplied, Munsell extraction is attempted with the configured provider.
fieldsheet	Optional path to a site-metadata field sheet (image or PDF).
pedon	Optional existing PedonRecord; when supplied, the function fills only the fields VLM extraction can fill (subject to the provenance-authority order).
provider	Either a provider name passed to vlm_provider (default "ollama") OR a pre-built ellmer chat object (when you want full control over system_prompt, api_key, ...).

model	Optional model identifier; passed through to vlm_provider() when provider is a string. Defaults to the per-provider default from default_model.
systems	Character vector listing which classification systems to run; subset of c("wrb", "sibcs", "usda"). Default: all three.
report	Optional output path for a self-contained report (.html or .pdf). When supplied, report is called on the classification results + pedon. Default NULL (no report file).
overwrite	When merging extracted values into an existing pedon, allow VLM-extracted attributes to clobber already-recorded ones. Default FALSE – the provenance authority order (measured > extracted_vlm) is enforced by PedonRecord$add_measurement().
verbose	Emit cli progress messages. Default TRUE.

Details

1. Constructs a vision-language provider chat object via vlm_provider (defaults to local Ollama with Gemma 4 edge for institutional independence and data sovereignty).
2. Extracts horizons from pdf via extract_horizons_from_pdf, Munsell colours from image via extract_munsell_from_photo, and site metadata from fieldsheet via extract_site_from_fieldsheet. Every extracted attribute is stamped source = "extracted_vlm" in the PedonRecord's provenance log.
3. Runs the three deterministic keys (classify_wrb2022, classify_sibcs, classify_usda). The VLM never classifies – the package's architectural invariant is preserved.
4. Optionally renders a one-pager HTML / PDF report via report.

At least one of pdf, image or fieldsheet must be supplied; you can also pass an existing partially-filled PedonRecord via pedon and let this function fill the gaps.

Value

A list with elements:

pedon The (mutated) PedonRecord.

classifications Named list with up to three ClassificationResult objects keyed by wrb, sibcs, usda.

report Path to the rendered report file (if report = ... was supplied), else NULL.

provider The chat-provider object actually used (useful for downstream debugging or cost accounting).

Why local-first by default

The default provider = "ollama" runs the entire VLM pipeline on the user's machine via Gemma 4 (edge variant, ~3 GB, multimodal text+image). No part of the soil description, photograph or field sheet ever leaves the local network. This is the recommended configuration for governmental surveys, indigenous land studies, and unpublished research data; it also makes the pipeline reproducible without an internet connection. Cloud providers ("anthropic", "openai", "google") remain one argument away when they are the right call.

Architectural invariants preserved

- The VLM never classifies. Every extracted value carries `source = "extracted_vlm"`; the deterministic keys consume the resulting `PedonRecord` unaware of how each value was obtained.
- Provenance is preserved end-to-end. The `evidence_grade` on each `ClassificationResult` reflects whether decisive attributes came from `measured`, `predicted_spectra`, `extracted_vlm`, `inferred_prior`, or `user_assumed` – so a caller always knows how robust the classification is.
- Authority order is enforced. A pre-existing `measured` value is never silently overwritten by a later `extracted_vlm` value (unless `overwrite = TRUE`).

See Also

`vlm_provider`, `extract_horizons_from_pdf`, `classify_wrb2022`, `report`.

Examples

```
## Not run:
# The simplest possible end-to-end call -- local Gemma 4 edge.
res <- classify_from_documents(
  pdf      = "perfil_042_descricao.pdf",
  image    = "perfil_042_parede.jpg",
  report   = "perfil_042.html"
)
res$classifications$wrb$name
#> "Geric Rhodic Ferralsol (Clayic, Dystric, Epic, Ferric, Humic)"

# Cloud provider for a one-shot, production run
res <- classify_from_documents(
  pdf      = "perfil_042_descricao.pdf",
  provider = "anthropic"
)

# Different Gemma 4 size on Ollama
res <- classify_from_documents(
  pdf      = "perfil_042_descricao.pdf",
  provider = "ollama",
  model    = "gemma4:31b"
)

## End(Not run)
```

---

classify_from_photos *Classify a soil profile from field photographs alone*

Description

A no-lab-data pipeline: profile photographs are sent to a vision-language model for Munsell-colour and (optionally) site-metadata extraction; the missing horizon attributes are back-filled from a SoilGrids depth prior; and the WRB 2022, SiBCS 5 and USDA Soil Taxonomy keys are run on the assembled `PedonRecord`.

Usage

```
classify_from_photos(
  images,
  lat = NULL,
  lon = NULL,
  country = NULL,
  provider = NULL,
  systems = c("wrb", "sibcs", "usda"),
  soilgrids = TRUE,
  depth_profiles = NULL,
  on_missing = "silent"
)
```

Arguments

`images`	Either a character vector of profile-photo paths, or a named list with elements `profile` (character vector, required) and `fieldsheet` (character vector, optional).
`lat, lon`	Optional decimal-degree coordinates. When supplied they seed `pedon$site` and are used for the SoilGrids fetch; a field sheet can also supply them through extraction.
`country`	Optional ISO-2 country code; passed through to the constructed pedon's site metadata.
`provider`	A vision-language provider: an **ellmer** chat object for live use, or a `MockVLMProvider` for testing and offline demos. Required – there is no default, so a real classification is never produced from canned data by accident.
`systems`	Character vector, any subset of `c("wrb", "sibcs","usda")`.
`soilgrids`	If `TRUE` (default) missing horizon attributes are back-filled from a SoilGrids depth prior via `apply_soilgrids_depth_prior`.
`depth_profiles`	Optional named list of six-slice SoilGrids depth profiles, forwarded to `apply_soilgrids_depth_prior`. Supplying it skips the network call.
`on_missing`	Forwarded to the classifiers; default `"silent"`.

Details

Because every value originates from a photograph or a spatial prior, the classification's evidence grade is low by construction (`D` for VLM-extracted attributes, `C` where a SoilGrids prior contributed). The result is a screening estimate, not a substitute for a described and sampled profile.

Value

A named list with one ClassificationResult per requested system ($wrb, $sibcs, $usda), the constructed $pedon, its $provenance ledger, and a one-row $summary data frame. If extraction yields no horizons the list instead carries $error and a NULL pedon.

See Also

extract_munsell_from_photo, apply_soilgrids_depth_prior, compute_per_attribute_evidence_grade.

Examples

```
## Not run:
# Live use with an ellmer chat:
res <- classify_from_photos(
  images   = list(profile = "profile.jpg", fieldsheet = "sheet.jpg"),
  lat = -22.7, lon = -43.6, country = "BR",
  provider = ellmer::chat_anthropic())
res$wrb$name
res$wrb$evidence_grade   # "D" or "C"

## End(Not run)
```

---

classify_hscs2024 *Classify a pedon under the Hungarian diagnostic soil classification*

Description

Keys a PedonRecord through the Hungarian diagnostic soil classification of Micheli et al. (2024, second approximation), following its key to the soil types. **Experimental**: this release keys the 15 types (Laptalaj, Antropogen talaj, Koves-sziklas vaztalaj, Kozethatasu talaj, Szolonyec talaj, Szoloncsak talaj, Duzzado agyagtalaj, Reti talaj, Karbonat talaj, Mezosegi talaj, Agyagbemosodasos talaj, Homoktalaj, Barnafold, Hordalektalaj, Foldeskopar). Since v0.9.316 the name also carries the subtype properties of the type's table (before the type, the first with a capital letter) and its variety properties (after it, in parentheses), in the table's order, with the depth modifiers of 8.7 (felszintol, sekelyen, kozepmelyen, melyben) on properties defined by a layer: e.g. "Savanyu, humuszszegeny Agyagbemosodasos talaj (sekelyen kozettormelekes)", with the accents.

Usage

```
classify_hscs2024(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set (load_rules("hscs2024")).
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Details

The names are the Hungarian singular, as the text's naming example writes them; levels$name_en gives an English gloss. The key reads the FAO (2006) horizon notation the text adopts (an l or r suffix shows reductive conditions; i slickensides; k carbonates; n sodium; p ploughing; t clay accumulation; u artefacts; w a cambic horizon), moist and dry Munsell colours, texture (Hungarian names of the USDA triangle the text prints, English names, or lab clay, silt and sand), organic carbon, bulk density, CaCO3, effervescence, secondary carbonates, base saturation, exchangeable Na, Mg and the CEC, structure, slickensides, cracks, clay films, coarse fragments, artefacts, the layer origin, the dipyridyl test (childs_test), redoximorphic features, and the total water-soluble salts (soluble_salts_pct, Hungarian method). From the site it reads land_use, methane_odour, geomembrane, mechanical_mixing_depth_cm and anthropic_truncation. Features a describer records when met (artefacts, mottles, secondary carbonates, cracks) are absent when not recorded. The choices where the text is silent are listed in system.file("rules", "hscs2024", "readings.yaml", package = "soilKey").

Value

A ClassificationResult with system_id "hscs2024", the full name in name and the type in rsg_or_order; levels has the type with its English gloss and one row per subtype and variety property named; qualifiers holds the subtypes, the varieties and status, every property of the type's table with its verdict (TRUE, FALSE, NA), depth and label; and the key trace and provisional_because (types before the assigned one that could not be decided).

References

Micheli E., Fuchs M., Lang V., Csorba A., Dobos E., Szegi T., Kele G. (2024). A diagnosztikus szemleletu hazai talajosztalyozasi rendszer. Utmutato. Masodik kozelites. Magyar Agrar- es Elettudomanyi Egyetem, Godollo.

See Also

make_hscs2024_example, soilkey_systems.

Examples

```
res <- classify_hscs2024(make_hscs2024_example("steppe"), on_missing = "silent")
res$name
res$levels
```

`classify_ka5` *Classify a pedon under the German KA5 soil systematics (2005)*

Description

Keys a `PedonRecord` to the divisions (Abteilungen), classes (Klassen) and types (Typen) of the German soil systematics of the Bodenkundliche Kartieranleitung, 5th edition (Ad-hoc-AG Boden 2005, KA5), by its key Tab. 51. **Experimental**: this release keys the 4 divisions, 21 classes and 56 types, and the 223 subtypes of the type texts (pp. 202-262: 138 of the terrestrial types, 85 of the semiterrestrial, subhydric and Moore types), and the varieties and subvarieties of Tab. 46-50 (pp. 263-282).

Usage

```
classify_ka5(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	Optional pre-loaded rule set (`load_rules("ka5")`).
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`.
`gapfill`	Opt-in within-pedon gap-fill, as in `classify_usda`.

Details

The subtypes (Hugo's decision Q21): inside a type the transitional subtypes (X-Y, a subtype of the type Y with the features of X) are tested first in printed order, then the deviation subtypes in printed order, the Norm subtype last as the fallback; the first that passes names the subtype, and every subtype of the type that passes is listed in `trace$subtypes_passing`. Each subtype is one of its printed formulas matched on the horizons (`match_horizon_sequence`), its depth windows from the GOF and its numbers (base saturation, clay differences, Cp/Fep, slope ...). The Rigosol and the Treposol take an open subtype in brackets for the type still recognisable under the R ("[BB-YY]", Q22); the Treposol has no Norm subtype, the Moorgley neither (an uH over a Gr takes the open "[Uebergangsmoorgley]", an hHm under the Erdhochmoor the open "[vermulmtes Erdhochmoor]").

The varieties (Hugo's decision Q1): an additive level after the subtype, rows of level `"variety"` in `levels` (the code of Tab. 49 with the digit of its subvariety, the word of the name, the gloss, the full symbol in a `symbol` column). A variety is offered where Tab. 50 marks the subtype, plus the six of its note 1 (og, c, z, k, v, ku) everywhere; each criterion of Tab. 46 is read on the horizons (the symbols written, or the recorded data), the humus forms only as recorded; one the data leave open is not named but listed in `trace$varieties`. The name becomes the full name (adjectives before the subtype's name, nouns in brackets before its deciding part: "podsolige Normbraunerde",

"(Acker)Gley"), the symbol (rules I-VII: `pBBn`, `g4.p2BBn`, `BB-rGG`) is in `trace$variety$symbol`. The further subtypes of the type that pass join as a combination (Tab. 46 no. 17; Q21: `PP-GG-BB`, `GGh-GGc`).

The overlays (Tab. 46 no. 18) are reported in `trace$overlay` and the warnings, and enter the name as a variety (`YK/BB`, Kolluvisol ueber Braunerde); the levels above it do not change. A mineral cover on a Moor keeps the Moor (Hugo's decision Q7): under 2 dm it is named as substrate only, 2 to < 3 dm "flach X ueber HN" (`X\HN`), 3 to < 4 dm "X ueber HN" (`X/HN`), the cover keyed as the mineral type it would be. A fersiallitic or ferrallitic palaeosol (a relict or fossil Bj or Bu) under the soil is named "... ueber Fersiallit / Ferrallit" (Q19); a buried soil (fossil horizons of a layer II or deeper) is keyed alone from its own top and named by the thickness of the soil above (under 3 dm "flach", 3 to < 7 dm; 7 dm or more the upper soil alone).

The key reads the horizons of the profile (`identify_horizons(pedon, "ka5")`): the KA5 notation where the profile is written in it (declared with `site$designation_system = "ka5"`, or recognised by a sign only it writes: the degree sign, a lower-case prefix before a master such as lC, fAh, aM, tmGo, Sw, Sd, Axh, Bfv ...), otherwise soilKey's order of tests on the recorded data. The last symbol of a designation carries the horizon (Bv-Go is a Go, Go-Bv a Bv). Depths are read from the GOF, the top of the first horizon that is not L, O, Of, Oh or Ovh (p. 79). The order and gates are Tab. 51's (the Moore, the subhydric soils, the semiterrestrial soils, the terrestrial soils), with the numbers of the type texts where the compressed key differs and two rules of the figures Abb. 22 and 23 (Hugo's decisions Q5 and Q6). A gate the data leave open makes the result provisional (`provisional_because`); a type that holds only on the designation, a measurement missing (the Fahlerde without its clay, C13), or by the key's order with the record missing (the Rigosol without the kind of deep cultivation, Q23), is listed there too.

Names as printed (German) in `name` and `levels$name`, with soilKey's English gloss in `levels$name_en`. The horizon panel (one row per layer: depths from the GOF, the horizon read, the route notation, data or both, the candidates, the evidence) is in `trace$horizon_panel`, the full frame in `trace$horizons`, the consistency flags of the figures in `trace$figures`. The readings where the text is silent are listed in `system.file("rules","ka5", "readings.yaml", package = "soilKey")`.

Value

A `ClassificationResult` with `system_id` `"ka5"`: the deepest class assigned in `name`, the division in `rsg_or_order`, the division, class and type in `levels`, the key trace, and `provisional_because`.

References

Ad-hoc-AG Boden (2005). Bodenkundliche Kartieranleitung. 5. Auflage. Bundesanstalt fuer Geowissenschaften und Rohstoffe, Hannover.

See Also

`make_ka5_example`, `identify_horizons`, `soilkey_systems`.

Examples

```
res <- classify_ka5(make_ka5_example("BB"), on_missing = "silent")
res$name
res$levels
```

`classify_kidpr2008` *Classify a pedon under the Russian soil classification (Field Guide, 2008)*

Description

Keys a `PedonRecord` to the trunks and divisions of the Russian soil classification, in the version of its Field Guide (Polevoi opredelitel' pochv Rossii, Dokuchaev Soil Science Institute, 2008). **Experimental**: this release keys the five trunks (key 2.2: the organogenic, synlithogenic, postlithogenic and weakly developed soils, and trunk E, not a soil: rock outcrops, technogenic surface formations, or a deposit that is not humified over a buried soil) and the 27 divisions (key 2.3), and the types of every division (the profile formulas of chapter 4: CA to CJ in v0.9.338, CK to CT in v0.9.339, trunks A, B and D in v0.9.342). The chemozems CU have no types: their name is built from the pollutants and the soil they were (v0.9.342). The subtypes of every division's list follow the type (v0.9.345).

Usage

```
classify_kidpr2008(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE,
  subtypes = TRUE,
  species = subtypes
)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	Optional pre-loaded rule set (`load_rules("kidpr2008")`).
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`.
`gapfill`	Opt-in within-pedon gap-fill, as in `classify_usda`.
`subtypes`	Name the subtypes after the type (default `TRUE`); `FALSE` stops at the type.
`species`	Name the species of chapter 5.1 after the subtypes (default: as `subtypes`).

Details

The keys read the diagnostic horizons of the profile: the Russian notation where the profile is written in it (declared with `site$designation_system = "kidpr"`, or recognised by an index only it writes: AY, EL, BEL, BHF, BCA, PU, TUR ...), otherwise the key to horizons 2.1 run on the recorded data (see `identify_horizons`). From them the keys read whether a horizon is present, whether it starts the profile (at the surface, or under a deposit of 40 cm or less), whether it lies directly over another, whether the profile consists of one horizon only, the depth of hard rock, a deposit at the surface (TPO and the deposits the key to horizons finds on the data included), the peat body of trunk A (more than 50 cm, the agro-peat horizons included) and the thickness of a

turbated horizon; of the site, a river floodplain on modern alluvium (`landform`, `landform_element`, `layer_origin`, `rock_origin`, `parent_material`, the alluvial mark of the notation; a terrace is no floodplain) and the contaminated horizon X (`contamination_hazard`). The ochre horizon BAN is formed in volcanic ash, so it makes the volcanic soils alone. With several organic and humus horizons at the surface the thickest is the diagnostic one (p. 37), for step 5 of the key to trunks as for the divisions; FAO and WRB transitions (AB, BC, E/B) in which the key on the data finds no diagnostic horizon are transitions, as those of the Russian notation; a missing depth leaves the step that needs it undecided.

The keys are branching trees; soilKey flattens them into ordered lists in which each gate is the conjunction of the conditions of the book's path to the class, and repairs only the gaps the book's own types prove; the other combinations the key leaves without an exit are "not provided by the Field Guide of 2008": the trunk is assigned and the division left unassigned, with a note. Names: an English gloss in `name`, the Russian of the book in `levels$name_native`. The readings where the text is silent are listed in `system.file("rules", "kidpr2008","readings.yaml", package = "soilKey")`.

The type is named by the formula of its profile: the profile is matched against every formula of its division (see `kidpr2008_types`), closed on the horizons the division's formulas write, most specific first and agro before natural. A profile no formula matches keeps its division, the type unassigned, with the nearest formulas in the warnings and in the trace (phase `types_nearest`).

A chemozem (CU, an X horizon: pollutants of the extremely hazardous category within 30 cm) is named, as p. 134 asks, by the kind of pollution and the original soil: the `pollutants` of the layers of the X horizon, and the soil the profile keys to with the X masked (the X overlays and the contamination records removed), down to its type. The type level is soilKey's `"[CU:<code of the original soil>]"`, named "Chemozems polluted with Cu, Ni, Co; original soil: Sod-podzolic"; the original soil keyed is in the trace (phase `chemozem`), and an original soil left undecided makes the name provisional.

The subtypes (v0.9.345): each division has one list of subtypes in chapter 4 (26 lists, 558 entries; the chemozems have none), each the index of a genetic feature of section 3.2, read by the feature tests (the recorded data first, the Russian notation where they say nothing). The soil is named by its type and every subtype of the list that passes, in the order of the list, comma-separated and in the plural as printed ("Podzolistye gleevatye, yazykovatye"): the subtypes are the rows of level `"subtype"` of `levels` (code: soilKey's key of the entry; `name` the English gloss, `name_native` the Russian), the full name in English in `name` and in Russian in `qualifiers$name_native`. When every entry is FALSE the soil is of the typical subtype, soilKey's "[tipichnye]" ("[typical]", code NA); when some are undecided the name is provisional, the undecided entries in `provisional_because` ("subtype:<key>") and in a warning. A feature the formula of the type already writes (the g of the gley types, the BF of CB05, the permafrost sign of the cryozems ...) is not named again, and a sentence of 3.2 that names a single type restricts its subtype to it. Each entry, with its evidence, is in the trace (phase `subtype`). A soil whose type is unassigned gets no subtypes. The lists are in `system.file("rules", "kidpr2008", "subtypes.yaml",package = "soilKey")`.

The species (v0.9.347): the twelve tables of chapter 5.1, the classes of the thickness of the humus, peat and agro horizons, of the depth and place of gleying, of the lower boundary of the bleached horizon, of the layer over the solonetzic horizon, of the depth of the carbonates and of the soluble salts and, in the organogenic trunk, of the decomposition, the botanical composition (`peat_botanical_composition`) and the thickness of the peat. Each table that applies to the type is read, and its class follows the subtypes in the name, in the order of the tables ("Dernovo-podzolistye [tipichnye], srednemoshchnye, glubokoelyuvial'nye"; the bleached species in the prefix form of p.

136): rows of level "species" of levels (code "<table>:<class>"), qualifiers$species and species_native. A depth not recorded, an undecided layer at the edge of a horizon, or a horizon that reaches the bottom of the description when a deeper class is possible, leaves the table undecided: it is not named, it is listed in qualifiers$species_undecided and in a note, and the name stays as it is. The tables are in system.file("rules","kidpr2008", "species.yaml", package = "soilKey"); the varieties and ranks (5.2, 5.3) are not keyed.

The 1977 analogue (v0.9.347): when the type is assigned, the rows of the correlation table of chapter 6 that stand for it are added as a note (a line in warnings) and as qualifiers$analogue_1977 (the 2004 name and formula, the page, whether the soils were distinguished in 1977 and the units of 1977 with their level and extent), never changing the classification; where the table says the soils were not distinguished or has no row for the type, the note says so. The table is in system.file("rules", "kidpr2008", "correlation_1977.yaml", package = "soilKey").

Value

A ClassificationResult with system_id "kidpr2008", the deepest class assigned in name (the type with its subtypes, else the division, else the trunk), the trunk in rsg_or_order, the trunk, division, type, subtypes and species in levels (with name_native, the Russian name), the key trace, the notes on a profile that is not a soil or whose combination of horizons the Field Guide does not provide for in warnings, and provisional_because (classes before the assigned one that could not be decided).

References

Polevoi opredelitel' pochv Rossii [Field Guide of Russian Soils] (2008). Dokuchaev Soil Science Institute, Moscow.

See Also

make_kidpr2008_example, identify_horizons, soilkey_systems.

Examples

```
res <- classify_kidpr2008(make_kidpr2008_example("CA"), on_missing = "silent")
res$name
res$levels
```

---

classify_nzsc4 *Classify a pedon under the New Zealand Soil Classification*

Description

Keys a PedonRecord through the New Zealand Soil Classification, 4th edition (Manaaki Whenua - Landcare Research 2026), following its Key to Soil Orders and keys to groups and subgroups. **Experimental**: this release keys the 15 orders (Organic, Gley, Ultic, Podzol, Allophanic, Pumice, Melanic, Semiarid, Oxidic, Granular, Pallic, Brown, Anthropic, Recent, Raw Soils), their 77 groups (since 0.9.298, e.g. Orthic Brown Soils, Perch-gley Pallic Soils) and their 336 subgroups (keyed

from 0.9.299 to 0.9.302, e.g. Peaty Orthic Gley Soils, Humose Orthic Podzol Soils, Mottled Fragic Pallic Soils, Typic Orthic Brown Soils); seven groups have no subgroups and keep the group name. The Cryic Soils of the appendix are not keyed.

Usage

```
classify_nzsc4(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set (load_rules("nzsc4")).
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Details

The NZSC reads horizon notations and field tests. The diagnostics read the New Zealand notation (Ah, E, Bw, Bt, Bg, Bh, Bs, Bo, Bx, Bw(g), Bhm, Bsm, Blm, CR, R, Of, Om, Oh, Lm, Lh, an e prefix for anthropic soil material; FAO designations are read by the same letters), moist Munsell colours, texture (field_texture words or codes, or lab clay, silt and sand), pH, P retention (phosphate_retention_pct), dry bulk density, volcanic glass, coarse fragments and their origin, structure, clay coatings, carbonate, ESP, and the field tests naf_test (reactive-aluminium test: none, slight, moderate, strong), childs_test, sensitivity_class, fluidity_class, stickiness and strength_moist. From the site it reads drainage_class, parent_material (Melanic Soils), anthropic_truncation and overthickened_a (the note to the key on overthickened A horizons), and designation_system. The groups also read the packing (packing_degree), Ksat (ksat_mm_h), penetration resistance, where the redox segregations are (redox_segregations_location), sulfate S, roots, a Maori cultural layer (maori_cultural_layer) and the colour of coatings; and, on the site, soil_temp_30cm_C, mean_annual_air_temp_C, surface_water, methane_odour, mechanical_mixing_depth_cm and artificially_drained. The subgroups also read sulfidic S (sulfidic_s_pct), the EC of a 1:5 extract (ec_1to5_dS_m), the share of Sphagnum in the fibres (sphagnum_fibres_pct), a shelly layer (shelly_layer), exchangeable Ca and Mg, clay films, the colour of coatings on pores (pore_coatings_hue_moist, pore_coatings_value_moist, pore_coatings_chroma_moist), the weathering of rock fragments (weathering_stage, weathered_fragments_pct) and the cementation class; on the site, mechanical_mixing_purpose and acid_neutralising_amendment (Mixed and Fill Anthropic Soils). Features a describer records when met (stratification, nodules, cracks, refuse or cultural layers, permafrost, surface water) are absent when not recorded. The choices where the text is silent are listed in system.file("rules", "nzsc4", "readings.yaml", package = "soilKey").

Value

A ClassificationResult with system_id "nzsc4", the deepest class in name (the subgroup, or the group where no subgroup is keyed yet), the order in rsg_or_order, the classes of each level in levels, the key trace, and provisional_because (classes before the assigned one that could not be decided).

References

Manaaki Whenua - Landcare Research (2026). New Zealand Soil Classification. 4th ed. doi:10.7931/dl1lrss42026.

See Also

make_nzsc4_example, soilkey_systems.

Examples

```
res <- classify_nzsc4(make_nzsc4_example("brown"), on_missing = "silent")
res$name
```

---

classify_profiles *Classify many profiles from one horizon table*

Description

Builds one PedonRecord per profile from a table with one row per horizon, and classifies each with classify_all. The column named by id says which profile a horizon belongs to; the other columns use soilKey's names (horizon_attribute_spec), or are renamed and converted by map. Site properties (site_attribute_spec), such as soil_moisture_regime, come from sites.

Usage

```
classify_profiles(
  horizons,
  sites = NULL,
  id = "profile_id",
  map = NULL,
  systems = "all",
  on_missing = "silent",
  ...
)
```

Arguments

horizons	A data.frame with one row per horizon and the column named by id.
sites	Optional site properties: a data.frame with one row per profile and the same id column, or a list of lists named by profile id. A missing value (NA) is read as not recorded. Profiles without an entry get only their id.
id	Name of the column that identifies the profile, in horizons and in a sites table (default "profile_id").
map	Optional data.frame with the columns from and to, and optionally factor: the column from of horizons becomes the column to, multiplied by factor (default 1) to convert its unit. For example, organic carbon in g/kg: from = "ORGC", to = "oc_pct", factor = 0.1.
systems, on_missing	Forwarded to classify_all; the default on_missing = "silent" keeps a batch quiet, and the missing column reports what the keys did not find.
...	Other arguments forwarded to classify_all, such as include_family.

Details

The names are checked once for the whole table: a horizon column or site property that no key reads is named in one warning (see unrecognised_attributes), not once per profile. Within a profile the horizons are put in order of top_cm. A profile that cannot be built or classified gets its reason in the error column, and the others are classified as usual.

Value

A data.frame with one row per profile, in the order of their first horizon in horizons: the id column, one column per system with the name of the class (as in the summary of classify_all), missing (the attributes the keys asked for and did not find, comma-separated) and error (NA, or why the profile could not be classified). The attributes "results" and "pedons" hold the full classify_all results and the pedons, in lists named by profile id.

See Also

classify_all for one pedon, classify_csv for one profile in a spreadsheet, unrecognised_attributes to check the names first.

Examples

```
h <- rbind(
  cbind(profile_id = "p1", as.data.frame(make_ferralsol_canonical()$horizons)),
  cbind(profile_id = "p2", as.data.frame(make_phaeozem_canonical()$horizons)))
s <- data.frame(profile_id = c("p1", "p2"),
                soil_moisture_regime = c("udic", "ustic"))
res <- classify_profiles(h, sites = s, systems = "usda")
res[, c("profile_id", "usda", "error")]


# Columns named and measured differently: rename and convert with `map`.
```

```
h2 <- h
names(h2)[names(h2) == "oc_pct"] <- "ORGC"
h2$ORGC <- h2$ORGC * 10                       # g/kg
m <- data.frame(from = "ORGC", to = "oc_pct", factor = 0.1)
classify_profiles(h2, sites = s, map = m)[, c("profile_id", "wrb", "sibcs", "usda")]
```

`classify_rp2008` *Attach a pedon to the references of the French Referentiel pedologique 2008*

Description

Attaches a `PedonRecord` to the references of the *Referentiel pedologique* 2008 (AFES; Baize and Girard coord.; Editions Quae, 2009). **Experimental**: this release tests the 110 references of the 34 groups of references (GER): alocrisols (2), andosols (6), anthroposols (5), arenosols (1), brunisols (2), the solums dominated by calcium or magnesium (7: rendosols, calcosols, dolomitosols, rendisols, calcisols, magnesisols, calcarisols), chernosols (3), colluviosols (1), cryosols (2), ferrallitisols and oxydisols (7), ferruginosols (4), fersialsols (4), fluviosols (4), grisols (3), gypsosols (2), histosols (5), leptismectisols (1), lithosols (1), luvisols (6), nitosols (1), organosols (4), pelosols (3), peyrosols (1), phaeosols (2), planosols (3), podzosols (8), rankosols (1), reductisols et redoxisols (3), regosols (1), salisols et sodisols (7), thalassosols (3), thiosols et sulfatosols (2), veracrisols (1) and vertisols (4). The horizons of the intertropical chapters (F, OX, RT, ND, FE, OXc, OXm) are read only in the intertropical domain, and the results of those chapters and of the nitosols say that the book calls them a first version (Q18); a fluviosol needs a recorded alluvial origin and is provisional while the position in the valley and the water table are not recorded (Q14). The histosols read their predominant H horizon between 40 and 120 cm, the luvisols their texture contrast by the bands of the chapter, the organosols the S/CEC of their Aho as a whole. The reductisols and redoxisols read the functional reductic or redox features starting within 50 cm and continuing over 50 cm (a G within 50 cm makes a reductisol whatever lies above), and their doubles with the luvisols, podzosols, pelosols, brunisols, colluviosols, cryosols and histosols are those the book names; the depth class of the hydromorphy is in `trace$hydromorphy`. The salisols and sodisols read the salic and sodic horizons within 60 cm, the thalassosols a recorded marine material (provisional while the tidal flooding or the polder is not recorded), the thiosols and sulfatosols, first in every case, the sulfur within 50 cm, the veracrisols their vermihumic Ah, the vertisols their SV and V, litho against topo by the site, else the SV.

Usage

```
classify_rp2008(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rules (load_rules("rp2008")).
on_missing	One of "warn" (default), "silent", "error".
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Details

The RP is not a key: a solum attaches to one, two or no reference, simply or imperfectly (pp. XIV, XVII). Every reference is tested every time: its sequences of reference horizons, matched on the horizon frame (the RP notation where the profile is written in it, else the horizons identified on the data, identify_horizons(pedon, "rp2008")), and the side tests of its chapter (depths, thicknesses, forbidden horizons, saturation, carbonates, permafrost, colluvial or anthropic materials). Each test is evaluated strictly and with the tolerances of p. 64 (depth and thickness 10 to 200 cm, pH +/- 0.2, CEC +/- 1, S/CEC +/- 5, organic carbon 8 +/- 2 and 30 +/- 5; any other threshold strict): a reference met only within tolerance, or by an imperfect case its chapter names (calcosol leptique, rendosol pachique, arenosol redoxique ...), is an imperfect attachment, never a simple one. The main name follows the book's explicit statements of precedence, then the degree, then the order of the rapid key (pp. 66-68); a double attachment is written only for the pairs the book prescribes (calcisol-calcosol, the anthroposols transformes); with no reference, the name is "[non rattache]" and the three nearest references are listed. The choice of one name is a convention of soilKey, and $candidates lists every reference that may attach the solum with its degree and the criteria that fail, pass only within tolerance or are undecided.

Names are French, as printed, in the singular (name, levels$name_native), with an English gloss (levels$name_en); the GER is a label (rsg_or_order, levels$ger), not a level. The qualifiers and the types come later. The readings where the text is silent are listed in system.file("rules", "rp2008","readings.yaml", package = "soilKey").

Value

A ClassificationResult with system_id "rp2008": the main name in name ("[non rattache]" when nothing attaches), the GER label in rsg_or_order, the references in the name in levels (level "references", code, name, name_native, name_en, ger, attachment: simple, imperfect, simple_or_imperfect, double), every reference that may attach in candidates (with the columns of ClassificationResult, group the GER, plus name_en, implemented and precedence, the book's statements that put a reference of the name before it or exclude it), the evaluation of every reference, the horizon frame and the rapid key in trace (with trace$position, the depth of the first ND, OXm or OXc of a ferrallitisol or oxydisol and its proximal or distal type), and provisional_because ("references:<code>": the references or GER that could take the name once their data are known).

References

Baize D., Girard M.-C. (coord.) (2009). Referentiel pedologique 2008. AFES, Editions Quae, Versailles.

See Also

make_rp2008_example, soilkey_systems, within_tolerance.

Examples

```
res <- classify_rp2008(make_rp2008_example("calcosols"), on_missing = "silent")
res$name
res$candidates[, c("code", "degree", "in_name")]
```

`classify_scsj2017` *Classify a pedon under the Soil Classification System of Japan*

Description

Keys a PedonRecord through the Soil Classification System of Japan (Japanese Society of Pedology 2017), following its key to great groups. **Experimental**: this release keys the ten great groups (Man-made soils, Organic Soils, Andosols, Podzols, Fluvic soils, Red-Yellow soils, Stagnic soils, Eutrosols, Brown Forest soils, Regosols) and their 26 groups (for instance Allophanic and Non-allophanic Andosols, Fluvic Paddy soils, Magnesian Eutrosols, Sandy Regosols), and their 120 subgroups (Haplic, Humic, Cumulic, Aquic ...); the soil phases follow. Each level gives the official English name and, in `levels$name_native`, the Japanese one.

Usage

```
classify_scsj2017(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set (`load_rules("scsj2017")`).
on_missing	One of `"warn"` (default), `"silent"`, `"error"`.
gapfill	Opt-in within-pedon gap-fill, as in classify_usda.

Details

The SCSJ reads the Japanese horizon notation (H, O, A, E, B, C, G with Go and Gr, R; p ploughed, g gleyed, t, w, s, h, ir, mn; a 2 or II prefix for a lithologic discontinuity, a trailing b for buried; the FAO/USDA notation by the same letters), moist Munsell colours, texture, organic carbon, the phosphate absorption coefficient (`phosphate_absorption_coef`, mg P2O5/100 g) or its substitutes (P retention, Alo + 1/2 Feo) and the field active-aluminium test (`active_al_test`), rubbed fibre, pH, base saturation, exchangeable Ca and Mg, Fe(d), Sio, Alp/Alo, the exchange acidity

(exchange_acidity_y1), the dipyridyl test (childs_test), fluidity and n_value, mottles and where they are, artefacts, and the recorded origin of the layers (layer_origin, rock_origin: fill, volcanic ejecta, alluvial deposits). From the site it reads land_use or flood_irrigated (paddy land), parent_material, drainage_class, artificially_drained, acid_neutralising_amendment and designation_system. The choices where the text is silent are listed in system.file("rules", "scsj2017", "readings.yaml", package = "soilKey").

Value

A ClassificationResult with system_id "scsj2017", the deepest class in name, the great group in rsg_or_order, the classes of each level in levels (with name_native), the key trace, provisional_because (classes before the assigned one that could not be decided), and in phases the soil phases of chapter V (temperature, man-made and forest phases; the class name is left as it is).

References

The Fifth Committee for Soil Classification and Nomenclature (2017). Soil Classification System of Japan. The Japanese Society of Pedology.

See Also

make_scsj2017_example, soilkey_systems.

Examples

```
res <- classify_scsj2017(make_scsj2017_example("andosol"), on_missing = "silent")
res$name
res$levels
```

---

classify_sibcs — *Classifica um pedon segundo o SiBCS 5a edicao (1o + 2o + 3o + 4o niveis)*

Description

v0.7 ligou as 13 ordens; v0.7.1 desce ao 2o nivel (subordens) via run_sibcs_subordem; v0.7.3 desce ao 3o nivel (Grandes Grupos) via run_sibcs_grande_grupo para as ordens progressivamente wiradas em inst/rules/sibcs5/grandes-grupos/<ordem>.yaml (Cap 14 Organossolos primeiro). Quando a subordem ainda nao tem bloco de Grandes Grupos, ou quando nenhum Grande Grupo passa (e nao ha catch-all default), a classificacao para no 2o nivel.

Usage

```
classify_sibcs(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
```

```
  include_familia = FALSE,
  gapfill = FALSE
)
```

Arguments

pedon	A PedonRecord.
rules	Conjunto de regras pre-carregado.
on_missing	Um de "warn" (default), "silent", "error".
include_familia	Quando TRUE (default FALSE), adiciona o 5o nivel categorico (Familia) via classify_sibcs_familia. O label textual da Familia aparece em $trace$familia_label, e a lista de FamilyAttributes em $trace$familia.
gapfill	Preenchimento opcional de lacunas por interpolacao intra-perfil, default FALSE (no-op; classificacao byte-identica). TRUE preenche celulas NA interiores dos atributos continuos por profundidade; um vetor de caracteres restringe aos atributos citados; uma lista nomeada e repassada a gapfill_within_pedon. Celulas preenchidas recebem proveniencia inferred_prior, baixando o grau de evidencia para "C". Opera sobre copia profunda – o pedon do chamador nunca e modificado.

Value

Um ClassificationResult cujo name eh o nome completo da classe atribuida no nivel mais profundo (Grande Grupo > Subordem > Ordem) e rsg_or_order eh o nome da ordem (e.g. "Organossolos"). Os codigos de cada nivel e o trace ficam em $trace.

Examples

```
pedon <- make_latossolo_canonical()
res <- classify_sibcs(pedon)
res$name
```

---

classify_sibcs_familia

Classifica um perfil no 5o nivel categorico do SiBCS (Familia)

Description

Aplica as dimensoes pertinentes a ordem do solo e devolve uma lista nomeada de FamilyAttribute. O label textual da Familia eh formado adicionando-se cada value nao-nulo apos a designacao do 4o nivel, separados por virgulas (Cap 18, p 281).

Usage

```
classify_sibcs_familia(
  pedon,
  ordem_code = NULL,
  sg_code = NULL,
  max_depth_cm = 200,
  nome_superior = NULL
)
```

Arguments

pedon	A PedonRecord.
ordem_code	Codigo da ordem (1 letra: "P", "L", ...). Se NULL, sera derivado de sg_code.
sg_code	Codigo do subgrupo do 4o nivel (e.g. "PVdAr"). Opcional; usado para ajustes especificos por SG (e.g. forcar subgrupamento textural em arenicos/espessarenicos).
max_depth_cm	Profundidade da secao de controle (default 200 cm).
nome_superior	Nome da classe nos niveis mais altos (e.g. o nome do subgrupo). Opcional; quando ja diz Eutroficos / Distroficos, a saturacao por bases nao e repetida, e quando ja diz Aluminicos, o termo "alico" nao e repetido (Cap 18, p 285).

Details

Esta funcao NAO eh uma chave determinista: cada perfil recebe SIMULTANEAMENTE todos os adjetivos pertinentes (multi-rotulo).

Value

Lista nomeada de FamilyAttribute.

Status v0.7.14.A

Implementadas 5 dimensoes – grupamento textural, subgrupamento textural, distribuicao de cascalhos, constituicao esqueletica, tipo de horizonte superficial. Outras dimensoes (prefixos epi/meso/endo, saturacao de bases, alico, mineralogia, atividade da argila, oxidos de ferro, andico, especificos de Organossolos) adicionadas em sub-commits subsequentes.

References

Embrapa (2018), SiBCS 5a ed., Cap 18, pp 281-288.

`classify_usda` *Classify a pedon under USDA Soil Taxonomy (13th edition)*

Description

Walks the canonical USDA key (Order -> Suborder -> Great Group -> Subgroup) using YAML rule files at:

- `inst/rules/usda/key.yaml`: Order key (12 entries)
- `inst/rules/usda/suborders/<order>.yaml`
- `inst/rules/usda/great-groups/<order>.yaml`
- `inst/rules/usda/subgroups/<order>.yaml`

Usage

```
classify_usda(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  include_family = FALSE,
  infer_temperature = TRUE,
  gapfill = FALSE
)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	Optional pre-loaded rule set.
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`.
`include_family`	If `TRUE`, derive and prepend the 5th-level family modifiers. Default `FALSE` (output byte-identical to earlier versions).
`infer_temperature`	When deriving the family, infer the soil temperature regime from latitude/elevation if `site$soil_temperature_regime` is absent (default `TRUE`). See `family_temperature_regime_usda`.
`gapfill`	Opt-in within-pedon depth gap-fill, default `FALSE` (no-op, classification stays byte-identical). `TRUE` fills interior `NA` cells of the continuous depth-trending attributes by linear interpolation from the profile's own measured horizons; a character vector restricts it to those attributes; a named list is passed to `gapfill_within_pedon`. Filled cells carry `inferred_prior` provenance, so the evidence grade drops to `"C"`. Runs on a deep copy – the caller's pedon is never mutated.

Details

With `include_family = TRUE` it additionally derives the 5th category, the **family** – a set of class modifiers (particle-size, mineralogy, CEC-activity, reaction, temperature regime, depth) PREPENDED to the subgroup name, e.g. *"fine, kaolinitic, isohyperthermic Rhodic Hapludox"*. See `classify_usda_family`.

Value

A ClassificationResult with deepest-level taxon name. Each level's trace is in $trace; the family attributes are in $trace$family.

References

Soil Survey Staff (2022). Keys to Soil Taxonomy, 13th edition. USDA Natural Resources Conservation Service.

Examples

```
pedon <- make_ferralsol_canonical()
res <- classify_usda(pedon)
res$name
# include the 5th (family) level:
classify_usda(pedon, include_family = TRUE)$name
```

---

classify_usda_family *Classify the USDA family (5th level) of a pedon*

Description

Runs the applicable family-modifier dimensions and returns them as a named list of FamilyAttribute objects (multi-label; each dimension is orthogonal). Mirrors classify_sibcs_familia.

Usage

```
classify_usda_family(
  pedon,
  order_code = NULL,
  subgroup_code = NULL,
  infer_temperature = TRUE
)
```

Arguments

pedon	A PedonRecord.
order_code	Optional USDA order code (selects applicable dimensions).
subgroup_code	Optional subgroup code (reserved for refinements).
infer_temperature	Passed to family_temperature_regime_usda.

Value

Named list of FamilyAttribute objects.

References

Soil Survey Staff (2022), KST 13th ed., Ch. 16–17.

See Also

`family_label_usda`, `classify_usda`.

`classify_via_smartsolos_api` *Classify a PedonRecord via Embrapa's SmartSolosExpert REST API*

Description

Sends a soilKey `PedonRecord` to the SmartSolosExpert REST endpoint maintained by Embrapa (Glauber Vaz's PROLOG-based implementation of the SiBCS classifier) and returns the resulting four-level classification (Ordem / Subordem / Grande Grupo / Subgrupo) wrapped in a soilKey `ClassificationResult`.

Usage

```
classify_via_smartsolos_api(
  pedon,
  api_key = Sys.getenv("AGROAPI_TOKEN"),
  endpoint = c("classification", "verification"),
  drenagem = NULL,
  reference_sibcs = NULL,
  base_url = "https://api.cnptia.embrapa.br/smartsolos/expert/v1",
  timeout_seconds = 30,
  post_fn = NULL,
  verbose = TRUE
)
```

Arguments

`pedon`	A `PedonRecord`.
`api_key`	Bearer token. Defaults to `Sys.getenv("AGROAPI_TOKEN")`. Required unless `post_fn` is supplied (test injection).
`endpoint`	One of `"classification"` (default; classify only) or `"verification"` (classify + compare against user-supplied `reference_sibcs`).
`drenagem`	Optional drainage class. Integer 1..8 or Portuguese string (`"bem drenado"` etc.).
`reference_sibcs`	Optional named list (`ordem`, `subordem`, `gde_grupo`, `subgrupo`) used by the `"verification"` endpoint as the user's reference.
`base_url`	Override base URL. Default `"https://api.cnptia.embrapa.br/smartsolos/expert/v1"`.

`timeout_seconds`
: HTTP timeout (default 30).

`post_fn`
: Internal: function with signature `function(payload) -> response_list` for unit tests. When supplied, the network is bypassed.

`verbose`
: If `TRUE` (default), emits a one-line summary.

Details

This is an **external classifier** – the package does not host or replicate the PROLOG rules. The function exists so soilKey users can cross-validate the local classifier against an authoritative Embrapa-hosted reference. Use the `"verification"` endpoint to compare against your own user-supplied reference classification (the API returns a per-level match `summary` with counters `L0..L4`).

Authentication: register a free AgroAPI account at `https://www.agroapi.cnptia.embrapa.br/portal/`, subscribe to the SmartSolosExpert API and generate an access token. Pass it via the `AGROAPI_TOKEN` environment variable or the `api_key` argument.

Value

A `ClassificationResult` with `system = "SiBCS 5a edicao (SmartSolosExpert API)"` and the four taxonomic levels in `rsg_or_order` (Ordem) and `qualifiers` (Subordem / GdeGrupo / Subgrupo). Verification-mode responses additionally carry `trace$smartsolos_summary` (the per-level match counters `L0..L4`).

References

Vaz, G. J., Silva Neto, L. de F. da, & Barbedo, J. G. A. (2025). SmartSolos Expert: an expert system for Brazilian soil classification. *Smart Agricultural Technology*, 10, 100735. doi:10.1016/j.atech.2024.100735.

Vaz, G. J., Silva Neto, L. de F. da, Lima, R. N., & Oliveira, S. R. de M. (2019). Uma API para a classificacao de solos do Brasil. In *Anais do 12 Congresso Brasileiro de Agroinformatica* (SBIAGRO 2019), pp. 63-72. Ponta Grossa.

Vaz, G. J., Silva Jr, A. F., & Silva Neto, L. de F. da (2023). Brazilian soil data for taxonomic classification. *Redape*, V1. doi:10.48432/PYKKA7.

See Also

`classify_sibcs` for the local PROLOG-free classifier; `compare_smartsolos` for a side-by-side comparison helper; `benchmark_redape` for the gold-standard curated dataset published by the same authors.

Examples

```
## Not run:
Sys.setenv(AGROAPI_TOKEN = "<your token>")
res <- classify_via_smartsolos_api(make_argissolo_canonical())
res$rsg_or_order      # "ARGISSOLO"
res$qualifiers
#> $subordem  "VERMELHO"
#> $gde_grupo "Distrofico"
```

```
#> $subgrupo  "tipico"

## End(Not run)
```

`classify_with_engine_heuristic` *Classify a pedon with the engine chosen by 'pick_engine()'*

Description

Convenience wrapper that routes `classify_wrb2022` / `classify_sibcs` / `classify_usda` through whichever engine the heuristic recommends for the specific pedon.

Usage

```
classify_with_engine_heuristic(
  pedon,
  system = c("wrb2022", "sibcs", "usda"),
  min_score = 3L,
  ...
)
```

Arguments

`pedon`	A `PedonRecord`.
`system`	One of `"wrb2022"`, `"sibcs"`, `"usda"`.
`min_score`	Forwarded to `pick_engine`.
`...`	Forwarded to the underlying classifier.

Value

The result of the chosen classifier (a `ClassificationResult`). The chosen engine is captured in `$trace$engine_used`.

`classify_with_uncertainty` *Posterior distribution over classification outcomes*

Description

Runs `n` Monte-Carlo perturbations of a pedon and tallies the resulting classes into an empirical posterior. Unlike `classification_robustness`, the perturbation magnitude of every `(horizon, attribute)` cell is scaled by its provenance evidence grade (see `get_perturbation_scale`): an A-grade measurement is nudged by a few percent, an E-grade assumption by a third of its value. The posterior therefore reflects not just how close the profile sits to a key boundary, but how trustworthy the inputs that placed it there actually are.

Usage

```
classify_with_uncertainty(
  pedon,
  n = 200L,
  system = c("wrb2022", "sibcs", "usda"),
  level = c("rsg", "name"),
  scales = NULL,
  sensitivity = TRUE,
  seed = 42L
)
```

Arguments

pedon	A PedonRecord.
n	Number of Monte-Carlo draws (default 200).
system	One of "wrb2022", "sibcs", "usda".
level	"rsg" (default; compare the RSG / order) or "name" (compare the full classification name, qualifiers included – strictly more uncertain).
scales	Optional named list overriding the default per-grade magnitudes; each element has the shape returned by get_perturbation_scale, keyed by grade letter.
sensitivity	If TRUE (default) also computes a leave-one-attribute-out sensitivity ranking. Set FALSE to skip that extra pass when only the posterior is needed.
seed	Random seed for reproducibility.

Value

A list of class "soilkey_uncertainty" with elements: posterior (named numeric vector summing to 1, sorted descending), top1 (the modal class), entropy (Shannon entropy of the posterior, natural log), sensitivity (a data.table of attribute / importance, or NULL), n_runs, n_success, baseline, system and level.

See Also

classification_robustness, get_perturbation_scale, compute_per_attribute_evidence_grade.

Examples

```
p <- make_ferralsol_canonical()
u <- classify_with_uncertainty(p, n = 5, system = "wrb2022")
u$posterior   # P(RSG = x)
u$entropy     # near 0 for a robust profile
```

`classify_wrb2022` *Classify a pedon under WRB 2022*

Description

High-level classification entry point. Pre-computes the implemented diagnostic horizons (argic, ferralic, mollic) for transparent reporting, runs the key, and assembles a `ClassificationResult` with the trace, ambiguities, missing-data hints, evidence grade, and (in future) prior sanity check.

Usage

```
classify_wrb2022(
  pedon,
  prior = NULL,
  prior_threshold = 0.01,
  on_missing = c("warn", "silent", "error"),
  rules = NULL,
  strict = NULL,
  specifiers = FALSE,
  gapfill = FALSE
)
```

Arguments

`pedon`	A `PedonRecord`.
`prior`	Optional spatial prior – a `data.table` with columns `rsg_code` and `probability`, typically the return value of `spatial_prior`. If supplied, the result records a `prior_check` entry from `prior_consistency_check`; an inconsistent prior also emits a warning. The deterministic key is NEVER overridden by the prior.
`prior_threshold`	Probability below which the prior triggers an "inconsistent" warning (default 0.01).
`on_missing`	One of `"warn"` (default), `"silent"`, `"error"`. Behaviour when the trace reports missing attributes.
`rules`	Optional pre-loaded rule set.
`strict`	Logical or `NULL`. Controls WRB Tier-3 strict mode for the per-RSG numerical gates (Vertisols, Andosols, Gleysols, Planosols, Ferralsols, Chernozems, Kastanozems). When `NULL` (default) the gates follow `getOption("soilKey.rsg_strict", FALSE)`. Passing `TRUE` or `FALSE` forces strict mode on or off for the duration of this call; see the individual RSG-gate help pages (e.g. `ferralsol`) for the strengthened thresholds.
`specifiers`	Logical. When `TRUE`, auto-attach WRB 2022 Ch 5 depth specifiers (Epi-/Endo-/Bathy-/Amphi-/Panto-/Kato-) to depth-anchored qualifiers based on the diagnostic feature's actual depth – e.g. a gleyic feature confined to 50–100 cm yields `Endogleyic` instead of `Gleyic`. Default `FALSE` keeps the canonical names byte-identical. Surface / epipedon qualifiers are excluded (their depth is definitional).

gapfill — Opt-in within-pedon depth gap-fill, default FALSE (no-op, classification stays byte-identical). TRUE fills interior NA cells of the continuous depth-trending attributes by linear interpolation from the profile's own measured horizons; a character vector restricts it to those attributes; a named list is passed to gapfill_within_pedon. Filled cells carry inferred_prior provenance, so the evidence grade drops to "C". Runs on a deep copy – the caller's pedon is never mutated.

Value

A ClassificationResult.

Examples

```
pedon <- make_ferralsol_canonical()
res <- classify_wrb2022(pedon)
res$name
```

---

classify_za1991 *Classify a pedon under the South African soil classification (1991)*

Description

Keys a PedonRecord to the soil forms of the South African soil classification of 1991 (Soil Classification Working Group, 1991; the "Blue Book"). **Experimental**: the book itself is not available to soilKey; the key is read from its open pages (the AGIS key and form pages, the general rules, the topsoil pages), SASRI (1999) and van Huyssteen (2020). It keys the 73 forms and, since v0.9.359, their 404 families.

Usage

```
classify_za1991(
  pedon,
  rules = NULL,
  on_missing = c("warn", "silent", "error"),
  gapfill = FALSE
)
```

Arguments

pedon — A PedonRecord.

rules — Optional pre-loaded rule set (load_rules("za1991")).

on_missing — One of "warn" (default), "silent", "error".

gapfill — Opt-in within-pedon gap-fill, as in classify_usda.

Details

Each layer is read as one of the 30 diagnostic horizons and materials (see identify_horizons): from the name written in diagnostic_za (or in the designations of a profile declared with site$designation_system = "za"), else by soilKey's tests on the recorded data, three-valued, with WRB numbers standing in, flagged, where no open source gives the book's (the podzol B, the placic pan, the dorbank, the carbonate horizons and the colours of the subsoil). The profile is first read by the general rules: only horizons whose top lies above 1500 mm count; a recent deposit thinner than 500 mm over a buried soil is set aside and the buried soil classified, a thicker one is classified alone (recent_deposit, else its origin, flagged); a profile with no topsoil whose first horizon is a red or yellow-brown apedal B, a neocarbonate B, a dorbank or a hard plinthic B is read as truncated, an orthic A supplied.

A form holds when its horizons are the first of the profile's, in order (layers of no diagnostic horizon skipped), and the horizon that follows is not one that would make another form (Hutton over hard rock is Hutton; Hutton over a soft plinthic B is Bainsvlei). A description that stops above 1500 mm leaves the longer forms undecided, so the shorter one is provisional. When no form passes and some are undecided, the name lists them in the order of the key, provisional ("[Undecided in the South African soil classification (1991): Inanda ([Ia]), Arcadia (Ar), or not classified]", v0.9.402). The 500 mm rules of the form pages are tested too (Valsrivier, Sepane, Swartland, the podzol forms; Witbank, Dundee, Namib over a buried soil). A sequence no form has gives no form, with the nearest forms in the warnings; an orthic A under 100 mm thick over nothing diagnostic is the land class "rock and lithosols" (general rules). Lichtenburg (orthic A, red apedal B, hard plinthic B) is named by van Huyssteen (2020) but not by the AGIS key of 1991: it is keyed, always provisional. The readings soilKey makes where the open sources are silent are in system.file("rules", "za1991", "readings.yaml",package = "soilKey").

The families (v0.9.359) are read by the 19 family criteria of the book, each on the horizon its family names (the B1 of the red apedal B, the E, the upper G, the layer under the placic pan ...), as the form key identified it: the first family whose classes all hold is the family; with a criterion undecided for lack of data, the result names the families left possible and soilKey's partial code, an "x" for each undecided digit ("Hutton, [Hu1x00]"). The name follows the identification of the book: the form, the family code and name, and the USDA texture class of the A when its clay, silt and sand are recorded (with the sand grade of sand_grade): "Hutton, Hu1200 Kelvin, sandy loam".

Value

A ClassificationResult with system_id "za1991": the form in rsg_or_order, the full name (form, family, texture) in name, the form and the family in levels (with name_en, soilKey's gloss of the form: the sequence of its horizons), the family read in qualifiers$za_family (its code and name, the partial code of an undecided family, the texture, and every criterion read: its class, the layers read, the stand-ins and the attributes missing), the sequence read in qualifiers$za_sequence and the identified layers in qualifiers$za_horizons (with za_truncated, za_land_class and za_nearest when they apply), the general rules applied, the stand-ins used and the nearest forms in warnings, and provisional_because.

References

Soil Classification Working Group (1991). Soil classification: a taxonomic system for South Africa. Memoirs on the Agricultural Natural Resources of South Africa No. 15. Department of Agricultural Development, Pretoria.

See Also

make_za1991_example, identify_horizons, soilkey_systems.

Examples

```
res <- classify_za1991(make_za1991_example("hutton"), on_missing = "silent")
res$name
res$qualifiers$za_sequence
```

---

clear_kst13_cache *Clear the in-memory KST13 cache*

Description

Useful after pointing `options(soilKey.kst13_dir)` at other copies of the files mid-session. Frees ~3.1 MB.

Usage

```
clear_kst13_cache()
```

Value

`NULL`, invisibly. Called for its side effect of emptying the KST 13th-edition lookup cache.

---

clear_ossl_cache *Clear the soilKey OSSL cache*

Description

Removes the per-region cache files written by download_ossl_subset. Useful when a stale cache is suspected or when disk space is tight.

Usage

```
clear_ossl_cache(region = NULL, cache_dir = NULL, verbose = TRUE)
```

Arguments

region	Optional character vector of regions to clear; the default `NULL` clears every cached file under ‘tools::R_user_dir("soilKey", "cache")‘.
cache_dir	Cache directory (defaults to the soilKey user-cache dir).
verbose	If `TRUE`, prints which files were removed.

Value

Invisibly, the character vector of files that were removed.

`cohesic` *Cohesic horizon (WRB 2022, Chapter 3.1.7)*

Description

A subsurface horizon with a massive or weak subangular blocky structure, poor in organic matter and iron oxides, with a kaolinite-dominated, low-activity clay fraction, and hard when dry. It is the diagnostic basis of the *Cohesic* qualifier and corresponds to the *caracter coeso* of the Brazilian SiBCS (typical of the coastal "Tabuleiros" Latossolos and Argissolos).

Usage

```
cohesic(
  pedon,
  max_oc = 0.5,
  min_clay = 15,
  max_cec_per_clay = 24,
  min_thickness = 10
)
```

Arguments

`pedon`	A `PedonRecord`.
`max_oc`	Maximum soil organic carbon (%), default 0.5.
`min_clay`	Minimum clay (%), default 15.
`max_cec_per_clay`	Maximum CEC in cmol_c per kg clay, default 24.
`min_thickness`	Minimum thickness in cm, default 10.

Details

Diagnostic criteria (all required): mineral material with (1) < 0.5% soil organic carbon; (2) >= 15% clay; (3) a CEC (by 1 M NH4OAc, pH 7) of < 24 cmol_c per kg clay; (4) a massive or weak subangular blocky structure; (5) not cemented; (6) a rupture-resistance class, when dry, of at least hard; and (7) a thickness of >= 10 cm.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 3.1.7, Cohesic horizon.

combine_priors *Combine multiple spatial priors via weighted geometric mean*

Description

Given a list of priors (each a data.table with `rsg_code,probability`), pools them into a single distribution using a weighted geometric mean and renormalises to sum to 1.

Usage

```
combine_priors(priors, weights = NULL, epsilon = 1e-06)
```

Arguments

`priors`	A list of `data.tables` with columns `rsg_code` and `probability`.
`weights`	Optional non-negative numeric vector of length `length(priors)`. Defaults to equal weights. Will be renormalised to sum to 1.
`epsilon`	Smoothing floor for classes missing from a prior (default 1e-6). Must be > 0 – otherwise any class missing from a single prior is suppressed entirely.

Details

Geometric pooling has two desirable properties for soil-class priors:

1. externally Bayesian (the pooled posterior under any common likelihood matches what one would get by individual updates), and
2. zero-preserving: a class assigned probability 0 by any prior is suppressed in the pooled distribution. To avoid that, classes absent from a given prior are imputed with the smoothing constant `epsilon`.

Value

A `data.table` with columns `rsg_code`, `probability`, sorted by descending probability.

compare_engines *Side-by-side comparison of soilKey vs aqp diagnostic engines*

Description

Runs the soilKey hand-coded diagnostic and the aqp wrapper on the same pedon, returns both results plus an agreement flag. Useful for A/B benchmarks and for choosing which engine to use per dataset.

Usage

```
compare_engines(pedon, diagnostic = c("argic", "cambic"))
```

Arguments

pedon	A PedonRecord.
diagnostic	One of "argic" or "cambic".

Value

A list with soilkey, aqp, agree.

---

compare_smartsolos *Cross-validate the local SiBCS classifier against the SmartSolosExpert API*

Description

Runs both classify_sibcs (local) and classify_via_smartsolos_api (remote PROLOG via Embrapa AgroAPI) on the same PedonRecord and tabulates agreement at each of the four SiBCS categorical levels.

Usage

```
compare_smartsolos(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Forwarded to classify_via_smartsolos_api.

Value

A list with local and remote ClassificationResults plus a one-row agreement data.frame with columns ordem, subordem, gde_grupo, subgrupo, n_match.

Examples

```
## Not run:
Sys.setenv(AGROAPI_TOKEN = "<your token>")
cmp <- compare_smartsolos(make_argissolo_canonical())
cmp$agreement

## End(Not run)
```

compute_ki *Ki (silica:alumina molar) – SiBCS Cap 1, p 32*

Description

Calcula o indice molar Ki = SiO2 / Al2O3 a partir de teores percentuais por ataque sulfurico-NaOH (Embrapa Manual de Metodos). Massas molares: 60.08 (SiO2), 101.96 (Al2O3):

Usage

```
compute_ki(sio2_pct, al2o3_pct)
```

Arguments

sio2_pct	Teor de SiO2 por ataque sulfurico (%).
al2o3_pct	Teor de Al2O3 por ataque sulfurico (%).

Details

Ki (molar) = (% SiO2 / 60.08) / (% Al2O3 / 101.96) $\approx 1.6973 \times$ (% SiO2 / % Al2O3)

Value

Ki molar (numeric); NA se algum input for NA ou Al2O3 ≤ 0.

References

Embrapa (2018), SiBCS 5a ed., Cap 1, p 32; Embrapa Manual de Metodos de Analise de Solo (3a ed., 2017).

Examples

```
compute_ki(sio2_pct = 18, al2o3_pct = 20)  # ~1.53, abaixo do limite latossolico
```

compute_kr *Kr (silica:sesquioxidos molar) – SiBCS Cap 1, p 32*

Description

Calcula o indice molar Kr = SiO2 / (Al2O3 + Fe2O3) usando massas molares 60.08 (SiO2), 101.96 (Al2O3) e 159.69 (Fe2O3):

Usage

```
compute_kr(sio2_pct, al2o3_pct, fe2o3_pct)
```

Arguments

sio2_pct	Teor de SiO2 por ataque sulfurico (%).
al2o3_pct	Teor de Al2O3 por ataque sulfurico (%).
fe2o3_pct	Teor de Fe2O3 por ataque sulfurico (%).

Details

Kr (molar) = (% SiO2 / 60.08) / (% Al2O3 / 101.96 + % Fe2O3 / 159.69)

Value

Kr molar (numeric); NA se algum input for NA ou denominador ≤ 0.

References

Embrapa (2018), SiBCS 5a ed., Cap 1, p 32.

Examples

```
compute_kr(sio2_pct = 18, al2o3_pct = 20, fe2o3_pct = 12)
```

---

compute_per_attribute_evidence_grade

Per-attribute provenance-aware evidence grade

Description

Resolves the evidence grade of every (horizon, attribute) cell that carries a provenance entry. Where a cell has more than one entry (a value re-sourced over the profile's lifetime) the most authoritative source wins, mirroring PedonRecord's own authority order.

Usage

```
compute_per_attribute_evidence_grade(pedon)
```

Arguments

pedon	A PedonRecord.

Details

Grades: A measured, B predicted from spectra, C inferred from a spatial prior, D extracted by a vision-language model, E user-assumed.

Value

A data.table with columns horizon_idx, attribute and grade, sorted by horizon then attribute. A pedon with no provenance entries yields a zero-row table.

See Also

classify_from_photos, the global evidence grade reported on every ClassificationResult.

Examples

```
p <- make_ferralsol_canonical()
compute_per_attribute_evidence_grade(p)   # all-measured -> all grade A
```

---

continuous_rock *Continuous rock (WRB 2022 Ch 3.2.5)*

Description

Consolidated material below the soil. v0.3.3: detects via designation R or Cr on the lowermost (or any) layer.

Usage

```
continuous_rock(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

coverage_report *Honest taxonomic-completeness report*

Description

Measures, by NAME, exactly which canonical taxa/qualifiers the package's deterministic rule base registers, replacing hand-maintained coverage claims with an auditable, reproducible diff. For "usda_subgroup" the canonical reference is the Soil Taxonomy 13th-edition subgroup set from kst13_codes; for "wrb_qualifiers" it is the WRB 2022 principal + supplementary qualifier set from wrb2022_canonical.

Usage

```
coverage_report(
  system = c("usda_subgroup", "usda_great_group", "usda_suborder", "wrb_qualifiers",
    "wrb_horizons", "wrb_properties", "wrb_materials", "wrb_rsg", "sibcs"),
  write = FALSE,
  report_dir = NULL
)
```

Arguments

system	Which axis to measure. USDA taxon levels against the Soil Taxonomy 13th-edition code set (kst13_codes): "usda_subgroup" (default), "usda_great_group", "usda_suborder". WRB 2022 against wrb2022_canonical and the canonical diagnostic reference (inst/extdata/canonical/wrb2022_diagnostics.csv): "wrb_qualifiers", "wrb_horizons" (40 diagnostic horizons), "wrb_properties" (17 diagnostic properties), "wrb_materials" (19 diagnostic materials), and "wrb_rsg" (32 Reference Soil Groups, diffed against inst/rules/wrb2022/key.yaml). For the qualifier and diagnostic axes "covered" means the mapped function exists *and* is a genuine implementation (not an unconditional passed = NA stub); the inert ones are returned in $stubs. "sibcs" has no external canonical class list, so it honestly reports registered class counts per level only (no percentage).
write	If TRUE, also write a Markdown summary to report_dir. Default FALSE.
report_dir	Directory for the Markdown report when write = TRUE. Defaults to inst/benchmarks/reports inside the installed package.

Value

Invisibly, a list with $overall (one-row data frame: system, level, canonical_n, registered_n, covered_n, missing_n, pct), $by_group (per order, or per principal/supplementary), $missing (canonical names not registered), $extra (registered names absent from the canonical set), and – for "wrb_qualifiers" – $stubs (functions that exist but are inert). A compact summary is printed as a side effect.

Examples

```
# WRB diagnostic horizons are measured against soilKey's own table, offline.
cov <- coverage_report("wrb_horizons")
cov$overall

# The USDA axes read the KST 13th edition from ncss-tech/SoilKnowledgeBase,
# downloaded on first use, so they need a connection the first time.
# Needs network access; skipped offline and in batch checks.
online <- interactive() && requireNamespace("httr", quietly = TRUE) &&
  !inherits(try(httr::HEAD("https://raw.githubusercontent.com", httr::timeout(5)),
                silent = TRUE), "try-error")
if (online) {
  usda <- try(coverage_report("usda_subgroup"), silent = TRUE)
  if (!inherits(usda, "try-error")) head(usda$missing)
}
```

---

cryic_conditions *Cryic conditions (WRB 2022)*

Description

Tests whether continuous frozen / permafrost material occurs within the upper `max_top_cm`. Two alternative paths qualify per WRB 2022:

1. **Permafrost temperature**: a layer at top_cm <= `max_top_cm` (default 100) with `permafrost_temp_C` <= `max_temp_C` (default 0 C).
2. **Designation pattern**: a layer at top_cm <= `max_top_cm` with designation containing suffix `"f"` (frozen) or matching `"^Cf"` / `"perma"`. A site that declares the Canadian notation (`site$designation_system = "cssc"`) writes a frozen layer `z`; the New Zealand and Australian notations have no frozen suffix, so there only `"perma"` counts. Used as a fallback when the temperature field is not in the pedon (typical of legacy survey data).

Either path qualifies. Diagnostic of Cryosols. When neither passes and the site has a soil temperature regime warmer than gelic (cryic, frigid, mesic, thermic, hyperthermic or an iso form; `site$soil_temperature_regime`), the answer is no: those regimes have no permafrost (KST 13, Ch. 3, p. 50).

Usage

```
cryic_conditions(pedon, max_top_cm = 100, max_temp_C = 0)
```

Arguments

`pedon`	A `PedonRecord`.
`max_top_cm`	Maximum top depth (cm) (default 100).
`max_temp_C`	Maximum mean annual permafrost-zone temperature (deg C) for the temperature path (default 0).

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 5, Cryosols.

`cryic_horizon` *Cryic horizon (WRB 2022, Chapter 3.1.8)*

Description

A perennially frozen horizon in mineral or organic material (permafrost). The permafrost logic is shared with `cryic_conditions`; this exposes it under the canonical WRB 2022 horizon name so that `coverage_report("wrb_horizons")` reports it, and returns a `DiagnosticResult` named `"cryic"`.

Usage

```
cryic_horizon(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Passed to cryic_conditions (e.g. max_top_cm, max_temp_C).

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 3.1.8, Cryic horizon.

DiagnosticResult *DiagnosticResult: structured outcome of a diagnostic test*

Description

DiagnosticResult: structured outcome of a diagnostic test

DiagnosticResult: structured outcome of a diagnostic test

Details

Returned by every WRB or SiBCS diagnostic function (e.g. argic, ferralic, mollic). A DiagnosticResult never reduces to a bare TRUE/FALSE — it always carries (a) which layers satisfied the criteria, (b) the per-sub-test evidence, (c) which attributes would have been required but are missing, and (d) the literature reference for the diagnostic definition.

passed is TRUE/FALSE/NA; NA means the test could not be evaluated because critical attributes were missing. This three-valued semantics propagates through the rule engine — an indeterminate test does not silently fail.

Public fields

name Character. Name of the diagnostic (e.g. "argic").

passed Logical. TRUE, FALSE, or NA.

layers Integer vector. Indices of horizons that satisfy the diagnostic.

evidence Named list. Sub-test results, each itself a list with at least passed, layers, and missing.

missing Character vector. Attribute names that would have been needed but were NA.

reference Character. Literature citation for this diagnostic.

notes Character. Free-form notes (interpretation choices, edge cases hit).

Methods

Public methods:

- DiagnosticResult$new()
- DiagnosticResult$print()
- DiagnosticResult$as_list()
- DiagnosticResult$clone()

Method `new()`**:** Build a DiagnosticResult.

Usage:

```
DiagnosticResult$new(
  name,
  passed = NA,
  layers = integer(0),
  evidence = list(),
  missing = character(0),
  reference = NA_character_,
  notes = NA_character_
)
```

Arguments:

`name` Diagnostic name.

`passed` TRUE/FALSE/NA.

`layers` Integer vector of horizon indices that satisfied.

`evidence` Named list of sub-test results.

`missing` Character vector of missing attribute names.

`reference` Citation string.

`notes` Free-form notes.

Method `print()`**:** Pretty-print the result with sub-test breakdown.

Usage:

```
DiagnosticResult$print(...)
```

Arguments:

`...` Ignored (S3 print signature compatibility).

Method `as_list()`**:** Return the result as a plain list (for serialization).

Usage:

```
DiagnosticResult$as_list()
```

Method `clone()`**:** The objects of this class are cloneable with this method.

Usage:

```
DiagnosticResult$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

distrofico — *Solo distrofico (SiBCS Cap 1, p 30)*

Description

Negacao operacional de `eutrofico`: V < 50% no horizonte diagnostico subsuperficial.

Usage

```
distrofico(pedon, max_v = 50)
```

Arguments

pedon	A `PedonRecord`.
max_v	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

dolomitic_material — *Dolomitic material (WRB 2022 Ch 3.3.5): \>= 2% Mg-rich carbonate, CaCO3/MgCO3 < 1.5. v0.3.3: detects via designation pattern* `kdo|do|magn` *as proxy when ratio data missing.*

Description

Dolomitic material (WRB 2022 Ch 3.3.5): \>= 2% Mg-rich carbonate, CaCO3/MgCO3 < 1.5. v0.3.3: detects via designation pattern `kdo|do|magn` as proxy when ratio data missing.

Usage

```
dolomitic_material(pedon)
```

Arguments

pedon	A `PedonRecord`.

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`download_bdsolos` *Download the BDsolos consulta-publica CSV (experimental, requires chromote)*

Description

Drives the Embrapa BDsolos web form via headless Chrome (`chromote`) to produce a CSV of all profiles + all attributes. Marked **experimental**: heavy queries (no UF filter) frequently overload the Embrapa server. Prefer `filter_uf` = batches of one or two states at a time and stitch the resulting CSVs.

Usage

```
download_bdsolos(
  out_path,
  accept_terms = FALSE,
  filter_uf = NULL,
  attributes = "default",
  timeout_seconds = 600,
  chromote_session = NULL,
  verbose = TRUE
)
```

Arguments

`out_path`	File path for the downloaded CSV.
`accept_terms`	Logical. Must be `TRUE` to proceed; the function aborts otherwise. Documents informed consent to the BDsolos terms (personal/academic use, ABNT citation).
`filter_uf`	Optional 2-letter UF code (e.g. `"RJ"`, `"SC"`). Strongly recommended – the full-table query often times out.
`attributes`	Character vector. Which attribute groups to request. Defaults to the full SiBCS-classification-relevant set (Identificacao + Localizacao + Classificacao for Pontos de Amostragem, Identificacao + Morfologicas + Fisicas + Quimicas for Horizontes; Mineralogicas excluded for performance). Pass `"all"` to include Mineralogicas.
`timeout_seconds`	Total timeout for the AJAX query. Default 600 (10 min).
`chromote_session`	Optional pre-built `chromote::ChromoteSession`. Useful to share a session across calls.
`verbose`	If `TRUE` (default), prints progress.

Details

Per the Embrapa terms-of-use, the data is licensed for personal / academic use and publications must cite the source per ABNT. **Set** `accept_terms = TRUE` **to acknowledge this and let the function click "Concordo" on your behalf.**

Value

File path to the downloaded CSV (invisible).

See Also

load_bdsolos_csv, inspect_bdsolos_csv.

Examples

```
## Not run:
# Single UF (fast, recommended)
download_bdsolos("soil_data/bdsolos/RJ.csv",
                 accept_terms = TRUE,
                 filter_uf    = "RJ")

# Stitch multiple UFs
for (uf in c("RJ", "SP", "MG", "ES")) {
  download_bdsolos(file.path("soil_data/bdsolos",
                             paste0(uf, ".csv")),
                   accept_terms = TRUE, filter_uf = uf)
}

# Then load all of them
csvs <- list.files("soil_data/bdsolos", "\\.csv$", full.names = TRUE)
all_pedons <- unlist(lapply(csvs, load_bdsolos_csv), recursive = FALSE)
length(all_pedons)

## End(Not run)
```

---

download_extdata_cache

Download one or more soilKey lazy-fetch caches from GitHub Release

---

Description

soilKey ships four large benchmark caches (KSSL, KSSL+NASIS, AfSP, WoSIS stratified) that are too large to embed in the CRAN source tarball. Since v0.9.94 they are pinned to a versioned GitHub Release and downloaded on demand into the user cache directory at `tools::R_user_dir("soilKey", "data")`.

Usage

```
download_extdata_cache(
  which = "all",
  release = NULL,
  overwrite = FALSE,
  verbose = TRUE
)
```

Arguments

which	Character vector of cache names to download. "all" (default) downloads every lazy-fetch cache. Valid names: "afsp_sample", "kssl_sample", "kssl_nasis_sample", "wosis_stratified_sample".
release	Release tag to pull from. NULL (the default) means soilKey has nowhere to pull from: since v0.9.199 the package distributes no third-party soil observations, and this function refuses rather than serving a copy of somebody else's data. Supply a tag only if you maintain your own mirror, obtained from the providers under their terms.
overwrite	If TRUE, redownload even if the file is already present in the user cache (default FALSE).
verbose	Print progress (default TRUE).

Details

On first call to any of load_kssl_sample(), load_kssl_nasis_sample(), load_afsp_sample(), or load_wosis_stratified_sample(), soilKey checks for the file in the user cache. If missing, the loader prompts (interactive sessions only) to download. Use download_extdata_cache() to eagerly populate the cache without prompting.

Value

Invisibly, a named character vector of local paths to the downloaded files.

Examples

```
## Not run:
# Download every lazy-fetch cache once, ahead of any benchmark run:
download_extdata_cache()

# Or just the WRB AfSP sample:
download_extdata_cache("afsp_sample")

## End(Not run)
```

---

download_ossl_subset *Download an OSSL subset and return an 'ossl_library' artefact*

Description

Fetches a region-filtered subset of the Open Soil Spectral Library (Sanderman et al. 2024) and assembles it into the 'list(Xr, Yr, metadata)' shape consumed by predict_ossl_mbl and predict_ossl_plsr_local. The result is cached under 'tools::R_user_dir("soilKey", "cache")' so subsequent calls in the same session (or future R sessions) skip the network.

Usage

```
download_ossl_subset(
 region = c("global", "south_america", "north_america", "europe", "africa", "asia",
    "oceania"),
 properties = c("clay_pct", "sand_pct", "silt_pct", "cec_cmol", "bs_pct", "ph_h2o",
    "oc_pct", "fe_dcb_pct", "caco3_pct"),
  wavelengths = 350:2500,
  endpoint = NULL,
  cache_dir = NULL,
  force = FALSE,
  verbose = TRUE
)
```

Arguments

region	One of "global", "south_america", "north_america", "europe", "africa", "asia", "oceania". Filters the OSSL training rows by their site coordinates' continent.
properties	Character vector of OSSL property names to keep in 'Yr' (drops other reference columns to keep the artefact small). Defaults to the WRB-relevant set used by fill_from_spectra.
wavelengths	Integer vector of wavelengths (nm) the returned Xr matrix will be interpolated to. Defaults to Vis-NIR/SWIR (350-2500 nm at 1-nm resolution, 2151 columns).
endpoint	OSSL HTTP endpoint serving the JSON manifest; overrideable via options(soilKey.ossl_endpoint = ...) for testing or for using a private mirror. The default is the public Soil Spectroscopy GG bucket.
cache_dir	Cache directory; defaults to tools::R_user_dir("soilKey", "cache").
force	If TRUE, re-fetches even when a cached subset exists.
verbose	If TRUE, emits a 'cli' summary of the fetch.

Details

This function intentionally does **not** fall back to the synthetic predictor on network failure – a missing OSSL artefact is a real condition that the caller must handle, and silent fallback would make benchmarks meaningless.

Value

A list with elements Xr (numeric matrix, rows = training profiles, columns = wavelengths in nm), Yr (data.frame with the requested property columns, rows aligned to Xr), and metadata (snapshot date, region, n profiles, source URL, and the SHA-256 of the cache file). Pass it as the ossl_library argument to fill_from_spectra or predict_ossl_mbl.

References

Sanderman, J., Savage, K., Dangal, S.R.S., Duran, G., Rivard, C., Cardona, M.T., Sandzhieva, A., Aramian, A. & Safanelli, J.L. (2024). Soil Spectroscopy for Global Good – the Open Soil Spectral Library (OSSL). https://soilspectroscopy.org/.

download_ossl_subset_with_labels
Download an OSSL subset and attach WRB / SiBCS / USDA labels

Description

Fetches a region-filtered slice of the Open Soil Spectral Library via download_ossl_subset and post-joins WRB Reference Soil Group labels from WoSIS GraphQL by spatial nearest-neighbour. The resulting artefact has the canonical `list(Xr, Yr, metadata)` shape – with extra columns in `Yr`: `wrb_rsg`, `wrb_label_source`, `wrb_label_distance_km`, plus optionally `sibcs_ordem` and `usda_order` when `translate_systems = TRUE`.

Usage

```
download_ossl_subset_with_labels(
 region = c("global", "south_america", "north_america", "europe", "africa", "asia",
    "oceania"),
  max_distance_km = 5,
  wosis_endpoint = NULL,
  translate_systems = TRUE,
  max_to_label = Inf,
  verbose = TRUE,
  query_fn = NULL,
  ...
)
```

Arguments

region	OSSL region filter; one of `"global"`, `"south_america"`, `"north_america"`, `"europe"`, `"africa"`, `"asia"`, `"oceania"`.
max_distance_km	WoSIS spatial-join tolerance in kilometres (default 5). Profiles whose nearest WRB-labeled WoSIS neighbour is farther than this are left unlabeled.
wosis_endpoint	Override for the WoSIS GraphQL endpoint (default `getOption("soilKey.wosis_graphql")`). The canonical value is `"https://graphql.isric.org/wosis/graphql"`.
translate_systems	If `TRUE` (default), also adds `sibcs_ordem` and `usda_order` columns derived from the WRB label via the Schad (2023) Annex Table 1 / SiBCS 5ª ed. Annex A correspondence. Those translations are 1:N for some classes; we pick the most-common partner and tag rows where the translation is genuinely ambiguous.
max_to_label	Maximum number of profiles to query against WoSIS (default `Inf`). WoSIS throttles aggressive queries; cap this when running interactive demos.
verbose	Emit `cli` progress messages.

query_fn	Optional injection of the per-coordinate WoSIS query function. Default uses .query_nearest_wosis_wrb. Tests pass a stub here to exercise the join logic without network.
...	Forwarded to download_ossl_subset.

Value

A list with Xr (numeric matrix), Yr (data frame with the labels attached), and metadata (list with the OSSL fetch metadata + the join statistics: number of profiles labeled, average / max distance, WoSIS endpoint, snapshot date).

Why this function exists

OSSL stores Vis-NIR / MIR spectra and lab data but typically lacks WRB Reference Soil Group labels on most profiles (KSSL data is USDA-flavoured; non-US contributions are inconsistent). WoSIS, by contrast, archives ~228 000 profiles with WRB labels but no spectra. This function bridges the two so the user can run classify_by_spectral_neighbours on a real-data OSSL library without having to do the spatial join themselves.

Caveats and provenance

WRB labels obtained via spatial join are **weak labels**. The same physical location may have been classified differently across surveys (different WRB editions, different interpretations). Each row carries:

- wrb_label_source = "wosis_spatial_join": label inherited from a WoSIS neighbour within max_distance_km.
- wrb_label_distance_km: the distance to that neighbour (NA when no neighbour was found within tolerance).
- wrb_label_source = "ossl_native": label was already present in OSSL Yr (rare; preserved verbatim).
- wrb_label_source = "missing": no neighbour within tolerance; the row stays unlabeled and will be skipped downstream.

Treat the labels as priors, not ground truth.

See Also

download_ossl_subset, classify_by_spectral_neighbours.

Examples

```
## Not run:
# Real OSSL South-America subset with WRB labels:
lib <- download_ossl_subset_with_labels(
  region          = "south_america",
  max_distance_km = 10
)
table(lib$Yr$wrb_rsg, useNA = "always")
```

```
table(lib$Yr$wrb_label_source)

# Drop into the spectral analogy classifier:
res <- classify_by_spectral_neighbours(
  spectrum     = my_query_spectrum,
  ossl_library = lib,
  k            = 25,
  region       = list(lat = -22.7, lon = -43.7,
                      radius_km = 500)
)

## End(Not run)
```

`download_redape_dataset` *Download the curated Redape GeoTab dataset (Vaz et al 2023)*

Description

Enumerates the dataset via the Dataverse API and downloads all JSON profile files (the structured / interoperable format used by the curators) into `dest_dir`. Skips files already present unless `overwrite = TRUE`.

Usage

```
download_redape_dataset(
  dest_dir,
  dataset_doi = .REDAPE_GEOTAB_DOI,
  include_rtf = FALSE,
  overwrite = FALSE,
  verbose = TRUE
)
```

Arguments

`dest_dir`	Destination directory for the JSON files.
`dataset_doi`	DOI of the dataset (default: the Vaz 2023 dataset).
`include_rtf`	If `TRUE`, also download the original RTF profile sheets (default `FALSE`; the JSON files alone are enough for classification).
`overwrite`	If `TRUE`, re-download files that already exist locally.
`verbose`	Print progress (default `TRUE`).

Value

Character vector of paths to the downloaded files.

References

Vaz, G. J., Silva Jr, A. F., & Silva Neto, L. de F. da (2023). Brazilian soil data for taxonomic classification. Redape, V1. doi:10.48432/PYKKA7.

`duric_horizon` *Duric horizon (WRB 2022)*

Description

Tests for >= 10% volume of duripan nodules (Si-cemented) within a horizon at least 10 cm thick. Diagnostic of Durisols.

Usage

```
duric_horizon(pedon, min_thickness = 10, min_duripan_pct = 10)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness (cm; default 10 per WRB 2022).
`min_duripan_pct`	Minimum duripan volume % (default 10 per WRB 2022).

Value

A `DiagnosticResult`.

v0.3.1: thresholds aligned with WRB 2022 Ch 3.1.7 (10%, 10 cm) – previous v0.3 used 15%/15 cm. Petroduric (cemented continuous duripan) detection still deferred and will be added in v0.4.

References

IUSS Working Group WRB (2022), Chapter 3.1.7 – Duric horizon (p. 41).

`duripa` *Duripa (SiBCS Cap 2, p 74; v0.7)*

Description

Reuso de `duric_horizon` (WRB Ch 3.1): subsuperficial cimentado por silica, continuo ou em \>= 50% volume.

Usage

```
duripa(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Reserved for future arguments.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

ensure_horizon_schema *Coerce a horizons-like data.frame to the canonical schema*

Description

Adds any missing canonical columns as NAs of the right type and reorders canonical columns first. Extra user-supplied columns are preserved at the end. Coerces character values to numeric where the schema requires it.

Column names are recognised tolerantly before coercion: an exact canonical name always wins, but names differing only in capitalisation or separators (pH_H2O, PH.H2O -> ph_h2o) and a small table of common analytical abbreviations (Clay -> clay_pct, SOC -> oc_pct, CEC -> cec_cmol, ...) are mapped to their canonical form rather than silently ignored. See inst/ATTRIBUTES.md for the complete list of recognised names and accepted aliases.

Usage

```
ensure_horizon_schema(h)
```

Arguments

h	Input data.frame or data.table.

Value

A data.table with the canonical horizon columns present, in canonical order, with extra columns preserved at the end.

Examples

```
h <- ensure_horizon_schema(data.frame(top_cm = 0, bottom_cm = 20))
"designation" %in% names(h)
# capitalisation and common abbreviations are recognised:
h2 <- ensure_horizon_schema(data.frame(pH_H2O = 5.4, Clay = 32, SOC = 1.1))
h2$ph_h2o
```

eutrofico *Solo eutrofico (SiBCS Cap 1, p 30)*

Description

Returns TRUE se a saturacao por bases (V%) >= 50% no horizonte diagnostico subsuperficial (B ou C). 65% para A chernozemico.

Usage

```
eutrofico(pedon, min_v = 50)
```

Arguments

pedon	A PedonRecord.
min_v	Numeric threshold or option (see Details).

Details

Since v0.9.257, in a profile with a B horizon the base saturation is judged as the key and the definitions of the classes ask for it, "na maior parte dos primeiros 100 cm do horizonte B (inclusive BA)": it is high when the B layers with V of min_v or more make more than half of the thickness of that section, among the layers where V is recorded. The BC is not part of the section. Until v0.9.256 one layer of B or of C with a high V was enough. A profile with no B is read as before, in its C layers, or in all of its layers when it has no C either.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

evaluate_rsg_tests *Evaluate the test block of a single RSG*

Description

Given a parsed tests block from a YAML key entry, evaluates the appropriate combinator and returns a list with passed, evidence, missing, and (optionally) notes.

Usage

```
evaluate_rsg_tests(pedon, tests)
```

Arguments

pedon	A PedonRecord.
tests	A tests block from the YAML.

Value

A list summarising the test outcome.

---

extract_horizons_from_pdf

Extract horizons from a soil description PDF

Description

Reads a PDF (typically a soil survey chapter, field-sheet scan, or thesis appendix), prompts the configured VLM to extract horizon attributes against inst/schemas/horizon.json, and merges the result into pedon. Every extracted attribute is recorded with source = "extracted_vlm" and the model's reported confidence and verbatim source quote.

Usage

```
extract_horizons_from_pdf(
  pedon,
  pdf_path = NULL,
  provider,
  max_retries = 3L,
  overwrite = FALSE,
  prompt_name = "extract_horizons",
  schema_name = "horizon",
  pdf_text = NULL
)
```

Arguments

pedon	A PedonRecord to merge into. Mutated in place AND returned invisibly.
pdf_path	Path to the PDF file. Either pdf_path or pdf_text must be supplied.
provider	A chat provider from vlm_provider (or a MockVLMProvider for testing).
max_retries	Integer; how many times to re-prompt on validation failure. Default 3.
overwrite	If TRUE, lower-authority values are allowed to clobber higher-authority ones. Default FALSE.
prompt_name	Override the default prompt template ("extract_horizons").
schema_name	Override the default schema ("horizon").
pdf_text	Optional alternative to pdf_path: the already-extracted description text. Useful for smoke tests, unit tests without pdftools, and for already-OCR'd field-sheet text.

Details

The PedonRecord's authority order guarantees that values already tagged `"measured"` are never silently overwritten by VLM extraction unless `overwrite = TRUE`.

If the PDF is long (more than ~30,000 characters), it is chunked page-by-page and each page is sent independently. This is a conservative-but-simple strategy; for very long surveys callers should pre-chunk and call this function once per profile.

Value

Invisibly, the (mutated) `pedon`. Carries a `"vlm_extraction"` attribute with the parsed response, number of attempts, and number of provenance entries added.

Failure modes

- If `pdftools` is not installed -> error.
- If the PDF cannot be read -> error.
- If the VLM response fails JSON parse / schema validation after `max_retries + 1` attempts -> error from `validate_or_retry`.

`extract_munsell_from_photo` *Extract Munsell color from a profile photo*

Description

Sends the photo to a multimodal VLM with a prompt that asks the model to estimate Munsell hue / value / chroma per visible horizon (when a Munsell reference card is in frame). Recorded as `extracted_vlm` with the model's self-reported confidence; photos without a reference card should yield confidence below 0.5 per the prompt specification.

Usage

```
extract_munsell_from_photo(
  pedon,
  image_path,
  provider,
  max_retries = 3L,
  overwrite = FALSE,
  prompt_name = "extract_munsell_from_photo",
  schema_name = "munsell"
)
```

Arguments

pedon	A PedonRecord.
image_path	Path to the image file (JPG / PNG).
provider	A chat provider from vlm_provider (or a MockVLMProvider for testing).
max_retries	Integer; how many times to re-prompt on validation failure. Default 3.
overwrite	If TRUE, lower-authority values are allowed to clobber higher-authority ones. Default FALSE.
prompt_name	Override the default prompt template ("extract_horizons").
schema_name	Override the default schema ("munsell", a colour-only subset of the horizon schema: geometry, designation and Munsell only). Since v0.9.194 – the full horizon schema was injected into the prompt and all but those fields discarded, which on a token-metered endpoint cost roughly 2,500 tokens per call for attributes a photograph cannot show.

Details

Quantitative non-color attributes (clay %, CEC, pH, etc.) are **never** extracted from photos, by prompt-level instruction. If the model returns one anyway, it is silently dropped.

Value

Invisibly, the mutated pedon, with the photo added to pedon$images.

---

extract_site_from_fieldsheet

Extract site metadata from a field-sheet image

Description

Sends a photographed / scanned field sheet to a multimodal VLM and merges the extracted site-level metadata (lat, lon, elevation, parent material, land use, etc.) into pedon$site. Existing fields are preserved unless overwrite = TRUE; only NULL fields are filled.

Usage

```
extract_site_from_fieldsheet(
  pedon,
  image_path,
  provider,
  max_retries = 3L,
  overwrite = FALSE,
  prompt_name = "extract_site_metadata",
  schema_name = "site"
)
```

Arguments

pedon	A PedonRecord.
image_path	Path to the field-sheet image.
provider	A chat provider from vlm_provider (or a MockVLMProvider for testing).
max_retries	Integer; how many times to re-prompt on validation failure. Default 3.
overwrite	If TRUE, lower-authority values are allowed to clobber higher-authority ones. Default FALSE.
prompt_name	Override the default prompt template ("extract_horizons").
schema_name	Override the default schema ("horizon").

Value

Invisibly, the mutated pedon.

---

familia_mineralogia_argila_geral

Familia: mineralogia da fracao argila (geral, nao-Latossolos)

Description

Classifica a mineralogia da argila para Argissolos, Cambissolos, Plintossolos, Luvissolos, Nitossolos, Vertissolos, Chernossolos, Planossolos, Gleissolos quando ha informacao quantitativa de atividade da argila e/ou Ki/Kr. Cobre as classes nao endereçadas por familia_mineralogia_argila_latossolo:

- esmectitica: T_argila >= ta_threshold (default 27 cmolc/kg argila), indicando dominancia de argilas 2:1 expansivas (esmectita / vermiculita / micas hidratadas).
- caulinitica: Ki >= ki_caulinitico_min (default 0.75) e Kr >= kr_caulinitico_min (default 0.75), alem de T_argila < ta_threshold.
- oxidica: Kr < kr_caulinitico_min, indicando predominancia de oxihidrooxidos de Fe e Al.
- mista: nenhum dos outros gates fechou conclusivamente – evidencia heterogenea ou incompleta.

Quando os tres atributos (T_argila, Ki, Kr) estiverem ausentes, o resultado fica NULL e os atributos faltantes sao reportados.

Usage

```
familia_mineralogia_argila_geral(
  pedon,
  max_depth_cm = 200,
  ta_threshold = 27,
  ki_caulinitico_min = 0.75,
  kr_caulinitico_min = 0.75
)
```

Arguments

pedon	A PedonRecord.
max_depth_cm	Profundidade da secao de controle (default 200).
ta_threshold	Limite cmolc/kg argila para esmectitica (default 27).
ki_caulinitico_min	Limite Ki para caulinitica (default 0.75).
kr_caulinitico_min	Limite Kr para caulinitica vs oxidica (default 0.75).

Value

FamilyAttribute.

References

Embrapa (2018), SiBCS 5a ed., Cap 18, p 286-287.

FamilyAttribute *Classe S4-like para atributos de Familia (5o nivel SiBCS)*

Description

Classe S4-like para atributos de Familia (5o nivel SiBCS)

Classe S4-like para atributos de Familia (5o nivel SiBCS)

Details

Estrutura categorica (em vez de booleana) que representa um adjetivo composto da Familia. value eh o adjetivo atribuido (string) ou NULL quando a dimensao nao se aplica ou nao foi possivel determinar.

Public fields

name Nome da dimensao (e.g. "grupamento_textural").

value Adjetivo atribuido (e.g. "argilosa") ou NULL.

evidence Lista nomeada com valores intermediarios.

missing Vetor de colunas necessarias mas indisponiveis.

reference String com referencia bibliografica.

Methods

Public methods:

- FamilyAttribute$new()
- FamilyAttribute$print()
- FamilyAttribute$clone()

Method `new()`**:** Build a FamilyAttribute.

Usage:

```
FamilyAttribute$new(
  name,
  value = NULL,
  evidence = list(),
  missing = character(0),
  reference = ""
)
```

Arguments:

`name` Nome da dimensao (e.g. "grupamento_textural").

`value` Adjetivo atribuido (e.g. "argilosa") ou `NULL`.

`evidence` Lista nomeada com valores intermediarios.

`missing` Vetor de colunas necessarias mas indisponiveis.

`reference` String com referencia bibliografica.

Method `print()`**:** Pretty-print the attribute.

Usage:

```
FamilyAttribute$print(...)
```

Arguments:

`...` Ignored (S3 print signature compatibility).

Method `clone()`**:** The objects of this class are cloneable with this method.

Usage:

```
FamilyAttribute$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

`fao_crosswalk` *The correspondence between the FAO-Unesco legends of 1974 and 1988*

Description

The official correspondence of Appendix 4 of the Revised Legend (FAO 1988, "Main changes in the revised legend", pp. 113-126) between the 26 major soil groups and 106 soil units of the Legend of 1974 and the 28 major soil groupings and 153 soil units of 1988, with the three changes the appendix does not flag added in brackets (Calcaric Gleysols to Calcic Gleysols is a change of definition; the Plinthic Luvisols have no 1988 row; the Thionic Histosols are new). For display and for reading labelled data only: it never classifies (`classify_fao` keys each edition by its own text).

Usage

```
fao_crosswalk(code = NULL, from = c("1974", "1988"))
```

Arguments

code	Optional character vector of codes to look up, of the edition from (a group, grouping or unit: "Lo", "L"; "LXh", "LX"). NULL (default) for the whole table.
from	"1974" (default): one row per 1974 class; the 1988 codes it goes to. "1988": the rows whose 1988 side names the code (as a target or as one of soilKey's alternatives), and the rows of the 1988 classes with no 1974 class (relation "new").

Value

A data.frame with level ("group" or "unit"), code_1974, name_1974, code_1988 and name_1988 (several joined by ", "; empty when deleted), relation ("same", "renamed", "amended", "renamed_amended", "merged", "split", "deleted", "new"), also and also_when (soilKey's alternative 1988 targets read from a group comment of the appendix, such as the Lixisols for a low activity Luvisol, and when they apply), page (printed page of 1988), bracketed (TRUE for a row the appendix does not print) and note (the comment, paraphrased).

References

FAO (1988). FAO/Unesco Soil Map of the World, Revised Legend. World Soil Resources Report 60, FAO, Rome; ISRIC reprint 1997, Appendix 4.

See Also

classify_fao.

Examples

```
fao_crosswalk(c("Lo", "Gc", "Vp", "Y"))[, c("code_1974", "code_1988", "relation", "also")]
fao_crosswalk("LXh", from = "1988")[, c("code_1974", "code_1988", "relation")]
```

---

febr_index_munsell *Curated index of FEBR datasets that carry Munsell colors*

Description

Returns a data.frame listing FEBR dataset IDs that have at least one Munsell-related column populated in their camada table, with metadata: n_horizons, n_finite_munsell, coverage, column_pattern.

Usage

```
febr_index_munsell(min_coverage = 0.1, refresh = FALSE, verbose = TRUE)
```

Arguments

`min_coverage`	Drop datasets whose Munsell coverage (fraction of horizons with non-NA hue) is below this. Default 0.1.
`refresh`	Logical. If `TRUE`, re-scan FEBR over the network instead of using the bundled May-2026 cache.
`verbose`	If `TRUE` (default), prints a one-line summary.

Details

Backed by a precomputed cache shipped in `R/sysdata.rda` (`.FEBR_MUNSELL_INDEX`; results of the May 2026 scan over 249 datasets). On first call after install, returns the cache instantly. Pass `refresh = TRUE` to re-scan FEBR live (slow, network-dependent; updates the in-memory copy but does not modify the bundled cache).

Value

A `data.frame` sorted by `n_finite_munsell` descending.

See Also

`read_febr_pedons`.

`ferralic` *Ferralic horizon (WRB 2022)*

Description

Tests whether any horizon meets the ferralic horizon criteria. The ferralic horizon is a subsurface horizon resulting from long and intense weathering, characterized by very low cation exchange capacity per unit clay – the canonical "low-activity clay" signal that defines the Ferralsol RSG.

Usage

```
ferralic(
  pedon,
  min_thickness = 30,
  max_cec = NULL,
  engine = NULL,
  cec_inclusive = FALSE
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness in cm (default 30).
max_cec	CEC (1M NH4OAc, pH 7) per kg clay that a ferralic horizon must stay below (default NULL = 16 in soilkey engine, 20 in aqp engine; see engine).
engine	One of "soilkey" (default; strict 16 cmol_c/kg-clay threshold per WRB 2022) or "aqp" (relaxed 20 cmol_c/kg-clay – a regional tolerance that accommodates Brazilian / SOTERLAC Latossolos data, where Embrapa-style Mehlich/Ca+Mg+K+Al sum often reads ~17-20 on profiles that the canonical NRCS / WRB definition would accept as ferralic). NULL reads getOption("soilKey.diagnostic_engine"). The numeric threshold can also be overridden directly via options(soilKey.ferralic_max_cec = ...).
cec_inclusive	FALSE (default): the CEC must be below max_cec, as WRB 2022 writes it ("< 16 cmolc kg-1 clay"). TRUE accepts a value equal to the limit, as USDA's oxic horizon does ("16 cmol(+) or less"); oxic_usda uses it.

Details

Sub-tests called:

- test_ferralic_texture – texture sandy loam or finer.
- test_cec_per_clay – CEC / clay < 16 (or < 20 under engine = "aqp") cmol_c/kg clay. Until v0.9.212 a value equal to the limit passed; WRB 2022 Ch 3.1.10 criterion 3 is "< 16 cmolc kg-1 clay".
- test_ferralic_thickness – thickness >= 30 cm.

v0.3.1 alignment with WRB 2022 Ch 3.1.10 (p. 44): the older "ECEC <= 12 cmol_c/kg clay" gate was removed because it is not in the canonical text – only CEC (1M NH4OAc, pH 7) < 16 is required.

v0.9.67 regional tolerance: BDsolos RJ benchmark (n=722 perfis) showed 88/115 Latossolos failing the strict 16-cmol gate because Embrapa lab methodology often reads CEC at 17-20 on profiles that are unambiguously Latossolos by every other criterion. The engine = "aqp" threshold of 20 closes that gap without redefining the WRB threshold itself; users targeting strict WRB 2022 fidelity should keep engine = "soilkey".

The weatherable-mineral test (<= 10% by volume), water-dispersible-clay test, and stratification / rock-structure exclusions remain deferred (they need mineralogical data outside the canonical horizon schema) and are refinements rather than gates.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition. International Union of Soil Sciences, Vienna. Chapter 3.1.10 – Ferralic horizon (p. 44).

`ferralsol` *Ferralsol RSG gate (WRB 2022 Ch 4, p 110)*

Description

WRB-canonical: ferralic horizon \<= 150 cm AND no argic horizon starting above (or at the upper limit of) the ferralic, UNLESS the argic in its upper 30 cm or throughout has one or more of:

- < 10% water-dispersible clay; OR
- DeltapH (pH_KCl - pH_water) \>= 0; OR
- \>= 1.4% soil organic carbon.

v0.3.4 enforces all three exception paths. The DeltapH check uses `ph_kcl` and `ph_h2o`; the WDC check uses `water_dispersible_clay_pct` (introduced in v0.3.3 schema), a clay content in % of the fine earth.

Usage

```
ferralsol(pedon, strict = NULL)
```

Arguments

`pedon`	A `PedonRecord`.
`strict`	Logical or `NULL`. When `NULL` (default) it resolves via `getOption("soilKey.rsg_strict", FALSE)`. `TRUE` requires two of the three argic exception paths.

Details

v0.9.310: the ferralic horizon must start <= 150 cm from the mineral soil surface; an argic horizon counts when it starts above or at the upper limit of the ferralic horizon (not only when it lies wholly above it); and an exception must hold throughout the upper 30 cm of the argic horizon (or all of it when it is thinner). Data that do not decide an exception leave the gate undecided (NA).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

Tier-3 strict mode (v0.9.98)

When an argic horizon sits above the ferralic, the default gate keeps the profile as a Ferralsol if *any one* of the three exception paths (WDC \< 10%, DeltapH \>= 0, SOC \>= 1.4%) holds. With `strict = TRUE` the gate requires *at least two* of the three – a single weak indicator no longer rescues a profile with a translocated-clay argic from being keyed out of Ferralsols.

ferric *Ferric horizon (WRB 2022)*

Description

v0.9.309: the criteria of WRB 2022 Ch 3.1.11 (p. 45). The horizon consists of mineral material, is >= 15 cm thick and is not part of a petroplinthic, pisoplinthic or plinthic horizon, and has, in one or more subhorizons, (a) >= 15% of its area as coarse (> 20 mm) masses inside aggregates that are black or have a hue redder than 7.5YR and a chroma >= 5, moist, or (b) >= 5% concretions or nodules, at least weakly cemented, reddish or blackish, > 2 mm (`ferric_nodules_pct`). Until v0.9.308 soilKey read it from Fe-dithionite >= 5%, which is not a criterion of the text: a red Ferralsol with 6% Fe-dithionite and no nodules became Ferric (ISRIC's KE009). soilKey records no size for mottles, so criterion (a) is met only by nodules; mottles of the right colours leave it undecided, and a layer with no mottles and no nodules recorded as such does not meet it.

Usage

```
ferric(pedon, min_thickness = 15, min_nodules_pct = 5)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness (cm; default 15).
`min_nodules_pct`	Minimum nodules, % (default 5).

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 3.1.11, Ferric horizon.

fetch_aafc_normals *Climate normals of Agriculture and Agri-Food Canada at given points*

Description

Queries the "Derived Normal Climate Data" of Agriculture and Agri-Food Canada (30-year normals on a 5 km grid, Open Government Licence - Canada) at each point: the monthly mean air temperature of every month, and the monthly minimum and maximum temperature and the precipitation from April to October (the growing season of these normals). The values come from the AAFC map service (one request per variable and per thousand points); no file is downloaded. Needs the **httr** and **jsonlite** packages and an internet connection.

Usage

```
fetch_aafc_normals(lat, lon, period = "1961-1990", id = NULL, verbose = TRUE)
```

Arguments

lat, lon	Decimal degrees (WGS 84).
period	One of "1961-1990", "1971-2000", "1981-2010", "1991-2020".
id	Optional identifiers, returned as the first column.
verbose	Print progress.

Value

A data frame with id, lat, lon and one column per variable (ta_m01 to ta_m12, tn_m04 to tn_m10, tx_m04 to tx_m10, ac_m04 to ac_m10; degrees Celsius and mm), with the period and the source as attributes. Points outside the grid get NA.

References

Agriculture and Agri-Food Canada. Derived Normal Climate Data. Open Government Portal (dataset 3a060b8f-e662-4a60-b297-3bed859ffc8a), Open Government Licence - Canada.

See Also

npdb_soil_climate, npdb_scenario

Examples

```
## Not run:
nm <- fetch_aafc_normals(lat = 52.13, lon = -106.67)
npdb_soil_climate(nm)

## End(Not run)
```

---

fetch_natsoil *Download NatSoil sites and profiles (CSIRO)*

Description

Reads NatSoil, the national soil site database of CSIRO, through its public API and saves the JSON it returns in dir: first every public site (pages of page_size sites, without horizons, in dir/sites), then, for each site whose licence allows external use (and, by default, that carries an order of the Australian Soil Classification), its horizons, samples and laboratory results (one file per site in dir/profiles). Files already in dir are not downloaded again, so an interrupted run can be resumed. The data are licensed per site, most under CC BY 4.0; cite them as each site's citation field says (kept in site$citation by load_natsoil_pedons).

Usage

```
fetch_natsoil(
  dir,
  classified_only = TRUE,
  max_sites = Inf,
  page_size = 200L,
  pause = 0.3,
  verbose = TRUE
)
```

Arguments

dir	Folder for the downloaded files (created when missing).
classified_only	Download profiles only for sites with an ASC order.
max_sites	Maximum number of profiles to download in this call.
page_size	Sites per page in the site pass.
pause	Seconds to wait between requests.
verbose	Print progress.

Details

The whole site pass is about 150 requests and the profile pass one request per site (about 9,800 with an ASC order), with a pause between them.

Value

Invisibly, a data frame of the sites selected for the profile pass: `agency_code`, `proj_code`, `s_id`, the ASC order given (`asc_order`), `file` and whether it is on disk (`fetched`).

References

CSIRO. CSIRO National Soil Site Database (NatSoil). https://doi.org/10.25919/sg5w-ny87

See Also

`load_natsoil_pedons`, `benchmark_natsoil`.

Examples

```
## Not run:
fetch_natsoil("natsoil")
peds <- load_natsoil_pedons("natsoil")
benchmark_natsoil(peds)

## End(Not run)
```

`fetch_sali` *Download Queensland SALI sites, reports and laboratory data*

Description

Reads Queensland's Soil and Land Information system (SALI) from its public services and saves what they return in `dir`: the site layer (Queensland Spatial, MapServer layer 101, the sites with an Australian Soil Classification label and laboratory data, one JSON file per page in `dir/layer`), then for each site its report (the SALI4320 site listing report, a PDF, in `dir/pdf/<PROJECT>_<SITE>.pdf`) and its laboratory samples with their results from the SALI soil chemistry API (OData `Samples` with `SiteLabMethodResults`, in `dir/lab/<PROJECT>_<SITE>_samples.json`). `read_sali_pedons` reads the folder.

Usage

```
fetch_sali(
  dir,
  sites = NULL,
  cap_bytes = 150000,
  workers = 1L,
  pause = 1,
  max_sites = Inf,
  verbose = TRUE
)
```

Arguments

`dir`	Folder for the downloaded files (created when missing).
`sites`	Optional sites to download: a data frame with `project_code` and `site_id`, or a character vector of `"<PROJECT>_<SITE>"`. By default every site of the layer whose label names an ASC order (the layer's "Unclassified" sites, which `read_sali_pedons` leaves out, are not downloaded).
`cap_bytes`	Largest report downloaded, in bytes (150 KB).
`workers`	Number of processes downloading at once (`parallel::mclapply`, each taking every `workers`-th site; always 1 on Windows).
`pause`	Seconds to wait after each request (in each process).
`max_sites`	Maximum number of sites downloaded in this call.
`verbose`	Print progress.

Details

Some reports embed full-size site photographs (one was over 50 MB): a report is downloaded only up to `cap_bytes`; a larger one is aborted as soon as it passes the cap, marked by a file `<PROJECT>_<SITE>.pdf.oversize` and not tried again. Files already in `dir` (and oversize marks) are not downloaded again, so an interrupted run can be resumed; a failed request is tried again on the

next run. Each request waits `pause` seconds, retries with a growing backoff, and the run stops (an error) on any page that asks for a login, a key or the acceptance of terms, or that a web application firewall rejected. The report host began refusing requests after some 14,000 of them (2 processes, a 1 s pause), so fetch in batches (`max_sites`) rather than all at once. Requests go only to the layer of Queensland Spatial and, per site, to the report host and the soil chemistry API; the report address is built from the project code and site number, not taken from the data.

The data are licensed under CC BY 4.0; cite them as `sali_citation` says.

Value

Invisibly, a data frame of the sites selected: `project_code`, `site_id`, `asc_label`, `stem`, `pdf` (`"ok"`, `"oversize"`, `"failed"` or `NA` when not attempted in this call) and `lab` (`"ok"`, `"failed"` or `NA`).

References

Queensland Government. Soil and Land Information (SALI). https://www.qld.gov.au/environment/land/soil/soil-data

See Also

`read_sali_pedons`, `benchmark_sali`, `sali_citation`.

Examples

```
## Not run:
fetch_sali("sali", max_sites = 20)
peds <- read_sali_pedons("sali")
benchmark_sali(peds)

## End(Not run)
```

---

fibrico *Material organico fibrico (SiBCS Cap 14)*

Description

Material organico pouco decomposto: >= 40% de fibras esfregadas OU indice de von Post H1-H4. Discrimina Organossolos Fibricos no 3o nivel.

Usage

```
fibrico(pedon)
```

Arguments

pedon	A `PedonRecord`.

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 14 (Organossolos), pp 224-226.

`fill_from_spectra` *Fill missing soil attributes from spectra via OSSL*

Description

Given a PedonRecord carrying a `spectra$vnir` matrix (rows = horizons, columns = wavelengths in nm), pre-processes the spectra, predicts the requested soil properties using the chosen OSSL-backed method, and writes the predictions into the pedon's horizons table via `pedon$add_measurement(..., source = "predicted_spectra")`. Each call updates the pedon's provenance log so that downstream classification can derive an evidence grade.

Usage

```
fill_from_spectra(
  pedon,
  library = "ossl",
  region = c("global", "south_america", "north_america", "europe", "africa"),
 properties = c("clay_pct", "sand_pct", "silt_pct", "cec_cmol", "bs_pct", "ph_h2o",
    "oc_pct", "fe_dcb_pct", "caco3_pct"),
  method = c("mbl", "plsr_local", "pretrained"),
  preprocess = "snv+sg1",
  k_neighbors = 100L,
  overwrite = FALSE,
  ossl_library = getOption("soilKey.ossl_library"),
  ossl_models = getOption("soilKey.ossl_models"),
  verbose = TRUE
)
```

Arguments

`pedon`	A PedonRecord with a `spectra$vnir` matrix.
`library`	Currently only `"ossl"` is supported.
`region`	One of `"global"`, `"south_america"`, `"north_america"`, `"europe"`, `"africa"`. Used to subset the OSSL training data when supported by the underlying backend.
`properties`	Character vector of OSSL-supported property names to predict. Default covers the most-requested WRB/SiBCS-relevant attributes.
`method`	One of `"mbl"`, `"plsr_local"`, `"pretrained"`.

preprocess	Pre-processing pipeline; passed to preprocess_spectra.
k_neighbors	Number of neighbours for memory-based methods.
overwrite	If FALSE (default), only fill cells whose existing provenance is weaker than predicted_spectra.
ossl_library	Optional OSSL library object (see predict_ossl_mbl). Defaults to getOption("soilKey.ossl_librar so a library set once reaches every call, including gap-fill from classify_all(gapfill = list(method = "spectra")). Without one the predictions are synthetic placeholders.
ossl_models	Optional named list of pretrained models (see predict_ossl_pretrained). Defaults to getOption("soilKey.ossl_models").
verbose	If TRUE, prints a cli summary.

Details

By default, predicted values do **not** overwrite measured values (the add_measurement() authority logic protects them). Setting overwrite = TRUE forces overwrite of any non-measured value.

Value

The mutated pedon, invisibly. Provenance entries with source = "predicted_spectra" are added per (horizon, property).

See Also

preprocess_spectra, predict_ossl_mbl, predict_ossl_plsr_local, predict_ossl_pretrained, pi_to_confidence.

---

fill_munsell_from_spectra

Fill missing Munsell colors on a PedonRecord from Vis-NIR spectra

---

Description

High-level helper that runs predict_munsell_from_spectra per horizon over the Vis-NIR spectra in pedon$spectra$vnir and writes the resulting hue / value / chroma back to the matching horizon rows via pedon$add_measurement(..., source = "predicted_spectra").

Usage

```
fill_munsell_from_spectra(pedon, overwrite = FALSE, verbose = TRUE)
```

Arguments

pedon	A PedonRecord that has $spectra$vnir populated (rows = horizons, cols = wavelengths).
overwrite	If TRUE, overwrite existing Munsell measurements. Default FALSE (only fills horizons whose Munsell is currently NA).
verbose	If TRUE (default), prints a per-horizon summary.

Details

This is the operational answer to the v0.9.35 Argissolo color confusion: when surveyor Munsell colors are missing and the user has Vis-NIR (e.g. from OSSL), call this helper, then re-run `classify_sibcs` – the v0.9.45 "color-undetermined" fallback will lift, and the classification will descend to subordem / grande grupo / subgrupo with proper `evidence_grade`.

Value

The pedon, invisibly. Provenance entries with `source = "predicted_spectra"` are appended.

`fluvic_material` *Fluvic material (WRB 2022)*

Description

Tests whether the profile shows fluvic material features: alternating textures across consecutive horizons within the upper 100 cm AND an irregular (non-monotone) organic carbon pattern with depth. Diagnostic of Fluvisols.

Usage

```
fluvic_material(pedon, max_top_cm = 100, min_clay_swing = 8)
```

Arguments

`pedon`	A `PedonRecord`.
`max_top_cm`	Maximum top depth (cm) considered (default 100).
`min_clay_swing`	Minimum absolute clay-percent change between consecutive layers required to count as alternation (default 8 percentage points).

Details

Sub-test: `test_fluvic_stratification`.

v0.3 limitations: WRB 2022 fluvic material also requires age (typically <100 years for sediment freshness), which v0.3 does not check (no temporal fields in the schema). The stratification proxy is conservative – truly heterogeneous floodplain profiles with dramatic texture swings will pass; subtle alluvial sequences may miss. v0.4 will refine.

Value

A `DiagnosticResult`.

v0.9.314

Fluvic material follows WRB 2022 Ch 3.3.6 (p. 87) layer by layer: fluviatile, marine or lacustrine origin (`rock_origin`, `layer_origin`, else the site's parent material); strata that are obvious (the clay reversal above, or a sand reversal of $\geq$ 15%) OR evidenced by two or more layers with more SOC than the layer directly above ($\geq$ 0.2%, $\geq$ 25% relative and $\geq$ 0.2% absolute, not of a natric or spodic horizon); and a single grain, massive, platy or weak subangular blocky structure (granular or subangular blocky in a layer of higher SOC). The layers returned are those that meet the three. An origin, a texture or a structure that is not recorded leaves the result undecided. The SiBCS carater fluvico keeps the earlier reading.

References

IUSS Working Group WRB (2022), Chapter 3, Fluvic material.

`folic` *Folic horizon (WRB 2022, Chapter 3.1.12)*

Description

A surface horizon of **well-aerated** organic material, i.e. organic material that is *not* water-saturated for long periods. It is the aerobic twin of the `histic_horizon`: both consist of organic material, but the histic horizon is water-saturated for >= 30 consecutive days in most years while the folic horizon is saturated for < 30 days and is not drained.

Usage

```
folic(pedon, min_oc = 20, max_saturation_days = 30, min_thickness = 10)
```

Arguments

`pedon`	A `PedonRecord`.
`min_oc`	Minimum soil organic carbon (%) for organic material, default 20 (WRB 2022 Ch 3.3.13; it was 12 until v0.9.312).
`max_saturation_days`	Maximum consecutive water-saturation days that still count as well-aerated, default 30.
`min_thickness`	Minimum thickness in cm, default 10.

Details

Diagnostic criteria (all required): organic material that (1) is saturated with water for < 30 consecutive days in most years and is not drained; and (2) has a thickness of >= 10 cm.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 3.1.12, Folic horizon.

`format_wrb_name` *Format a WRB 2022 soil name with qualifiers*

Description

Format a WRB 2022 soil name with qualifiers

Usage

```
format_wrb_name(
  rsg_name,
  principal = character(0),
  supplementary = character(0)
)
```

Arguments

`rsg_name`	Full RSG name (e.g. "Ferralsols").
`principal`	Character vector of principal-qualifier names, in the order they are written (as `resolve_wrb_qualifiers` returns them).
`supplementary`	Character vector of supplementary-qualifier names, in the order they are written.

Value

The name as WRB 2022 Chapter 2.2 writes it, e.g. "Geric Rhodic Ferralsol (Clayic, Eutric, Ferric, Humic)".

`fragic` *Fragic horizon (WRB 2022): a high-bulk-density horizon with restricted rooting. v0.3.3: detects via bulk_density_g_cm3 >= 1.65 AND structure grade massive/very firm OR designation pattern* `x/Bx`.

Description

Fragic horizon (WRB 2022): a high-bulk-density horizon with restricted rooting. v0.3.3: detects via bulk_density_g_cm3 >= 1.65 AND structure grade massive/very firm OR designation pattern `x/Bx`.

Usage

```
fragic(pedon, min_thickness = 15, min_bd = 1.65)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_bd	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

fragipa *Fragipa (SiBCS Cap 2, p 73-74; v0.7)*

Description

Reuso de fragic (WRB v0.3.3): horizonte subsuperficial endurecido quando seco, baixa MO, BD elevada, quebradicidade.

Usage

```
fragipa(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Reserved for future arguments.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

from_aqp *Convert an aqp SoilProfileCollection back to a list of PedonRecord*

Description

Inverse of as_aqp. Walks each profile in the SPC, renames aqp's canonical horizon column names back to soilKey's (top -> top_cm, name -> designation, clay -> clay_pct, ...), assembles a PedonRecord per profile, and returns the list.

Usage

```
from_aqp(spc)
```

Arguments

spc	A aqp::SoilProfileCollection.

Details

Round-trip property: from_aqp(as_aqp(pedon)) reproduces pedon modulo column ordering.

Value

A list of PedonRecord objects (length = length(spc)).

See Also

as_aqp, the forward conversion.

Examples

```
## Not run:
pedons <- list(make_ferralsol_canonical(), make_luvisol_canonical())
spc <- as_aqp(pedons)
pedons2 <- from_aqp(spc)
identical(pedons[[1]]$horizons$clay_pct, pedons2[[1]]$horizons$clay_pct)
#> [1] TRUE

## End(Not run)
```

---

gapfill_by_predicted_taxon

Fill missing horizon attributes from the predicted taxon's mean profile

---

Description

Classifies pedon with NO fill to get a provisional taxon, then fills its missing cells from taxon_profiles[[<that taxon>]] (built by build_taxon_profiles). Non-circular: the fill is keyed on the model's own prediction, not the reference. Each fill is written with source = "inferred_prior" (grade C). Reachable via gapfill = list(method = "taxon", taxon_profiles = <...>).

Usage

```
gapfill_by_predicted_taxon(
  pedon,
  taxon_profiles,
  system = c("sibcs", "wrb2022", "usda"),
  attrs = NULL,
  confidence = 0.55
)
```

Arguments

pedon	A PedonRecord.
taxon_profiles	Output of build_taxon_profiles.
system	One of "sibcs" (default), "wrb2022", "usda".
attrs	Attributes to fill (default: those present in the matched profile).
confidence	Provenance confidence (default 0.55, below a coordinate prior).

Value

Invisibly, the mutated pedon; attribute "gapfill_by_predicted_taxon" records the taxon + cells filled.

See Also

build_taxon_profiles, apply_soilgrids_depth_prior

---

gapfill_derive_horizon

Fill horizon attributes derivable BY DEFINITION from the same horizon

---

Description

Recovers cells that are exact closures of other measured columns in the same horizon (not statistical estimates): the texture third (clay/silt/sand) when the other two are present and sum to \< 100; effective CEC as sum(bases) + al; aluminium saturation as 100 * al / ecec; and base saturation as 100 * sum(bases) / cec. Every fill is written with source = "inferred_prior" so the PedonRecord authority order keeps it from displacing a measured value and the evidence grade drops to "C". Companion to gapfill_within_pedon (depth interpolation) and apply_soilgrids_depth_prior (external prior); reachable via the gapfill = list(method = "derive") argument of the classifiers.

Usage

```
gapfill_derive_horizon(pedon, overwrite = FALSE)
```

Arguments

pedon	A PedonRecord.
overwrite	If FALSE (default) only NA target cells are filled.

Value

Invisibly, the mutated pedon; attribute "gapfill_derive_horizon" records the count filled.

See Also

gapfill_within_pedon, apply_soilgrids_depth_prior

gapfill_within_pedon *Fill interior missing horizon attributes by within-pedon depth interpolation*

Description

For each requested attribute, builds a depth profile from the horizons in which that attribute is *measured* (non-NA) and linearly interpolates the value at the mid-depth of every horizon where it is missing – but only for horizons whose mid-depth falls strictly between the shallowest and deepest measured layer. Cells above the top or below the bottom measured layer are left NA: the function interpolates, it never extrapolates. Each fill is written with `source = "inferred_prior"`, so the PedonRecord authority order keeps it from displacing a measured, spectra-predicted or VLM-extracted value, and any downstream `compute_evidence_grade` call reports grade `"C"`.

Usage

```
gapfill_within_pedon(pedon, attrs = NULL, confidence = 0.6, overwrite = FALSE)
```

Arguments

pedon	A PedonRecord with at least two horizons.
attrs	Character vector of horizon columns to fill. Defaults to the continuous depth-trending attributes a linear interpolation can reasonably estimate (clay/silt/sand, pH, organic carbon, CEC/ECEC, base/aluminium saturation, bulk density).
confidence	Numeric in \[0, 1\] recorded as the provenance confidence of each interpolated cell. Defaults to `0.6` – below a measured value but anchored on the profile's own data, hence above the `0.5` used for an external SoilGrids prior.
overwrite	If FALSE (default) only NA cells are filled. If TRUE, non-measured cells are re-interpolated (measured cells are still never overwritten, and the provenance authority order is always respected).

Details

This is the within-pedon companion to apply_soilgrids_depth_prior (which fills from an external SoilGrids profile rather than from the profile's own measured layers). It is the mechanism behind the opt-in `gapfill` argument of classify_wrb2022, classify_sibcs, classify_usda and classify_all.

Note that this mutates pedon in place (as `apply_soilgrids_depth_prior` does). The `gapfill` argument of the classifiers operates on a deep copy instead, so a classification call never alters the caller's pedon.

Value

Invisibly, the mutated pedon. An attribute `"gapfill_within_pedon"` on the return value records how many cells were filled and for which attributes.

See Also

apply_soilgrids_depth_prior, classify_all

Examples

```
h <- data.frame(
  top_cm    = c(0, 20, 40, 60),
  bottom_cm = c(20, 40, 60, 90),
  clay_pct  = c(15, NA, 35, 40)
)
p <- PedonRecord$new(horizons = h)
gapfill_within_pedon(p, attrs = "clay_pct")
p$horizons$clay_pct   # second horizon filled to ~25 by interpolation
```

---

get_perturbation_scale

Monte-Carlo perturbation scale for an evidence grade

Description

Returns the noise magnitudes used by classify_with_uncertainty for a cell of the given evidence grade. A measurement (grade A) is perturbed only slightly; a user-assumed value (grade E) is perturbed heavily, reflecting how little is actually known about it.

Usage

```
get_perturbation_scale(grade = c("A", "B", "C", "D", "E"))
```

Arguments

grade	One of "A" (measured), "B" (spectra-predicted), "C" (prior-inferred), "D" (VLM-extracted) or "E" (user-assumed).

Value

A list with three elements: pct (the half-width of the multiplicative perturbation, applied to most numeric attributes), ph_abs (the half-width of the additive perturbation applied to pH columns) and munsell_abs (the additive half-width for Munsell value / chroma columns).

Examples

```
get_perturbation_scale("A")$pct   # 0.03 -- measured values barely move
get_perturbation_scale("E")$pct   # 0.30 -- assumptions move a lot
```

`gleyic_properties` *Gleyic properties (WRB 2022)*

Description

Tests whether the profile shows gleyic properties (WRB 2022 Ch 3.2.6, p. 75) in a layer starting within the upper `max_top_cm` (default 50). Gleyic properties qualify the Gleysols and many other RSGs (Gleyic, Endogleyic, Epigleyic qualifiers).

Usage

```
gleyic_properties(
  pedon,
  max_top_cm = 50,
  min_redox_pct = 5,
  stagnic_decay_factor = 3
)
```

Arguments

`pedon`	A `PedonRecord`.
`max_top_cm`	Maximum top depth (cm) of a candidate layer (default 50).
`min_redox_pct`	The oximorphic features of criterion 2 are more than this % of the layer (default 5).
`stagnic_decay_factor`	Not used since v0.9.312 (the decay test was removed); kept so that existing calls still work.

Details

Since v0.9.312 the layer reading follows the text and Hugo's reading of it for field descriptions that do not say where the mottles sit: a layer has gleyic properties when its matrix has a reductimorphic colour, moist (criterion 1: hue N, 10Y, GY, G, BG, B or PB, or hue 2.5Y or 5Y with a chroma of 2 or less; with more than 5% oximorphic features these are read as the gleyic features of criterion 2 over a reduced matrix); when it has more than 5% oximorphic features and lies directly on a layer with such a matrix (criterion 3); or when it has them and the site records groundwater within 150 cm that is not perched (`site$groundwater_kind`, `site$groundwater_depth_cm`). A mottled layer with none of these is `NA`, unless the site records no groundwater, a perched water table or a deeper one, or a claric layer lies directly above it: its features are then stagnic. Continuous rock has no gleyic properties. Until v0.9.311 the features also had to continue with depth (a "decay" test that is not in WRB 2022 and failed the textbook oximorphic-over-reductimorphic sequence).

Value

A `DiagnosticResult`.

v0.9.72 designation morphological inference (opt-in)

Field-described Brazilian Gleissolos profiles (e.g.\ the Embrapa Redape curated dataset) routinely encode gleyic properties via the designation suffix `g` (e.g.\ `Cg`, `Cg1`, `Cgn`, `Apg`) plus low-chroma Munsell colours (chroma \<= 2), without recording `redoximorphic_features_pct` as a numeric percent.

With `options(soilKey.gleyic_designation_inference = TRUE)` the function accepts a layer as gleyic when the reading above does not pass, the designation matches `[A-Z][a-z0-9]*g` and `munsell_chroma_moist <= 2` (or no Munsell colour). Default is `FALSE`.

References

IUSS Working Group WRB (2022), Chapter 3.2.6, Gleyic properties.

`gleysol` *Gleysol RSG gate (WRB 2022 Ch 4, p 103)*

Description

WRB-canonical (multi-path):

1. Layer \>= 25 cm starting \<= 40 cm with gleyic properties throughout AND reducing conditions in some parts of every sublayer; OR
2. Mollic/umbric > 40 cm thick with reducing conditions some parts of every sublayer 40 cm below mineral surface to lower limit, AND directly underneath a layer \>= 10 cm with lower limit \>= 65 cm having gleyic properties + reducing conditions; OR
3. Permanent saturation by water \<= 40 cm.

Path 1 reads the layers of `gleyic_properties` (since v0.9.312: the matrix colour, the layer underneath and the site's groundwater, without the old decay-with-depth test) and of `reducing_conditions`; since v0.9.312 the 25 cm may be made of adjoining sublayers (Bg1 over Bg2), counted from the mineral soil surface, and a layer whose redox reading is undecided leaves the gate `NA` when it could complete the 25 cm. Path 3 is read from the designation (W / saturated marker).

Usage

```
gleysol(pedon, strict = NULL)
```

Arguments

`pedon`	A `PedonRecord`.
`strict`	Logical or `NULL`. When `NULL` (default) it resolves via `getOption("soilKey.rsg_strict", FALSE)`. `TRUE` tightens path 1 and disables path 3.

Details

Path 2 (since v0.9.315; until then it was not read) takes the horizon of `mollic` or `umbric_horizon`: more than 40 cm of its layers, reducing conditions in every layer of it that reaches below 40 cm from the mineral soil surface, and directly underneath it (top within 1 cm of its lower limit) adjoining layers, gleyic with reducing conditions as in path 1, at least 10 cm thick and reaching 65 cm or more from the mineral soil surface. Where layers below the horizon's sure layers are undecided, its lower limit is undecided too, and every lower limit it could have is tried: the path is then `NA` at most. A horizon with no layer recorded underneath it does not meet item 2.b.

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

Tier-3 strict mode (v0.9.98)

With `strict = TRUE` the path-1 gleyic+reducing layer must start within the upper 25 cm (instead of 40 cm), and the path-3 designation-only fallback (a "W" / aquic marker) is disabled: strict mode requires measured gleyic and reducing evidence.

`grapes-or-or-grapes` *Default-value-for-NULL operator*

Description

Returns the left-hand side if it is non-NULL, otherwise the right-hand side. Re-exported so that downstream code can use the same idiom soilKey itself uses internally.

Usage

```
a %||% b
```

Arguments

`a`	The candidate value.
`b`	The fallback used when `a` is NULL.

Value

Either `a` or `b`.

GSM_DEPTHS *Default GlobalSoilMap depth intervals (cm)*

Description

GSM standard per Arrouays et al. (2014) "GlobalSoilMap: Toward a fine-resolution global grid of soil properties". Boundaries: 0-5, 5-15, 15-30, 30-60, 60-100, 100-200 cm.

Usage

```
GSM_DEPTHS
```

Format

An object of class `numeric` of length 7.

gypsic *Gypsic horizon (WRB 2022)*

Description

Tests whether any horizon meets the gypsic horizon criteria. The gypsic horizon is a horizon of secondary gypsum accumulation, diagnostic for Gypsisols.

Usage

```
gypsic(pedon, min_thickness = 15, min_gypsum_pct = 5)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness in cm (default 15).
`min_gypsum_pct`	Minimum gypsum percent in fine earth (default 5).

Details

Sub-tests called:

- `test_caso4_concentration` – gypsum >= 5%.
- `test_minimum_thickness` – thickness >= 15 cm.

v0.2 limitations: the WRB rule that gypsum content must exceed the underlying horizon by 1% (absolute) is not enforced. Petrogypsic (cemented) horizons are not yet detected. Both deferred to v0.3.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition. International Union of Soil Sciences, Vienna. Chapter 3 – Gypsic horizon.

---

gypsiric_material *Gypsiric material (WRB 2022 Ch 3.3.7): \>= 5% gypsum that is primary (not secondary). Without a "secondary fraction" schema column, v0.3.3 treats any layer with caso4_pct >= 5 as gypsiric unless it explicitly carries gypsic-horizon designation.*

Description

Gypsiric material (WRB 2022 Ch 3.3.7): \>= 5% gypsum that is primary (not secondary). Without a "secondary fraction" schema column, v0.3.3 treats any layer with caso4_pct >= 5 as gypsiric unless it explicitly carries gypsic-horizon designation.

Usage

```
gypsiric_material(pedon, min_caso4_pct = 5)
```

Arguments

pedon	A PedonRecord.
min_caso4_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

harmonize_to_gsm *Harmonise pedons to GlobalSoilMap depth intervals*

Description

Runs `mpspline2::mpspline_tidy()` on each requested numeric horizon attribute, producing a new PedonRecord per input pedon whose horizons table covers the canonical GSM intervals (GSM_DEPTHS). Categorical attributes (designation, Munsell hue) are propagated by mode-over-depth-overlap.

Usage

```
harmonize_to_gsm(
  pedons,
 attributes = c("clay_pct", "silt_pct", "sand_pct", "ph_h2o", "oc_pct", "cec_cmol",
    "base_saturation_pct", "munsell_value_moist", "munsell_chroma_moist",
    "redoximorphic_features_pct"),
  depths = GSM_DEPTHS,
  lam = 0.1,
  verbose = TRUE
)
```

Arguments

pedons	A list of PedonRecord objects.
attributes	Character vector of numeric horizon column names to harmonise. Default covers the chemistry / texture / Munsell numeric columns the soilKey diagnostics use.
depths	Numeric vector of GSM depth boundaries (n+1 values for n intervals). Default GSM_DEPTHS.
lam	Smoothing parameter for the spline (default 0.1, per Bishop et al. 1999 recommendation).
verbose	If TRUE (default), emits cli progress.

Value

A list of new PedonRecord objects with harmonised horizons.

Why mass-preserving

The Bishop et al. (1999) spline conserves the integral of the attribute over depth: if the original pedon has 30 g/kg OC over 0-15 cm, the harmonised pedon will report 30 g/kg integrated over 0-15 cm (split between 0-5 and 5-15 in proportion to the spline-implied gradient). This is a critical property for benchmark integrity: simple linear interpolation does not preserve mass and biases means upward / downward systematically.

Categorical handling

designation and munsell_hue_moist (and other character columns in the horizon schema) cannot be splined. Instead, for each target GSM interval, we pick the modal value weighted by the depth-overlap fraction with the input horizons. Ties broken by uppermost-input-horizon precedence.

References

Bishop, T.F.A., McBratney, A.B., Laslett, G.M. (1999). "Modelling soil attribute depth functions with equal-area quadratic smoothing splines." *Geoderma* 91: 27-45.

Arrouays, D. et al. (2014). "GlobalSoilMap: Toward a fine-resolution global grid of soil properties." *Advances in Agronomy* 125: 93-134.

See Also

mpspline2::mpspline_tidy, GSM_DEPTHS.

---

hemico *Material organico hemico (SiBCS Cap 14)*

Description

Material organico em decomposicao intermediaria: 17-40% de fibras esfregadas OU indice de von Post H5-H6. Discrimina Organossolos Hemicos no 3o nivel.

Usage

```
hemico(pedon)
```

Arguments

pedon	A PedonRecord.

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 14 (Organossolos), pp 224-226.

---

histic_horizon *Histic horizon (WRB 2022)*

Description

A horizon of organic material (WRB 2022 Ch 3.1.15): organic material (oc_pct >= min_oc, default 20, Ch 3.3.13) that is at least min_thickness cm thick (default 10). The thickness is that of the adjoining organic layers that run from the surface, taken together. A layer whose organic carbon is not recorded may be organic material and leaves the answer undecided (NA) where it could change it. The water-saturation criterion (saturated for 30 or more consecutive days, or drained) is not read.

Usage

```
histic_horizon(pedon, min_thickness = 10, min_oc = 20, surface_top_cm = 0)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness (cm), default 10.
min_oc	Minimum soil organic carbon (%) of organic material, default 20 (WRB 2022 Ch 3.3.13).
surface_top_cm	Maximum top depth (cm) of the uppermost layer of the horizon (default 0, the soil surface).

Details

Since v0.9.313 this is the horizon only: the Histosols are keyed by `histosol()`, which reads the Chapter 4 key as worded. Until then this function was also the Histosols gate, with organic material read at 12% organic carbon and a second path of 40 cm of organic material within the upper 80 cm.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 3.1.15, Histic horizon, and Chapter 3.3.13, Organic material.

---

horizon_attribute_spec

Metadata of the horizon attributes

Description

One row per canonical horizon column (see horizon_column_spec), in its order, with the section it belongs to, its unit, a description and the alternative column names that are recognised for it. The metadata come from `inst/schemas/horizon-attributes.yaml`, which also records, where known, which classification keys read an attribute, the values a categorical attribute takes and the range of a numeric one.

Usage

```
horizon_attribute_spec()
```

Value

A `data.frame` with the columns `name`, `type` (`"numeric"` or `"character"`), `section`, `unit`, `description`, `aliases`, `systems` and `allowed` (comma-separated, empty when none), and `min` and `max` (`NA` when not recorded).

See Also

horizon_column_spec, ensure_horizon_schema.

Examples

```
a <- horizon_attribute_spec()
a[a$name %in% c("clay_pct", "cec_cmol"), c("name", "unit", "description")]
```

---

horizon_column_spec *Canonical horizon column specification*

Description

Returns the schema for the `horizons` `data.table` carried by a PedonRecord: an ordered named list mapping column names to their R type (`"numeric"` or `"character"`). Adding a new attribute means editing this single function.

Usage

```
horizon_column_spec()
```

Value

Named list of column types in canonical order.

Examples

```
spec <- horizon_column_spec()
head(names(spec))
```

---

hortic *Hortic horizon (WRB 2022): garden / kitchen-midden topsoil. Diagnostic criteria: thickness \>= 20 cm, dark colour (mollic-like), high P (Mehlich-3 P >= 100 mg/kg or P2O5_1pct_citric >= 175 mg/kg), high SOC.*

Description

Hortic horizon (WRB 2022): garden / kitchen-midden topsoil. Diagnostic criteria: thickness \>= 20 cm, dark colour (mollic-like), high P (Mehlich-3 P >= 100 mg/kg or P2O5_1pct_citric >= 175 mg/kg), high SOC.

Usage

```
hortic(pedon, min_thickness = 20, min_oc = 1, min_p_mehlich3 = 100)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_oc	Numeric threshold or option (see Details).
min_p_mehlich3	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

humus_form_rp *The humus form of the Referentiel pedologique 2008*

Description

Attaches the epipedon ("episolum humifere") of a pedon to a humus form of Annex 1 of the French Referentiel pedologique 2008 (pp. 327-355): the four functional types (mull, moder, mor, amphimus) and their forms. The horizons are read as classify_rp2008 reads them (the RP notation where written, else the data), then, in order: H horizons at the top leave the form to the histosols (p. 351); an OHta makes a tangel, an An an anmoor; O horizons directly over an X, R or D give the forms without fine mineral earth (pp. 348-349); site$land_use or a ploughed L gives the open-land forms (p. 337); otherwise the key of Figure 2 (p. 354) is walked on the O and A horizons, and its outcome is made hydromorphic when the first mineral layer is a G or bears g (p. 350).

Usage

```
humus_form_rp(pedon)
```

Arguments

pedon	A PedonRecord.

Details

The key reads the kind of the O horizons (OL, OLn, OLv, OF, OH: the notation, or fine_organic_matter_pct), their zoogenic character (the suffixes zo and noz, or o_zoogenic), their lateral cover (lateral_cover_pct, or the parentheses of the RP notation), the thickness of the OH and the biological structure of the A (biogenic_aggregates_pct, else the structure). Every test is three-valued: what is not recorded leaves the form undecided among its candidates. A recorded site$humus_form decides when the horizons leave the form undecided (even outside the forms they leave possible, the conflict reported); when the horizons decide another form, they prevail and the conflict is reported.

Value

A list: `form` (the code of the form, NA when undecided), `name` (its name as printed), `type` (mull, moder, mor or amphimus, NA when undecided or for the anmoor), `label` (the name of the form when decided, else of the type when decided, else NA: what follows "a" in the name of the solum), `group` (aerated, hydromorphic, no_fine_earth, open_land, histic or NA), `decided`, `route` ("horizons", "recorded" or "none"), `candidates` (the codes of the forms still possible), `no_form_possible` (TRUE when the horizons may also fit no form), `recorded` (the recorded `site$humus_form`), `conflict` (the recorded form disagrees with the horizons), `qualifiers` (a data frame: `qualifier`, `group`, `value`, `degree` (pass, imperfect, undecided), `page`, `note`), `path` and `answers` (the walk of the key), `missing` (the attributes that would decide it) and `notes`.

References

Baize D., Girard M.-C. (coord.) (2009). Referentiel pedologique 2008. AFES, Editions Quae, Versailles. Annex 1, pp. 327-355.

See Also

`classify_rp2008`.

Examples

```
humus_form_rp(make_rp2008_example("organosols_holorganiques"))$label
```

`hydragric` *Hydragric horizon (WRB 2022): subsoil hydric horizon under anthraquic. v0.3.3 detects via designation pattern* `Bg|Brg` *immediately below an anthraquic-like topsoil.*

Description

Hydragric horizon (WRB 2022): subsoil hydric horizon under anthraquic. v0.3.3 detects via designation pattern `Bg|Brg` immediately below an anthraquic-like topsoil.

Usage

```
hydragric(pedon, min_thickness = 20)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`hypersulfidic_material`

Hypersulfidic material (WRB 2022 Ch 3.3.8): \>= 0.01% inorganic sulfidic S, pH \>= 4, capable of severe acidification on aerobic incubation.

Description

Hypersulfidic material (WRB 2022 Ch 3.3.8): \>= 0.01% inorganic sulfidic S, pH \>= 4, capable of severe acidification on aerobic incubation.

Usage

```
hypersulfidic_material(pedon, min_s_pct = 0.01, min_pH = 4)
```

Arguments

`pedon`	A `PedonRecord`.
`min_s_pct`	Numeric threshold or option (see Details).
`min_pH`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`hyposulfidic_material`

Hyposulfidic material (WRB 2022 Ch 3.3.9): same inorganic sulfidic S and field pH as hypersulfidic but does NOT consist of hypersulfidic (criterion 3 – does not acidify to pH < 4 on aerobic incubation, usually self-neutralised by carbonate). Reachable from v0.9.128: when `incubation_ph` *is measured, a sulfidic + pH>=4 layer that stays >= 4 on incubation is the set-complement of* `hypersulfidic_material` *and is reported here. Without an incubation pH the two cannot be told apart, so this returns empty (the layer is reported as potential hypersulfidic instead).*

Description

Hyposulfidic material (WRB 2022 Ch 3.3.9): same inorganic sulfidic S and field pH as hypersulfidic but does NOT consist of hypersulfidic (criterion 3 – does not acidify to pH < 4 on aerobic incubation, usually self-neutralised by carbonate). Reachable from v0.9.128: when `incubation_ph` is measured, a sulfidic + pH>=4 layer that stays >= 4 on incubation is the set-complement of `hypersulfidic_material` and is reported here. Without an incubation pH the two cannot be told apart, so this returns empty (the layer is reported as potential hypersulfidic instead).

Usage

```
hyposulfidic_material(pedon, min_s_pct = 0.01, min_pH = 4)
```

Arguments

pedon	A PedonRecord.
min_s_pct	Numeric threshold or option (see Details).
min_pH	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

inspect_bdsolos_csv *Diagnostic inspection of a BDsolos CSV before loading*

Description

Reads the CSV header, attempts to map each column to the soilKey horizon schema via .bdsolos_match_column, and prints three sections:

Usage

```
inspect_bdsolos_csv(path, sep = NULL)
```

Arguments

path	Path to the CSV downloaded from BDsolos.
sep	Field separator (default ","; some BDsolos exports use ";" or tab).

Details

- **Mapped columns** – BDsolos name -> soilKey name
- **Unmapped columns** – columns the loader will ignore (review these before running load_bdsolos_csv to make sure no critical attribute is silently dropped)
- **Munsell coverage** – whether matiz / valor / croma are present in either umido or seco variants

Run this before load_bdsolos_csv on any new CSV from BDsolos, especially if the export schema looks unfamiliar (BDsolos has shipped multiple schema versions over the years).

Value

Invisibly, a list with mapped, unmapped, munsell_present, taxon_column.

`irragric` *Irragric horizon (WRB 2022): topsoil thickened by irrigation deposits.*

Description

v0.9.317: the criteria of WRB 2022 Ch 3.1.18 (p. 50-51), tested on every stack of contiguous layers from the mineral soil surface (a surface horizon of mineral material), each against the layer directly below it:

1. soil aggregate structure or cloddy structure in $\geq$ 90% by volume;
2. a clay content $\geq$ 10% (relative) and $\geq$ 3% (absolute) higher than that of the layer directly buried; the fine clay alternative is not recorded, so a smaller increase leaves it undecided;
3. differences between subhorizons of < 20% (relative) or < 4% (absolute) in silt, clay and carbonates (also in the medium, fine and very fine sand, which soilKey does not record: undecided for more than one layer unless a recorded difference fails);
4. $\geq$ 0.3% SOC in each layer and $\geq$ 0.5% on a weighted average;
5. $\geq$ 25% traces of soil animal activity (`worm_holes_pct`, weighted);
6. evidence that the land surface has been raised: soilKey records no such evidence, so this criterion is undecided;
7. $\geq$ 20 cm thick (`min_thickness`).

The horizon is therefore never established from the data: `FALSE` when every stack fails a criterion, `NA` otherwise. Until v0.9.316 any Ap or Au of 20 cm passed.

Usage

```
irragric(pedon, min_thickness = 20)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

isis_citation *Citation for data read from ISIS*

Description

Returns the acknowledgement for profiles of the ISRIC World Soil Reference Collection read through the ISIS GraphQL service, with the date they were read. ISRIC asks users to cite the data it supplies in their products, web services, papers and reports.

Usage

```
isis_citation(accessed = Sys.Date())
```

Arguments

accessed	Date the data were read (default today).

Value

A character string.

Examples

```
isis_citation(as.Date("2026-10-29"))
```

isis_profile_to_pedon *Turn one ISIS profile into a PedonRecord*

Description

Builds a PedonRecord from one profile as the ISIS GraphQL service returns it (one element of the profiles list; see read_isis, which calls this for every profile it reads). Laboratory values are converted to soilKey units, FAO field codes to soilKey's words, and each value records its ISIS field in the provenance. The site carries the WRB and USDA names ISRIC gives the profile (reference_wrb, reference_wrb_full, reference_usda, reference_usda_subgroup and their editions), the soil moisture and temperature regimes and the groundwater (groundwater_kind, groundwater_depth_cm) when ISIS records them, the licence and the citation.

Usage

```
isis_profile_to_pedon(
  profile,
  absent_unless_recorded = TRUE,
  accessed = Sys.Date()
)
```

Arguments

`profile`	One ISIS profile, a list as parsed from the service's JSON.
`absent_unless_recorded`	If `TRUE` (default), mottles, slickensides, pressure faces, plinthite, a duripan and artefacts that the field description of a horizon does not mention are recorded as absent in that horizon; rows with laboratory data only stay unknown. ISIS has one cutans field per layer, so slickensides and pressure faces are read as absent only where that field is free (v0.9.313). `FALSE` leaves everything not recorded unknown.
`accessed`	Date the data were read (default today).

Value

A `PedonRecord`.

`kastanozem` *Kastanozem RSG diagnostic (WRB 2022)*

Description

Tests whether a profile satisfies the Kastanozem RSG criteria: a mollic horizon plus secondary carbonates plus NOT-Chernozem colour (chroma (moist) > 2 in the upper 20 cm).

Usage

```
kastanozem(pedon, max_chroma_upper = 2)
```

Arguments

`pedon`	A `PedonRecord`.
`max_chroma_upper`	Maximum moist chroma to qualify as Chernozem (default 2). Kastanozem requires the upper-20-cm chroma to EXCEED this value.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 5, Kastanozems.

kastanozem_strict *Kastanozem RSG gate (strengthened, WRB 2022 Ch 4, p 112)*

Description

Same structure as `chernozem_strict` but using the mollic horizon (no chernic gate) and starting \<= 70 cm of mineral soil surface.

Usage

```
kastanozem_strict(pedon, min_bs = 50, max_top_cm = 70, strict = NULL)
```

Arguments

pedon	A `PedonRecord`.
min_bs	Minimum base saturation % above the carbonate layer (default 50).
max_top_cm	Maximum depth of the carbonate layer below the mineral soil surface, cm (default 70).
strict	Logical or `NULL`. When `NULL` (default) it resolves via `getOption("soilKey.rsg_strict", FALSE)`. `TRUE` raises the base-saturation floor to 75%.

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

Tier-3 strict mode (v0.9.98)

With `strict = TRUE` the base-saturation floor above the carbonate-bearing layer is raised from 50% to 75%. The 70 cm carbonate-depth window is left unchanged.

key_trace_table *Flatten a classification key trace into a tabular form*

Description

Normalises the system-dependent decision trace carried by a `ClassificationResult` into a single, ordered data frame of display rows. WRB 2022 stores a flat list of reference-soil-group test steps; the hierarchical SiBCS and USDA keys store a nested list of phases (orders, suborders, great groups, subgroups, family, ...), each holding candidate steps, an assigned-taxon record, family attributes, or a bare label. This function walks all of those shapes and returns one row per step, in the order the key was evaluated, so every consumer (`print()`, the HTML and PDF reports, the Shiny app) can render the trace uniformly.

Usage

```
key_trace_table(x)
```

Arguments

x	A ClassificationResult, or the trace list taken from one.

Value

A data.frame with one row per trace step and columns:

- phase Key phase / level the step belongs to (e.g. "orders", "subgrupos"); empty for the flat WRB trace.
- code Taxon or attribute code.
- name Taxon or attribute name (or attribute value).
- status One of "passed", "failed", "indeterminate" (a test that could not be evaluated for want of data), "selected" (the taxon assigned at a level), or "info" (a family attribute or label, not a pass/fail test).
- missing Comma-separated attributes that were missing for the step (empty when none).
- n_missing Integer count of missing attributes.

A zero-row data frame with those columns when the trace is empty.

Examples

```
res <- classify_sibcs(make_ferralsol_canonical())
head(key_trace_table(res))
```

---

kst13_canonical *Keys to Soil Taxonomy 13th edition canonical reference*

Description

Convenience wrapper for canonical_reference("ST_criteria_13th"). Returns a nested list of 3,153 parsed Keys-to-Soil-Taxonomy clauses per chapter / page / key / taxon / code / clause / logic.

Usage

```
kst13_canonical(prefer_pkg = TRUE)
```

Arguments

prefer_pkg	Retained for compatibility and no longer used. Until v0.9.203 FALSE selected a copy bundled with soilKey; there is no bundled copy any more, so the data always come from SoilTaxonomy.

Details

Source: NCSS-tech `SoilTaxonomy` R package. Original: USDA-NRCS (2022). *Keys to Soil Taxonomy*, 13th edition.

Value

The canonical *Keys to Soil Taxonomy* (13th ed.) criteria reference (a list / data.frame).

`kst13_codes` *Load the canonical KST 13ed code -> taxon-name lookup table*

Description

Returns the 3,153-row data.frame from `2022_KST_codes.json` in NCSS-tech/SoilKnowledgeBase (GPL-3), which soilKey downloads on first use and caches (it ships no copy; see `clear_kst13_cache`). Each row is a (code, name) pair.

Usage

```
kst13_codes()
```

Details

Code structure:

- Single letter (`"A"`-`"L"`): Soil Order (Gelisols, Histosols, ..., Entisols)
- Two letters (`"AB"`, `"AC"`, ...): Suborder
- Three letters: Great Group
- Four letters: Subgroup

Value

A data.frame with columns `code, name`.

See Also

`kst13_criteria`, `kst13_canonical`.

kst13_criteria *Load the canonical KST 13ed criteria for a single taxon code*

Description

Returns the parsed clause data.frame for one code (e.g. `"A"` for Gelisols, `"ABA"` for Histels.Folistels, etc.). Each row is one clause of the diagnostic text with `content`, `chapter`, `page` columns.

Usage

```
kst13_criteria(code)
```

Arguments

`code`	Character. Taxon code in the KST 13ed code system (e.g. `"A"` for Gelisols, `"ABCDA"` for the Lithic Folistels subgroup).

Details

For the full 3,153-element nested list (all codes), use `kst13_canonical` (which loads the SoilTaxonomy R-package RDA equivalent).

Value

A data.frame with the parsed clauses for that code, or `NULL` if the code is not present.

See Also

`kst13_codes`, `kst13_canonical`.

leptic_features *Leptic features (WRB 2022)*

Description

Tests whether continuous rock or rock-like material occurs within `max_depth` cm of the surface. Two alternative paths qualify per WRB 2022:

1. **Designation**: a layer at depth <= `max_depth` with designation matching `"^R"` or `"^Cr"` (continuous rock or weathered rock-like substrate).
2. **Coarse fragments**: a layer at depth <= `max_depth` with coarse_fragments_pct >= `min_coarse_pct` (default 90% by volume), interpreted as rock-dominated even when not R / Cr-designated.

Either path qualifies.

Usage

```
leptic_features(pedon, max_depth = 25, min_coarse_pct = NULL, engine = NULL)
```

Arguments

pedon	A PedonRecord.
max_depth	Maximum depth (cm) at which continuous rock or rock-dominated material must appear (default 25).
min_coarse_pct	Minimum coarse-fragment percent for the coarse-fragments path (default 90 in soilkey engine, 50 in aqp engine; NULL picks a default per engine).
engine	One of "soilkey" (default; strict 90\ cfvo threshold) or "aqp" (LUCAS-friendly relaxed 50\ requiring positive evidence of rock contact – v0.9.66 tightening). The thin-topsoil path fires only when a horizon ending within max_depth also satisfies *at least one* of: (a) designation contains "R" (e.g.\ AR, BR, Cr, R, Rk), (b) coarse_fragments_pct >= 30 (gravelly), or (c) a deeper horizon is R/Cr-designated. Users with a strong external prior (e.g.\ a parent-material survey that documents rock < 25 cm but did not record it in the horizon table) can opt back into the original v0.9.65 loose behaviour with options(soilKey.leptic_assume_rock_below = TRUE). NULL (the default) reads getOption("soilKey.diagnostic_engine").

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 5, Leptosols.

---

limnic_material — *Limnic material (WRB 2022 Ch 3.3.10): subaquatic deposits (coprogenous earth, diatomaceous earth, marl, gyttja). v0.3.3: detects via* rock_origin %in% c("lacustrine", "marine") *or designation pattern.*

Description

Limnic material (WRB 2022 Ch 3.3.10): subaquatic deposits (coprogenous earth, diatomaceous earth, marl, gyttja). v0.3.3: detects via rock_origin %in% c("lacustrine", "marine") or designation pattern.

Usage

```
limnic_material(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`limonic` *Limonic horizon (WRB 2022 Ch 3.1)*

Description

From Greek *leimon* = meadow. A subaqueous / wet-meadow horizon showing accumulation of secondary Fe/Mn (oxi)hydroxides from fluctuating redox cycles. Distinct from *limnic material* (Ch 3.3.10), which is the parent material; the limonic horizon is the *soil* horizon derived from such material plus subsequent pedogenesis.

Usage

```
limonic(pedon, min_thickness = 5, min_redox_pct = 5)
```

Arguments

`pedon`	A PedonRecord.
`min_thickness`	Numeric threshold or option (see Details).
`min_redox_pct`	Numeric threshold or option (see Details).

Details

v0.3.5 detection: redoximorphic_features_pct \>= 5 AND designation pattern `Bm` / `Bjm` / `m` as proxy for past meadow wetness.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

lithic_discontinuity *Lithic discontinuity (WRB 2022 Ch 3.2.7)*

Description

Significant abrupt change in parent material between two layers. v0.3.3 simplified: detects via large discontinuity in coarse_fragments_pct (>= 10pp absolute jump) OR rock_origin difference between consecutive layers.

Usage

```
lithic_discontinuity(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

lixisol *Lixisol RSG diagnostic (WRB 2022)*

Description

The Lixisols of the WRB 2022 key (Ch 4, p. 120): an argic horizon starting $\leq$ 100 cm from the mineral soil surface and a CEC of < 24 cmolc/kg clay in some subhorizon of the argic horizon within 150 cm. The key reaches them after the Acrisols.

Usage

```
lixisol(pedon, max_cec = 24, min_bs = 50)
```

Arguments

pedon	A PedonRecord.
max_cec	CEC per kg clay below which the clay is of low activity (default 24).
min_bs	Unused since v0.9.310, kept so that existing calls still run.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 4, Lixisols (p. 120).

load_afsp_pedons *Load Africa Soil Profiles (AfSP) v1.2 as PedonRecord objects*

Description

Reads the AfSP DBase tables shipped inside `AF-AfSP1.2.zip` (downloadable from `https://files.isric.org/public/af` and converts each profile + its horizons to a soilKey `PedonRecord`. Filters to profiles with a populated WRB 2006 RSG code (i.e.\ classifiable; AfSP has ~7000 of these of the total 18,533).

Usage

```
load_afsp_pedons(
  afsp_dir,
  max_n = NULL,
  countries = NULL,
  wrb_codes = NULL,
  verbose = TRUE
)
```

Arguments

`afsp_dir`	Directory containing the extracted AfSP DBase tables (`AfSP012Qry_Profiles.dbf`, `AfSP012Qry_Layers.dbf`).
`max_n`	Optional integer; take a random sample of this size from the classifiable profiles.
`countries`	Optional character vector of ISO country codes to keep (e.g.\ `c("MW", "ET", "TZ")`).
`wrb_codes`	Optional character vector of WRB 2006 RSG codes to keep (e.g.\ `c("VR", "FR", "AC")`).
`verbose`	Print progress.

Details

The structure type, grade and size are AfSP class codes (FAO-ISRIC 1990, FAO 2006). Since v0.9.312 they are read by the AfSP dictionary (`AfSP012Qry_DictioClassValues`) into soilKey's words: a range of grades by its lower end, a range of sizes with both ends, and a code the dictionary does not define as not recorded. Each code goes to the provenance.

Since v0.9.397 the loader also reads, as recorded: `Mottling` into `mottles_present` (Y `TRUE`, N `FALSE`; the amount and the colours of the mottles are not recorded by AfSP), `ClyMinera` into `clay_mineralogy` (montmorillonitic and SM as smectite, kaolinitic, illitic, vermiculitic, interstratified; chloritic, sesquioxidic, allophane, KA/IL and gibbsite have no `clay_mineralogy` word and are read as not recorded; the value is an estimate, recorded with the provenance `inferred_prior`, so the evidence grade of a result drops to C), and `Sticknss` into `stickiness` (sticky as moderately sticky). Since v0.9.399 it reads on the site `Drain` into `drainage_class`, and `ParMat` and `LndForm` into `parent_material` and `landform` by the words of the AfSP dictionary when `AfSP012Qry_DictioClassValues.dbf` is in `afsp_dir`: a parent material with the words of its ancestors ("volcanic rocks, acid"; "clayey, fluvial, calcareous"), the codes of the unconsolidated

deposits the dictionary does not list decoded by its pattern; only the landforms that say something of the origin of the material (depression, low-gradient footslope, valley floor).

Value

A list of PedonRecord objects.

References

Leenaars, J. G. B., van Oostrum, A. J. M., & Ruiperez Gonzalez, M. (2014). Africa Soil Profiles Database, Version 1.2. ISRIC Report 2014/01. ISRIC – World Soil Information, Wageningen. Project page: https://isric.org/projects/africa-soil-profiles-database-afsp.

---

load_afsp_sample *Load the bundled AfSP stratified sample (v0.9.77)*

Description

Returns a 120-profile snapshot from AfSP v1.2 stratified by WRB RSG (5 profiles per RSG x 24 RSGs), pre-built so users can run the African WRB benchmark offline without the 35 MB ZIP download.

Usage

```
load_afsp_sample()
```

Details

This is the African analogue of load_wosis_stratified_sample (global WoSIS) and load_kssl_nasis_sample (US KSSL+NASIS).

The snapshot was built in v0.9.77 and holds the AfSP structure codes; since v0.9.312 they are read into words when it is loaded, as load_afsp_pedons reads them.

Value

A list with pedons (the PedonRecord objects), pulled_on, source, filter. benchmark_afsp takes this list as it is, or its $pedons.

Reference

Leenaars, J. G. B., van Oostrum, A. J. M., & Ruiperez Gonzalez, M. (2014). Africa Soil Profiles Database, Version 1.2. ISRIC Report 2014/01.

load_bdsolos_csv *Load a BDsolos CSV export as a list of PedonRecord objects*

Description

Reads the long-format BDsolos CSV (one row per horizon, with a profile-id key) and returns a list of PedonRecord objects. Auto-detects the column-name convention via inspect_bdsolos_csv and maps to the soilKey horizon schema. Texture (argila / silte / areia) is converted from g/kg to percent (BDsolos canonical unit).

Usage

```
load_bdsolos_csv(path, sep = NULL, verbose = TRUE)
```

Arguments

path	Path to the BDsolos CSV.
sep	Field separator. Default ",; BDsolos sometimes exports with ";" or tab – pass it explicitly.
verbose	If TRUE (default), prints a one-line summary.

Details

Iron (v0.9.221): "CDB - Ferro (g/kg)" (dithionite-citrate- bicarbonate, free Fe as the element) goes to fe_dcb_pct, which the WRB Fe-dith criteria read (nitic horizon, Ferric, Ferritic); "Ataque sulfurico - Fe2O3" (total Fe of the sulfuric attack, as the oxide) goes to fe2o3_sulfuric_pct, which the SiBCS ferrico / perferrico classes read; "Oxalato de Amonio - Ferro" goes to fe_ox_pct. All three are converted from g/kg to %. Until v0.9.220 the sulfuric Fe2O3 filled fe_dcb_pct, unconverted.

Silica and alumina (v0.9.243): "Ataque sulfurico - SiO2" and "- Al2O3" go to sio2_sulfuric_pct and al2o3_sulfuric_pct, converted from g/kg to %. The Ki index (molecular SiO2 / Al2O3), which SiBCS asks of the B latossolico, is computed from them; it reproduces the Ki column of the export.

Text of the horizon description (v0.9.245): the column "Descricao Original" holds the field description of each horizon as the survey wrote it. Two things are read from it where the structured columns are empty, and the text itself is not kept: the amount of slickensides written beside the term ("''slickenside'' comum e moderado" fills slickensides with "Comum"; a mention with no amount fills nothing), and a wedge-shaped or parallelepipedic structure ("cuneiforme", "paralelepipedica", added to structure_type; a structure "com tendencia a paralelepipedica" is not one). SiBCS knows the vertic horizon by these two features, and the surveys that predate the structured sheet describe them only there. Since v0.9.251 the amount and the grade of the cerosidade are read the same way ("cerosidade comum e moderada" fills clay_films_amount with "Comum" and clay_films_strength with "Moderada"), each where its own cell is empty.

Reading (v0.9.222): the file is read with utils::read.csv2. The export quotes every field, and its free-text fields hold line breaks and, in some states, quotation marks of their own; read.csv2 returns the export's records for all 27 state files. data.table::fread, used until v0.9.221, split

such records at the line breaks (in BA, GO, PI and RS, where the lines of text became profiles of their own) or stopped with an error that left the next `fread()` of the R session waiting forever (after DF, MT, PA, PB or SP). An empty text field is `NA` in every file; a quotation mark inside a text field is dropped.

Kind of sampling point (v0.9.242): the column `"Tipo"` of the export goes to `site$sampling_type`, as written: `"Perfil completo"`, `"Perfil parcial"` or `"Amostra extra"`. An extra sample is, in most cases, two sampled layers of a point (a topsoil and a subsoil sample, often with a gap between them), not a described profile: the thickness of its layers is that of the samples. About a third of the records of the national export are extra samples; `benchmark_bdsolos` reports the kinds apart.

Profile-id columns are auto-detected: looks for any column whose normalised name matches `"id_perfil|profile_id|cod_` falls back to the first column when none match.

Value

A list of `PedonRecord` objects. Each pedon has `site$id` from the profile-id column, the taxonomic reference (when present) at `site$reference_sibcs`, and one horizon row per CSV row matching the profile id.

See Also

`inspect_bdsolos_csv`, `download_bdsolos`.

`load_embrapa_pedons` *Load Embrapa dadosolos pedons with reference SiBCS classification*

Description

Reads the Embrapa BDsolos CSV export (or the dadosolos R package data frame, if present). Assembles a list of `PedonRecord` objects with the SiBCS classification attached as `pedon$site$reference_sibcs`.

Usage

```
load_embrapa_pedons(csv_path, head = NULL, verbose = TRUE)
```

Arguments

`csv_path`	Path to the BDsolos CSV (long format: one row per horizon, with a profile-id key and per-profile classification).
`head`	Optional integer for parser validation.
`verbose`	If `TRUE` (default), emits a summary.

Details

The dadosolos / BDsolos archive ships with ~5k profiles in PT-BR with full SiBCS classification, lab data, and horizon morphology – the primary validation set for Brazilian-context use. Available from `https://www.bdsolos.cnptia.embrapa.br/`.

Value

A list of PedonRecord objects.

---

load_febr_pedons *Load the Embrapa FEBR superconjunto into a list of PedonRecords*

Description

Reads the FEBR febr-superconjunto.txt export (one row per camada / horizon, with the profile-level classification denormalised onto every row), groups rows by (dataset_id, observacao_id), and returns a list of PedonRecord objects with all three reference taxa attached on $site: reference_sibcs (raw FEBR string, e.g. "LATOSSOLO VERMELHO"), reference_usda, reference_wrb.

Usage

```
load_febr_pedons(
  path,
  head = NULL,
  require_classification = c("sibcs", "any", "wrb", "usda"),
  verbose = TRUE
)
```

Arguments

path	Path to febr-superconjunto.txt.
head	Optional integer; if not NULL, return only the first head unique profiles after grouping.
require_classification	One of c("any", "sibcs", "wrb","usda"). Default "sibcs": drop profiles whose chosen classification is NA. "any" keeps profiles with at least one of the three.
verbose	If TRUE (default), summarises the load.

Details

Drops profiles whose taxon_sibcs is NA (no usable reference). Drops camadas with no horizon-depth information (no profund_sup / profund_inf).

Value

A list of PedonRecord objects.

`load_kssl_nasis_sample` *Load the bundled KSSL + NASIS morphological-enriched sample (v0.9.75)*

Description

Returns a 99-profile snapshot built by joining the NCSS Lab Data Mart (`ncss_labdata.gpkg`) with the companion NASIS Morphological sqlite (`NASIS_Morphological_*.sqlite`) via `load_kssl_pedons_with_nasis`. Pre-annotated with derived WRB Reference Soil Group via `usda_to_wrb_rsg`.

Usage

```
load_kssl_nasis_sample()
```

Details

Compared to `load_kssl_sample` (KSSL lab tables only), this sample carries the morphological evidence that several WRB diagnostic horizons need:

| Field | KSSL-only | KSSL + NASIS | |———-|———-:|———-:| | munsell_hue_moist | 0 | munsell_value_moist | 0 | munsell_chroma_moist | 0 | munsell_hue_dry | 0 | structure_grade | 0 | structure_type | 0 | clay_films_amount | 0 | slickensides | 0

First-ever benchmark on this enriched sample (soilKey v0.9.75, full v0.9.69-72 fallback stack):

- Top-1 baseline: 19.1\ **+3.5pp lift purely from NASIS morphology**)
- Top-1 full stack: 20.6\
- Phaeozem: 1/33 -> 2/33 (Munsell-driven mollic detection)
- Podzol: 0/15 -> 1/15

Remaining ceiling driven by attributes neither dataset preserves: Solonetz needs Na/ESP, Vertisols need slickensides + cracks (NASIS records 1.7 on subsoil samples NASIS often lacks.

Value

A list with `pedons`, `pulled_on`, `source`, `join_helper`, `cross_walk`.

Reference

Beaudette, D., Skovlin, J., Roecker, S., Brown, A. (2024). aqp: Algorithms for Quantitative Pedology. R package version 2.x. `https://github.com/ncss-tech/aqp`.

Examples

```
## Not run:
s <- load_kssl_nasis_sample()
length(s$pedons)
#> 99
# Munsell now populated (KSSL-only sample had 0%):
mean(vapply(s$pedons,
            function(p) any(!is.na(p$horizons$munsell_hue_moist)),
            logical(1)))
#> 0.99

## End(Not run)
```

`load_kssl_pedons` *Load NCSS / KSSL pedons with reference USDA Soil Taxonomy classification*

Description

Reads the KSSL pedon CSV export (typically named `NCSS_Pedon_Layer.csv` or similar) plus the lab-data CSV, joins on `pedon_key`, and assembles a list of `PedonRecord` objects. The published USDA Soil Taxonomy classification (from the `Series` or `Subgroup` field) is attached as `pedon$site$reference_usda`.

Usage

```
load_kssl_pedons(pedon_csv, layer_csv, head = NULL, verbose = TRUE)
```

Arguments

`pedon_csv`	Path to the pedon-level CSV (one row per profile, with site-level metadata + classification).
`layer_csv`	Path to the layer-level CSV (one row per horizon, with horizon properties).
`head`	Optional integer; if not `NULL`, returns only the first `head` pedons (useful for parser validation).
`verbose`	If `TRUE` (default), emits a summary of the load.

Details

KSSL is the de-facto standard for validating USDA Soil Taxonomy keys (~50k profiles, lab-grade analytical data, professional pedon descriptions). Get the export from the USDA-NRCS NCSS Lab Data Mart (`ncsslabdatamart.sc.egov.usda.gov`).

Value

A list of `PedonRecord` objects.

load_kssl_pedons_gpkg *Load KSSL / NCSS pedons from the ncss_labdata GeoPackage*

Description

Reads the ‘lab_combine_nasis_ncss‘ / ‘lab_site‘ / ‘lab_layer‘ / ‘lab_chemical_properties‘ / ‘lab_physical_properties‘ views from the NCSS Lab Data Mart GeoPackage and assembles a list of PedonRecord objects. Each pedon has its USDA Soil Taxonomy Order attached as `site$reference_usda`, normalised to match ‘classify_usda()‘ output ("Mollisols", "Alfisols", ...).

Usage

```
load_kssl_pedons_gpkg(
  gpkg,
  head = NULL,
  require_b_horizon = TRUE,
  verbose = TRUE
)
```

Arguments

gpkg	Path to `ncss_labdata.gpkg`.
head	Optional integer; load only the first N classified pedons. Useful for parser validation.
require_b_horizon	If `TRUE` (default), drops pedons whose deepest horizon's bottom_cm < 30. Most non-Entisol Order gates need a B horizon.
verbose	If `TRUE` (default), emits progress messages.

Value

A list of PedonRecord objects.

load_kssl_pedons_with_nasis *Load KSSL pedons enriched with NASIS morphology*

Description

Joins the NCSS Lab Data Mart GeoPackage with the NASIS Morphological SQLite to produce PedonRecord objects whose horizons table has BOTH lab chemistry + physics AND field morphology (Munsell, structure, clay films, slickensides, cracks). Required for the morphological-evidence diagnostics (`argic` clay-films, `vertic_horizon` slickensides, `mollic_epipedon_usda` Munsell, etc.) to fire on KSSL profiles – the lab gpkg alone has none of those.

Usage

```
load_kssl_pedons_with_nasis(
  gpkg,
  sqlite,
  head = NULL,
  require_b_horizon = TRUE,
  verbose = TRUE
)
```

Arguments

gpkg	Path to ncss_labdata.gpkg.
sqlite	Path to NASIS_Morphological_*.sqlite.
head	Optional integer; load only the first N classified pedons. Useful for parser validation / scaling.
require_b_horizon	If TRUE (default), drops pedons whose deepest horizon's bottom_cm < 30.
verbose	If TRUE (default), emits progress messages.

Value

A list of PedonRecord objects.

---

load_kssl_sample *Load the bundled KSSL/NCSS lab-data sample (v0.9.74)*

Description

Returns a 100-profile snapshot from the NCSS Lab Data Mart (KSSL gpkg, head = 100) pre-annotated with derived WRB Reference Soil Group via usda_to_wrb_rsg.

Usage

```
load_kssl_sample()
```

Details

This is the bundled offline counterpart to load_kssl_pedons_gpkg – use this for tests and demos when the 5.5 GB gpkg is not available locally.

Each pedon has BOTH:

- site$reference_usda (Order, Suborder, Greatgroup, Subgroup) – the canonical KSSL classification.
- site$reference_wrb_from_usda – the derived WRB RSG via the IUSS WRB 2022 Annex 6 cross-walk.

First-ever KSSL WRB benchmark (soilKey v0.9.74, full v0.9.69-72 fallback stack):

- Top-1 accuracy: 20.1\
- Calcisol 69\
- Phaeozem / Kastanozem / Solonetz 0\ data not in KSSL lab tables (in companion NASIS).

Value

A list with `pedons`, `pulled_on`, `source`, `cross_walk`.

Reference

Beaudette, D., Skovlin, J., Roecker, S., Brown, A. (2024). aqp: Algorithms for Quantitative Pedology. R package version 2.x. `https://github.com/ncss-tech/aqp`.

Examples

```
## Not run:
s <- load_kssl_sample()
length(s$pedons)
#> 100
table(vapply(s$pedons, function(p) p$site$reference_wrb_from_usda,
             character(1)))

## End(Not run)
```

---

load_lucas_pedons *Load EU-LUCAS / ESDB pedons with reference WRB classification*

Description

Reads the EU-LUCAS topsoil dataset joined with the ESDB profile archive (the v3 release produced by JRC). Assembles a list of `PedonRecord` objects with the WRB Reference Soil Group attached as `pedon$site$reference_wrb`.

Usage

```
load_lucas_pedons(lucas_csv, head = NULL, verbose = TRUE)
```

Arguments

`lucas_csv`	Path to the LUCAS topsoil CSV.
`head`	Optional integer for parser validation.
`verbose`	If `TRUE` (default), emits a summary.

Details

LUCAS is harvested every 3-6 years on a regular grid; the ESDB classification is updated synchronously. ~28k profile cells with WRB labels in the 2015-2018 release.

Value

A list of PedonRecord objects.

`load_lucas_soil_2018` *Load the LUCAS Soil 2018 Topsoil release as a list of PedonRecord objects*

Description

Reads the canonical European Soil Data Centre (ESDAC) release of LUCAS Soil 2018 Topsoil chemistry as published in the JRC report (ESDAC dataset `https://esdac.jrc.ec.europa.eu/content/lucas-2018-tops` The release ships ~18,984 European topsoil samples at 0-20 cm with pH (H2O and CaCl2), EC, OC, CaCO3, P, N, K and oxalate-extractable Al / Fe; a separate `BulkDensity_2018_final-2.csv` carries bulk density at 0-10 / 10-20 / 20-30 / 0-20 cm for ~6,272 of those points and is joined automatically when present.

Usage

```
load_lucas_soil_2018(
  path,
  attach_bulk_density = TRUE,
  countries = NULL,
  max_n = NULL,
  verbose = TRUE
)
```

Arguments

`path`	Folder containing `LUCAS-SOIL-2018.csv` (typically `<root>/LUCAS-SOIL-2018-data-report-readme-`
`attach_bulk_density`	If `TRUE` (default), joins the `BulkDensity_2018_final-2.csv` sister file on `POINTID` when present.
`countries`	Optional character vector of NUTS_0 codes (e.g. `c("ES", "FR")`) to filter pedons. Default `NULL` (all countries).
`max_n`	Optional integer cap on the number of pedons returned (after country filter). Useful for development.
`verbose`	If `TRUE` (default), prints a summary line.

Details

What's NOT in the release (and how to fill it):

- **Texture (clay / sand / silt)** – not in this CSV. Pass `benchmark_lucas_2018(..., fill_texture_from = "soilgrids")` to fill from ISRIC SoilGrids 250m via `lookup_soilgrids`.
- **Munsell colors** – not collected by LUCAS Soil 2018. If the user has Vis-NIR spectra (release separate ~83 GB), use `predict_munsell_from_spectra` (v0.9.47).
- **Vis-NIR spectra** – distributed separately by ESDAC. Once downloaded and attached to the pedon's `$spectra`, `predict_from_spectra` (v0.9.46) fills clay / sand / silt / pH / OC / CEC.
- **Taxonomic reference** – not in the LUCAS release; `benchmark_lucas_2018` attaches the canonical WRB Reference Soil Group via `lookup_esdb` (v0.9.44) at the pedon's coordinates.

Unit conversions applied (LUCAS -> soilKey schema):

- OC, N, CaCO3, Ox_Al, Ox_Fe: g/kg ->
- EC: mS/m -> dS/m (* 0.01)
- P, K: mg/kg unchanged
- pH: unitless

Special LUCAS string values `"< LOD"`, `"<LOD"`, empty cells and `"n.d."` / `"ND"` are converted to `NA` before numeric coercion.

Value

A list of `PedonRecord` objects (one per LUCAS point). Each pedon has a `site$id` matching the LUCAS `POINTID`, `site$lat` / `site$lon` in WGS84, and either one or two horizons (the second being 20-30 cm when the subsoil OC / CaCO3 columns are populated). Provenance entries from the loader use `source = "measured"`.

See Also

`benchmark_lucas_2018`, `lookup_esdb`, `lookup_soilgrids`.

Examples

```
## Not run:
path <- "soil_data/eu_lucas/LUCAS-SOIL-2018-data-report-readme-v2/LUCAS-SOIL-2018-v2"
pedons <- load_lucas_soil_2018(path, countries = c("ES", "PT"),
                                max_n = 100)
length(pedons)
pedons[[1]]

## End(Not run)
```

load_natsoil_pedons *Read NatSoil profiles as pedons*

Description

Turns the profile files saved by fetch_natsoil (or any JSON of the NatSoil API with horizons, samples and laboratory results) into PedonRecord objects for classify_asc3 and the other keys. Sites whose licence does not allow external use are left out.

Usage

```
load_natsoil_pedons(
  path,
  classified_only = TRUE,
  max_n = NULL,
  seed = NULL,
  verbose = TRUE
)
```

Arguments

path	A folder (searched recursively for .json and .json.gz files) or a vector of files.
classified_only	Keep only sites with an ASC order.
max_n	Optional maximum number of files read (a random sample; a profile file of fetch_natsoil holds one site).
seed	Optional seed for that sample.
verbose	Print a summary.

Details

The horizons keep their designation in the Australian notation (lithologic prefix, master, subdivision and suffixes), depths in cm, the field texture grade, boundary distinctness and shape, the grade and type of the first structure (lenticular when any structure is), the moist and the dry matrix colours (a colour with no moisture status is read as moist), slickensides, crack width, coarse fragments (abundance classes at their middle, added up to at most 100 %; unknown when a record has no abundance), calcareous segregations (a record with only its nature is a template and is not read; one with a form or size but no abundance counts as very few, 1 %), cementation of pans, effervescence, plasticity type and the moisture when described.

Laboratory results are matched to the horizons by depth: a sample counts for the horizon it is filed under where they overlap, and for another one when it lies mostly within it, weighed by the overlap (a sample with no usable depth counts for the horizon it is filed under). Codes are matched whatever their case, the "not recorded" methods included. A value below the detection limit, written "<x", is read as half the limit, unless the limit lies above a limit the keys read for that property (0.5 % CaCO3, 0.18 cmol/kg Na, an ESP of 6, 5 % free Fe, 12 % organic C, 10 % clay), where the value is unknown and the next method is tried. pH in water is the lab 1:5 value (method 4A1, or a pH

whose method is not recorded, often the field pH itself); the field pH is used only where the lab has no pH at all. Organic C is the measured value, or an uncorrected Walkley-Black value times 1.3; half the loss on ignition only in peat (an O or P horizon, or a peat field texture). Exchangeable bases and CEC are read whatever the extraction method; a laboratory ESP is kept as `esp_pct`. Free iron is read only from dithionite extractions; most profiles have none, which leaves the Ferrosols undecided.

The site keeps the surveyors' classification (`reference_asc_order`, `reference_asc_suborder`, `reference_asc_great_group`, `reference_asc_subgroup`, the confidence `reference_asc_confidence`), the drainage class, the landform pattern and element (`landform`, `landform_element`), the surface condition, the microrelief, coordinates (GDA94), the licence and the citation. Mottles (not biological or mechanical mixing, nor substrate inclusions) give `redoximorphic_features_pct` (abundance classes at their middle, added up) and `mottle_contrast` (the strongest, faint when distinct and prominent ones are 10 % or less on their own and faint ones make up the rest); calcareous segregations and calcrete fragments give `carbonate_soft_pct` (soft segregations, root linings, tubules, veins and crystals) and `carbonate_hard_pct` (nodules, concretions, fragments and laminae, and calcrete coarse fragments), kept only where what is recorded leaves the amount within one carbonate class of the ASC glossary (a segregation of unknown form may be soft or hard, a fragment of unknown lithology calcrete; coarse fragments not recorded count as none; segregations not recorded leave the soft amount unknown, and the hard one unless calcrete alone exceeds 50 %). Ferruginous and bauxite nodules give `ferric_nodules_pct` and `bauxitic_nodules_pct`; since 0.9.285 manganiferous nodules give `manganiferous_nodules_pct`, and gypseous and saline segregations `gypsum_visible_pct` and `halite_visible_pct`. Only what is recorded counts (a segregation of unrecorded nature or form is not one of them); the amount is NA when no segregation is recorded, and NaN when the recorded classes leave it on either side of 20 % (50 % for halite), which the key reads as undecided. Abundance classes of several records are added as ranges (each class strictly between its limits). Laboratory EC of a 1:5 extract gives `ec_1to5_dS_m`; the soil strength gives `strength_dry` or `strength_moist` by the moisture of its record; the size class of the first structure gives `ped_size_max_mm`. Since 0.9.292, for the families: `gravel_pct` (the coarse fragments of a gravel size class, 2-60 mm, or of no size recorded; NaN, unknown, when such a record has no abundance), `water_repellence`, and at the site the coarse fragments of the ground surface (`surface_coarse_fragments_pct`, all sizes, and `surface_gravel_pct`; NA when none is recorded, NaN when a record has no abundance) and the surveyors' family codes by position (`reference_asc_family`, `"B;E;M;O;V"`).

Value

A named list of `PedonRecord` objects (names are `agency_project_site`).

References

CSIRO. CSIRO National Soil Site Database (NatSoil). https://doi.org/10.25919/sg5w-ny87

See Also

`fetch_natsoil`, `benchmark_natsoil`.

load_npdb_pedons *Read the National Pedon Database of Canada*

Description

Reads the pre-packaged CSV files of the National Pedon Database (NPDB, version 3, Agriculture and Agri-Food Canada, CanSIS) into PedonRecord objects, with the survey's classification under the Canadian System of Soil Classification as the reference.

Usage

```
load_npdb_pedons(dir, max_n = NULL, seed = NULL, verbose = TRUE)
```

Arguments

dir	Folder with the NPDB CSV files.
max_n	Optional maximum number of pedons (a random sample).
seed	Optional seed for that sample.
verbose	Print progress.

Details

The NPDB is not distributed with soilKey. It is published on the Open Government Portal of Canada under the Open Government Licence - Canada (dataset "National Pedon Database"); download the pre-packaged CSV files and unzip them into a folder. Info.csv and Horizons.csv are required; Morphology.csv, Physical.csv, Chemical.csv and Profiles.csv are read when present.

What the reader does with the records:

- records are split on CR-LF only, since some fields hold a lone carriage return ("subangular\r blocky");
- the designation is rebuilt in the Canadian notation from its parts (lithologic numeral, master, suffixes, numeric modifier: "IICk1"), and the site declares designation_system = "cssc";
- depths: organic and water horizons recorded upward from the mineral surface (top above bottom) become negative depths; a mineral horizon with its depths reversed or negative keeps |top| with no bottom when it is the deepest layer, and no depths otherwise; a missing top is taken from the layer above; every profile is shifted to begin at 0 cm, and site$npdb_depth_order_mismatch flags profiles whose depth order differs from the order of their layers;
- colours: the codes ("100yr4030", "10yr2010", "025y 4020", "10.0yr4.0/3.0", "n 50") are read with hue numbers above 0 and up to 10; of the two matrix colours, the one that is never darker within a profile (and lighter at least once) is read as the dry colour and the other as the moist one; when they do not separate, COLOUR_1 is read as the moist colour and no dry colour is read;
- total N: the value 9 (a no-data code) and a value repeated over three or more horizons are not read; the units are told apart by the median C:N of each pedon (of its dataset when the pedon has too few values): 5 to 40 is read as percent, 0.5 to 4 as g/kg (divided by 10), anything else is not read; a horizon whose C:N then lies outside 3 to 60 is not read either;

- base saturation: computed from the exchangeable bases and the CEC (at most 100 %) when they are recorded, the recorded value being kept when it agrees with it (one dataset records something else in that column);
- pH: a dataset whose pH in CaCl2 is mostly above its pH in water has the two columns swapped back;
- values outside their possible range are not read;
- the survey's order, great group and subgroup words are turned into the names of the 3rd edition (1998) where they match one (`reference_cssc_order`, `reference_cssc_great_group`, `reference_cssc_subgroup` including the subgroups the 1998 edition merged (Solodic into Solonetzic); names of earlier editions with no 1998 equivalent are left out, and the raw words are kept in `reference_cssc_raw`.

The NPDB records no soil climate (mean annual soil temperature, moisture subclass), the chemistry of the podzolic B (pyrophosphate-extractable Fe and Al) or the contrast of mottles. The Chernozemic order and the Gray Brown Luvisols, which ask for the climate, are undecided on these records (see `benchmark_npdb`).

Value

A named list of `PedonRecord` objects (names are the NPDB `PEDON_ID`).

References

Agriculture and Agri-Food Canada (2021). National Pedon Database, version 3. Open Government Portal, Open Government Licence - Canada.

Examples

```
## Not run:
peds <- load_npdb_pedons("path/to/npdb_v3_csv")
b <- benchmark_npdb(peds, level = "order")
b$accuracy

## End(Not run)
```

`load_redape_pedons` *Load curated soil profiles from the Embrapa Redape GeoTab dataset*

Description

Reads the structured JSON files (one profile per file) published by Vaz et al. 2023 at the Embrapa Redape repository (DOI `10.48432/PYKKA7`) and converts each one to a soilKey `PedonRecord`.

Usage

```
load_redape_pedons(json_dir, max_n = NULL, verbose = TRUE)
```

Arguments

`json_dir`	Directory containing the GeoTab JSON files (or a character vector of file paths).
`max_n`	If non-`NULL`, take a random sample of this size.
`verbose`	Print progress (default `TRUE`).

Details

The dataset is unique in two ways:

- Every profile was hand-reviewed by experienced pedologists (the curation note and author list are preserved on each pedon site record), so it is suitable as a gold-standard benchmark.
- Unlike BDsolos, all profiles ship the full exchange complex (Ca, Mg, K, Na, Al *and H*), so `cec_cmol` (Valor T = S + H + Al) is computed directly without any fallback option.

Value

A list of `PedonRecord` objects.

Reference

Vaz, G. J., Silva Jr, A. F., & Silva Neto, L. de F. da (2023). Brazilian soil data for taxonomic classification. Redape, V1. doi:10.48432/PYKKA7.

See Also

`download_redape_dataset`, `benchmark_redape`.

`load_rules` *Load a soilKey rule set (YAML)*

Description

Load a soilKey rule set (YAML)

Usage

```
load_rules(system = "wrb2022", package = "soilKey")
```

Arguments

`system`	Id of a registered classification system (see `soilkey_systems`): `"wrb2022"` (WRB 2022, the default), `"sibcs5"` (SiBCS 5th edition) or `"usda"` (Keys to Soil Taxonomy, 13th edition). Aliases such as `"wrb"` and `"sibcs"` are accepted.
`package`	Package owning the rule files (default `"soilKey"`).

Value

A parsed YAML list with elements `version`, `source`, and a system-specific taxa list (`rsgs`, `orders`, or `ordens`), with the lists of the lower levels read from their folders and, where a level file has one, its `<level>_from` block (classes whose list at that level is taken from another class of the top level).

---

load_wosis_sample *Load the bundled WoSIS South-America sample*

Description

Returns a 40-profile snapshot of WoSIS GraphQL data pulled on 2026-05-03 with `continent = "South America"`. The data is a frozen artefact – do NOT use it for current paper-grade benchmarks (the WoSIS database is updated periodically; the bundled snapshot is for reproducible tests and offline development only).

Usage

```
load_wosis_sample()
```

Details

For up-to-date benchmarks, call `run_wosis_benchmark_graphql()` directly against the live ISRIC GraphQL endpoint.

Value

A list as described above.

Returned data

A list with elements:

- `profiles_raw` – the parsed GraphQL response (one element per profile; nested layer arrays).
- `pedons` – `PedonRecord` objects ready for classification (one per profile).
- `pulled_on` – `Date` of the snapshot.
- `endpoint`, `filter`, `n_pulled` – metadata.

Examples

```
## Not run:
sample <- load_wosis_sample()
length(sample$pedons)
#> 40
classify_wrb2022(sample$pedons[[1]])$rsg_or_order

## End(Not run)
```

`load_wosis_stratified_sample` *Load the bundled WoSIS stratified RSG-balanced sample (v0.9.73)*

Description

Returns a 130-profile snapshot of WoSIS GraphQL data pulled on 2026-05-09 with **stratified sampling by WRB Reference Soil Group**: 5 profiles per RSG across 26 RSGs (Acrisol, Andosol, Arenosol, Calcisol, Cambisol, Chernozem, Cryosol, Ferralsol, Fluvisol, Gleysol, Gypsisol, Histosol, Kastanozem, Leptosol, Luvisol, Nitisol, Phaeozem, Planosol, Plinthosol, Podzol, Regosol, Solonchak, Solonetz, Stagnosol, Umbrisol, Vertisol).

Usage

```
load_wosis_stratified_sample()
```

Details

This is the recommended cache for global WRB benchmarking. Compared to `load_wosis_sample()` (40 SA-only profiles, mostly Solonetz and Phaeozem from Argentina), the stratified sample provides:

- Even coverage across the 26 most important RSGs.
- Richer analytical attributes – CEC available on 26 ECEC on 37 in the SA snapshot).
- Geographic diversity (Angola, Brazil, USA, China, Russia, South Africa, Indonesia, Argentina, etc.).

First-ever benchmark on this sample (soilKey v0.9.73, full v0.9.69-72 fallback stack):

- Overall top-1 accuracy: 16.2\
- Histosol 100\ from 20\ Cambisol 60\
- 18 RSGs at 0\ expose (Munsell colours, base saturation, sodium for Solonetz, slickensides for Vertisols, etc.). Documented data ceiling.

For the live API, call `run_wosis_benchmark_graphql()` or the `read_wosis_profiles_graphql(wrb_rsg = "...", n_max = N)` helper (small RSG-filtered queries are tractable; large unfiltered pulls time out as of 2026-05).

Value

A list with:

- `pedons`: list of 130 `PedonRecord` objects.
- `meta`: named integer vector of profiles per RSG.
- `pulled_on`: pull date.
- `endpoint`: ISRIC GraphQL endpoint URL.
- `filter`: pull strategy metadata.
- `n_pulled`: 130.

Reference

Batjes, N. H., Ribeiro, E., van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth System Science Data*, 12, 299-320. doi:10.5194/essd122992020.

Examples

```
## Not run:
s <- load_wosis_stratified_sample()
length(s$pedons)
#> 130
table(vapply(s$pedons, function(p) p$site$wosis_rsg, character(1)))
#> 5 of each: Acrisol, Andosol, ... Vertisol

## End(Not run)
```

`lookup_esdb` *Look up an ESDB raster value at WGS84 coordinates*

Description

Loads the requested attribute raster, reprojects WGS84 lat/lon input to the raster's native CRS (typically LAEA Europe, EPSG:3035), and extracts the value(s). When a Value Attribute Table ('.vat.dbf') is available, the integer raster value is decoded to its coded string (e.g. '21' -> '"LV"' -> Luvisol).

Usage

```
lookup_esdb(coords, attribute, raster_root, decode = TRUE)
```

Arguments

`coords`	A two-column matrix or data.frame with 'lon' and 'lat' (WGS84 decimal degrees) – in that order. A single `c(lon, lat)` vector is also accepted.
`attribute`	Name of the ESDB attribute folder, e.g. `"WRBLV1"` or `"WRBFU"`. See `available_esdb_attributes`.
`raster_root`	Path to the unpacked ESDB raster directory.
`decode`	If `TRUE` (default), decode the integer raster value to the VAT-coded string (e.g. `"21"` -> `"LV"`). If `FALSE`, return the raw integer.

Details

Coordinates outside the European raster footprint return 'NA' silently (rather than erroring) so vectorised calls degrade gracefully.

Value

Character vector (decoded codes) or numeric vector (raw values) of the same length as `nrow(coords)`. `NA` for points outside the raster footprint.

See Also

`available_esdb_attributes`

Examples

```
## Not run:
root <- "~/data/ESDB-Raster-Library-1k-GeoTIFF-20240507"

# Single point: Wageningen, Netherlands (5.66 E, 51.97 N)
lookup_esdb(c(5.66, 51.97), "WRBLV1", root)
#> [1] "GL"   # Gleysol per the ESDB 1km raster

# Vector: Lisbon + Berlin + Helsinki
coords <- rbind(c(-9.14, 38.72), c(13.40, 52.52), c(24.94, 60.17))
lookup_esdb(coords, "WRBLV1", root)
#> [1] "CM" "LV" "PZ"   # Cambisol, Luvisol, Podzol

## End(Not run)
```

`lookup_mapbiomas_solos` *Look up a MapBiomas Solos raster value at WGS84 coordinates*

Description

MapBiomas Solos (Project MapBiomas, Brazil) distributes a national raster of SiBCS classes at 30 m, downloadable from `https://mapbiomas.org/en/produtos`. This helper mirrors the shape of `lookup_esdb` but is local-file only: pass the path of the unpacked GeoTIFF and the function reprojects the user's WGS84 lat/lon to the raster's native CRS, extracts the pixel and (optionally) decodes the integer class code via a user-supplied legend.

Usage

```
lookup_mapbiomas_solos(coords, raster_path, legend = NULL)
```

Arguments

`coords`	A 2-column matrix or data.frame with `lon`, `lat` (WGS84 decimal degrees), or a length-2 numeric vector for a single query.
`raster_path`	Path to the unpacked MapBiomas Solos GeoTIFF.
`legend`	Optional two-column data.frame (first column = numeric value, second = SiBCS class name). When provided, the integer raster value is decoded; when `NULL`, the raw integer is returned.

Details

MapBiomas does not bundle a '.vat.dbf'; the canonical legend is published as a CSV / dictionary on their website. Pass it via `legend` as a two-column data.frame (`value`, `class_name`) to enable decoding.

Value

Character vector of decoded class names (when `legend` is supplied) or numeric vector of raster values. Same length as `nrow(coords)`. `NA` for points outside the raster footprint.

See Also

`lookup_esdb`, `lookup_soilgrids`.

Examples

```
## Not run:
tif <- "~/data/mapbiomas_solos_collection2_2023.tif"
legend <- data.frame(
  value = c(1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L, 11L, 12L, 13L),
  class_name = c("Latossolo Vermelho-Amarelo",
                 "Latossolo Amarelo",
                 "Argissolo Vermelho-Amarelo",
                 "Argissolo Amarelo",
                 "Neossolo Quartzarenico",
                 "Cambissolo Haplico",
                 "Espodossolo",
                 "Gleissolo",
                 "Nitossolo",
                 "Planossolo",
                 "Plintossolo",
                 "Vertisolo",
                 "Outros")
)
lookup_mapbiomas_solos(c(-43.0, -22.0), tif, legend)

## End(Not run)
```

---

lookup_soilgrids *Look up a SoilGrids 250m soil property at WGS84 coordinates*

Description

Reads ISRIC SoilGrids 250m (Hengl et al. 2017, 2021) directly from the ISRIC Cloud-Optimized GeoTIFF (COG) endpoint at `https://files.isric.org/soilgrids/latest/data/` – no download required, only the pixel under each query coordinate is transferred over HTTPS.

Usage

```
lookup_soilgrids(
  coords,
 property = c("clay", "sand", "silt", "phh2o", "soc", "cec", "bdod", "nitrogen", "ocd",
    "ocs", "cfvo"),
  depth = c("0-5cm", "5-15cm", "15-30cm", "30-60cm", "60-100cm", "100-200cm"),
  quantile = c("mean", "Q0.05", "Q0.5", "Q0.95", "uncertainty"),
  baseurl = "https://files.isric.org/soilgrids/latest/data",
  raw = FALSE
)
```

Arguments

coords	A 2-column matrix or data.frame with lon, lat (WGS84 decimal degrees), or a length-2 numeric vector for a single query.
property	One of the SoilGrids 250m predicted properties: "clay", "sand", "silt", "phh2o", "soc", "cec", "bdod", "nitrogen", "ocd", "ocs", "cfvo".
depth	Depth interval. One of "0-5cm", "5-15cm", "15-30cm", "30-60cm", "60-100cm", "100-200cm".
quantile	Output quantile. One of "mean" (default), "Q0.05", "Q0.5", "Q0.95", "uncertainty".
baseurl	Base URL of the SoilGrids COG endpoint. Default is the canonical ISRIC location; override only for a local mirror.
raw	If TRUE, returns the integer raster value without scaling. Default FALSE (returns the value in conventional units).

Details

SoilGrids stores integer rasters scaled per property; this helper applies the canonical conversion factor so the returned value is in conventional soil units (%, pH, g/kg, cmol(c)/kg, g/cm^3).

Value

Numeric vector of length nrow(coords). NA outside the SoilGrids footprint or on network errors.

See Also

lookup_esdb, lookup_mapbiomas_solos.

Examples

```
## Not run:
# Single point
lookup_soilgrids(c(-43.0, -22.0),
                 property = "phh2o",
                 depth = "0-5cm",
                 quantile = "mean")

# Vector + multiple properties
```

```
coords <- rbind(c(-43.0, -22.0), c( -9.14, 38.72))
lookup_soilgrids(coords, "clay",  "0-5cm", "mean")
lookup_soilgrids(coords, "phh2o", "0-5cm", "mean")

## End(Not run)
```

`luvisol` *Luvisol RSG diagnostic (WRB 2022)*

Description

The Luvisols of the WRB 2022 key (Ch 4, p. 122): "Other soils having an argic horizon starting $\leq$ 100 cm from the mineral soil surface".

Usage

```
luvisol(pedon, min_cec = 24, max_al_sat = 50)
```

Arguments

`pedon`	A `PedonRecord`.
`min_cec, max_al_sat`	Unused since v0.9.310, kept so that existing calls still run.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 4, Luvisols (p. 122).

`make_acrisol_canonical`

Build the canonical Acrisol fixture

Description

Synthetic tropical-humid Acrisol on weathered gneiss: argic horizon at Bt1 with low-activity clay (CEC/clay ~ 17 cmol_c/kg clay) and low base saturation (BS ~ 25%). v0.9.310: in Bt1 and Bt2 the exchangeable Al (0.4 until then) exceeds the exchangeable bases (lowered from 1.15 to 0.72 cmolc/kg, so that the ECEC stays below 6 cmolc/kg clay, Geric), as the WRB 2022 Acrisols ask for in 50-100 cm (Ch 4, p. 119); with Al below the bases the profile is a Lixisol. By construction:

- `argic`: PASSES on Bt1.
- `acrisol`: PASSES (CEC low, exchangeable Al > bases).
- `lixisol`, `alisol`, `luvisol`: FAIL.
- Other diagnostics: FAIL.

Usage

```
make_acrisol_canonical()
```

Value

A PedonRecord.

`make_alisol_canonical` *Build the canonical Alisol fixture*

Description

Synthetic humid-tropical Alisol on weathered shale: argic horizon at Bt1 with high-activity clay (CEC/clay ~ 34) AND high Al saturation (Al sat ~ 70%); the canonical "young weathering on a 2:1 clay parent that has not yet released enough Al into the precipitate-stabilised pool". By construction:

- argic: PASSES on Bt1.
- alisol: PASSES (CEC high, Al sat high).
- acrisol, lixisol, luvisol: FAIL.

Usage

```
make_alisol_canonical()
```

Value

A PedonRecord.

`make_andosol_canonical`
Build the canonical Andosol fixture

Description

Synthetic Andosol on volcanic tephra: very dark surface with low bulk density (0.7 g/cm^3) and high active Al + Fe (Al_ox + 0.5 * Fe_ox = 2.25%). By construction andic_properties passes.

Usage

```
make_andosol_canonical()
```

Value

A PedonRecord.

`make_anthrosol_canonical` *Build the canonical Anthrosol fixture*

Description

Synthetic Anthrosol with a hortic horizon – a long-cultivated dark surface from sustained organic-matter additions (typical of centuries-old kitchen-garden / homegarden soils). By construction `anthric_horizons` passes via the designation pattern.

Usage

```
make_anthrosol_canonical()
```

Value

A `PedonRecord`.

`make_arenosol_canonical` *Build the canonical Arenosol fixture*

Description

Synthetic coastal-dune Arenosol: sandy throughout the upper 100 cm (silt + 2*clay « 30). By construction `arenic_texture` passes uniformly while every clay-dependent diagnostic fails.

Usage

```
make_arenosol_canonical()
```

Value

A `PedonRecord`.

make_argissolo_canonical *Perfil canonico de Argissolo (SiBCS 5a ed., Cap 5)*

Description

B textural com gradiente significativo, argila ativ baixa ou alta + V baixa. Catch-all final na chave – tipica do Brasil tropical.

Usage

```
make_argissolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_asc3_example *Example pedons for the Australian orders*

Description

Small, plausible profiles written for the tests and examples, one per order of the Australian Soil Classification, with designations in the Australian notation (Field Handbook), field textures and the properties the order key reads. They are not real profiles.

Usage

```
make_asc3_example(
  order = c("anthroposols", "organosols", "podosols", "vertosols", "hydrosols",
   "kurosols", "sodosols", "chromosols", "calcarosols", "ferrosols", "dermosols",
    "kandosols", "arenosols", "rudosols", "tenosols")
)
```

Arguments

order	One of the 15 orders, in lower case and plural: "anthroposols", "organosols", "podosols", "vertosols", "hydrosols", "kurosols", "sodosols", "chromosols", "calcarosols", "ferrosols", "dermosols", "kandosols", "arenosols", "rudosols", "tenosols".

Value

A PedonRecord.

Examples

```
p <- make_asc3_example("kandosols")
p$horizons[, c("designation", "top_cm", "bottom_cm", "field_texture")]
```

`make_calcisol_canonical`

Build the canonical Calcisol fixture

Description

Synthetic semi-arid Calcisol on calcareous loess: A horizon with modest secondary carbonate; a thick Bk1 with the diagnostic calcic horizon (35% CaCO3 over 40 cm); deepening accumulation in Bk2. By construction:

- `calcic`: PASSES on Bk1 and Bk2.
- `gypsic`, `salic`: FAIL.
- `argic`, `ferralic`, `mollic`: FAIL.

Usage

```
make_calcisol_canonical()
```

Value

A `PedonRecord`.

`make_cambisol_canonical`

Build the canonical Cambisol fixture

Description

Synthetic temperate-zone Cambisol on weathered colluvium: modest subsurface alteration in Bw without meeting argic clay-increase or ferralic CEC criteria. By construction:

- `cambic`: PASSES on Bw (thickness 35 cm, sandy clay loam, no argic / no ferralic).
- `argic`, `ferralic`, `mollic`, `calcic`, `gypsic`, `salic`: FAIL.

Usage

```
make_cambisol_canonical()
```

Value

A `PedonRecord`.

make_cambissolo_canonical

Perfil canonico de Cambissolo (SiBCS 5a ed., Cap 6)

Description

Reusa fixture WRB Cambisol – B incipiente sem ser plintico, vertico, planico, etc.

Usage

```
make_cambissolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_chernossolo_canonical

Perfil canonico de Chernossolo (SiBCS 5a ed., Cap 7)

Description

Reusa fixture WRB Chernozem – A chernozemico + Bk com argila Ta + V alta. SiBCS strictos exigem (a) Bi/Bt + Ta + V alta, OR (b) calcico/petrocalcico/carbonatico + A chernozemico.

Usage

```
make_chernossolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_chernozem_canonical *Build the canonical Chernozem fixture*

Description

Synthetic Ukrainian / Russian steppe Chernozem on loess: thick dark Ah, granular structure, secondary carbonates accumulating in the Bk. By construction:

- mollic: PASSES on horizon Ah1 (moist value 2, chroma 1, dry value 3; SOC 4%; BS 89%; thickness 30 cm; strong granular structure).
- argic: FAILS (essentially no clay differentiation; ratios all close to 1).
- ferralic: FAILS (CEC/clay ~ 120 cmol_c/kg clay – high-activity 2:1 clay).

Usage

```
make_chernozem_canonical()
```

Value

A PedonRecord.

make_cpcs1967_example *Example pedons for the CPCS 1967 classes*

Description

Small, plausible profiles written for the tests and examples, one per class of the French soil classification (CPCS 1967), with designations in the old French notation and the properties the key reads. They are not real profiles.

Usage

```
make_cpcs1967_example(
  class = c("raw_mineral", "weakly_developed", "vertisol", "andosol", "calcimagnesic",
    "isohumic", "brunified", "podzolised", "sesquioxide", "ferrallitic", "hydromorphic",
     "sodic")
)
```

Arguments

class	One of "raw_mineral", "weakly_developed", "vertisol", "andosol", "calcimagnesic", "isohumic", "brunified", "podzolised", "sesquioxide", "ferrallitic", "hydromorphic", "sodic".

Value

A PedonRecord.

Examples

```
classify_cpcs1967(make_cpcs1967_example("podzolised"), on_missing = "silent")$name
```

`make_cryosol_canonical` *Build the canonical Cryosol fixture*

Description

Synthetic Arctic Cryosol on weathered shale with permafrost at 50 cm: thawed A horizon over a frozen Bf horizon. By construction `cryic_conditions` passes via the designation pattern.

Usage

```
make_cryosol_canonical()
```

Value

A PedonRecord.

`make_cscj2011_example` *Example pedons for the CSCJ great groups*

Description

Small, plausible profiles written for the tests and examples, one per great group of the Comprehensive Soil Classification System of Japan (2011), with designations in the Japanese notation. Most are the `make_scsj2017_example` profiles of the great group of the same concept; the Dark Red soil (from limestone) is the CSCJ's own. They are not real profiles.

Usage

```
make_cscj2011_example(
 great_group = c("man_made", "organic", "podzol", "andosol", "dark_red", "lowland",
    "red_yellow", "stagnic", "brown_forest", "regosol")
)
```

Arguments

great_group	One of "man_made", "organic", "podzol", "andosol", "dark_red", "lowland", "red_yellow", "stagnic", "brown_forest", "regosol".

Value

A PedonRecord.

Examples

```
p <- make_cscj2011_example("dark_red")
p$horizons[, c("designation", "top_cm", "bottom_cm")]
```

`make_cssc3_example` *Example pedons for the Canadian orders*

Description

Small, plausible profiles written for the tests and examples, one per order of the Canadian System of Soil Classification, with designations in the Canadian notation and the properties the order, great-group and subgroup keys read: a Brunisolic Dystric Turbic Cryosol, a Terric Humic Mesisol, an Orthic Vertisol, an Orthic Humo-Ferric Podzol, an Orthic Humic Gleysol, a Dark Brown Solonetz, an Orthic Dark Brown Chernozem, an Orthic Gray Luvisol, an Orthic Eutric Brunisol and an Orthic Regosol. They are not records of real pedons.

Usage

```
make_cssc3_example(
 order = c("cryosolic", "organic", "vertisolic", "podzolic", "gleysolic", "solonetzic",
    "chernozemic", "luvisolic", "brunisolic", "regosolic")
)
```

Arguments

`order`	One of `"cryosolic"`, `"organic"`, `"vertisolic"`, `"podzolic"`, `"gleysolic"`, `"solonetzic"`, `"chernozemic"`, `"luvisolic"`, `"brunisolic"`, `"regosolic"`.

Value

A PedonRecord.

Examples

```
p <- make_cssc3_example("chernozemic")
p$horizons[, c("designation", "top_cm", "bottom_cm", "oc_pct")]
```

`make_cst2001_example` *Example pedons for the suborders, groups and subgroups of the Chinese Soil Taxonomy (2001)*

Description

Small, plausible profiles written for the tests and examples, one per suborder of the Chinese Soil Taxonomy (2001), with FAO designations, the regimes on the site and the data the key reads; each keys to its own suborder, not provisionally (with the group it reaches). Since v0.9.374 a group code (`"A1.1"` ... `"N4.7"`, and soilKey's `"[A1.4]"` and `"[A2.5]"`) gives one profile per group, 140 in all, each keying to its own group, not provisionally; since v0.9.375 a subgroup code (`"A1.1.1"` ... `"N4.7.6"`, and `"[A1.3.x]"`) gives one profile per subgroup (orders A to G since v0.9.375, H to N since v0.9.376), 588 in all, each keying to its own subgroup, not provisionally. They are not real profiles. With `near_miss = TRUE`, fourteen of them (one per order) change one property so that the profile misses its class and lands in another: A1 without permafrost (Orthic Histosols), B2 with a cumulic layer of 35 cm (cumulic evidence only: Orthic Primosols), C1 with 40 g/kg carbon in the Bhs (Orthic Spodosols), D1 frigid (Udic Andosols), E1 with a CEC of 18 cmol/kg clay (a LAC-ferric horizon: Udic Ferrosols), F1 with rock at 45 cm (Aquic Cambosols), G1 without the crust and pavement (Udic Cambosols, with the note on the residual's name), H1 with an ESP of 22 (alkalic evidence: Ustic Argosols), I1 without permafrost (Orthic Gleyosols), J2 with its humus near the surface (no isohumic property: Ustic Cambosols), K1 with a CEC of 26 cmol/kg clay (Ustic Argosols), L1 gelic (Udic Argosols), M3 aridic and not irrigated (Udic Cambosols, with the note) and N2 with 2 g/kg carbon (Orthic Primosols).

Usage

```
make_cst2001_example(suborder = .CST_EXAMPLE_CODES, near_miss = FALSE)
```

Arguments

`suborder`	The code of the suborder, as printed (`"A1"` ... `"N4"`), of a group (`"A1.1"` ... `"N4.7"`) or of a subgroup (`"A1.1.1"` ... `"N4.7.6"`).
`near_miss`	Give the near miss of the suborder (one per order: `"A1"`, `"B2"`, `"C1"`, `"D1"`, `"E1"`, `"F1"`, `"G1"`, `"H1"`, `"I1"`, `"J2"`, `"K1"`, `"L1"`, `"M3"`, `"N2"`).

Value

A `PedonRecord`.

See Also

`classify_cst2001`.

Examples

```
p <- make_cst2001_example("E1")
p$horizons[, c("designation", "top_cm", "bottom_cm", "clay_pct", "cec_cmol")]
classify_cst2001(make_cst2001_example("E1", near_miss = TRUE), on_missing = "silent")$name
classify_cst2001(make_cst2001_example("K3.2"), on_missing = "silent")$levels
classify_cst2001(make_cst2001_example("E1.3.4"), on_missing = "silent")$name
classify_cst2001(make_cst2001_example("L4.2.1"), on_missing = "silent")$name
```

`make_durisol_canonical` *Build the canonical Durisol fixture*

Description

Synthetic semi-arid Durisol with a Si-cemented subsurface horizon (35% duripan nodules over 45 cm). By construction `duric_horizon` passes on Bdu.

Usage

```
make_durisol_canonical()
```

Value

A `PedonRecord`.

`make_empty_horizons` *Build an empty horizons data.table with the canonical schema*

Description

Build an empty horizons data.table with the canonical schema

Usage

```
make_empty_horizons(n = 0L)
```

Arguments

`n`	Number of rows (default 0).

Value

A `data.table` with all canonical horizon columns filled with NAs of the correct type.

Examples

```
h <- make_empty_horizons(3)
nrow(h)
```

`make_espodossolo_canonical` *Perfil canonico de Espodossolo (SiBCS 5a ed., Cap 8)*

Description

Reusa fixture WRB Podzol – B espodico imediatamente abaixo de E.

Usage

```
make_espodossolo_canonical()
```

Value

A `PedonRecord` populated with the canonical horizons and site metadata for this reference profile.

`make_fao_example` *Example pedons for the FAO-Unesco legend*

Description

Small synthetic profiles written for the tests and examples, one per major soil grouping and one per soil unit of the Revised Legend (FAO 1988), built from the key's numbers (no profile of the book is copied). A grouping's example keys to its own grouping without an undecided gate before it; a unit's example (built on its grouping's) keys to its own unit without an undecided grouping or unit before it. With `near_miss = TRUE`, the same profile with one printed number of its gate crossed (a Luvisol with a base saturation of 45 % in its lower B, a Histosol with 35 cm of peat, a Lithic Leptosol with its rock at 11 cm), which keys elsewhere; a unit whose gate has no printed number has no near miss. They are not real profiles.

Usage

```
make_fao_example(code = "HS", edition = c("1988", "1974"), near_miss = FALSE)
```

Arguments

`code`	The code of a major soil grouping of 1988 ("HS", "AT", "LP", "VR", "FL", "SC", "GL", "AN", "AR", "RG", "PZ", "PT", "FR", "PL", "SN", "GR", "CH", "KS", "PH", "PD", "GY", "CL", "NT", "AL", "AC", "LV", "LX", "CM"), or of one of its 153 soil units ("HSi", "LVx", "CMe" ...); with `edition = "1974"`, of a major soil group of 1974 ("O", "I", "V", "J", "Z", "G", "T", "Q", "R", "U", "E", "P", "F", "W", "S", "M", "C", "K", "H", "D", "X", "Y", "N", "A", "L", "B") or of one of its 106 soil units ("Ox", "Lc", "Be" ...).
`edition`	`"1988"` (default) or `"1974"` (since v0.9.389).
`near_miss`	`TRUE` for the near miss of the grouping.

Value

A PedonRecord.

See Also

classify_fao.

Examples

```
p <- make_fao_example("PH")
p$horizons[, c("designation", "top_cm", "bottom_cm", "bs_pct")]
classify_fao(make_fao_example("LVx"), on_missing = "silent")$name
classify_fao(make_fao_example("Fx", edition = "1974"), edition = "1974",
             on_missing = "silent")$name
```

---

make_ferralsol_canonical

Build the canonical Ferralsol fixture

Description

Synthetic but realistic Brazilian Latossolo Vermelho (Ferralsol per WRB 2022): deeply weathered, kaolinitic / oxidic, with the canonical "low-activity clay" signature. Diagnostic outcomes are deterministic by construction:

- ferralic: PASSES on horizons Bw1 and Bw2 (CEC/clay = 8.3 cmol_c/kg clay; ECEC/clay = 3.6 cmol_c/kg clay; texture sandy clay / clay; thickness >= 30 cm).
- argic: FAILS (gradual clay increase, all pairwise ratios < 1.2; absolute increment too small for the >= 40% rule).
- mollic: FAILS (chroma > 3, BS < 50%, A horizon < 20 cm thick).

Usage

```
make_ferralsol_canonical()
```

Value

A PedonRecord.

make_fluvisol_canonical

Build the canonical Fluvisol fixture

Description

Synthetic floodplain Fluvisol: stratified textures across consecutive C horizons, OC pattern non-monotone with depth (because C2 is more recently deposited, OC-richer than C1). By construction `fluvic_material` passes.

Usage

```
make_fluvisol_canonical()
```

Details

v0.9.310: the strata are numbered as deposits of their own (C, 2C, 3C; FAO 2006 Guidelines), the lithic discontinuities that their particle-size distributions show (WRB 2022 Ch 3.2.7, criterion 2.a.iv), and the first two are massive, as unaltered alluvium is. Written C1, C2, C3, the clay-richer second stratum met the argic horizon (WRB 2022 Ch 3.1.3), and the profile keyed out as a Luvisol once the Luvisols asked only for an argic horizon starting <= 100 cm (Ch 4, p. 122); with the weak subangular blocky structure they had, the C below the Ah (chroma 1 unit higher) was a cambic horizon (Ch 3.1.5, criterion 3.b.iii) and the profile a Fluvic Cambisol.

Value

A `PedonRecord`.

make_gleissolo_canonical

Perfil canonico de Gleissolo (SiBCS 5a ed., Cap 9)

Description

Reusa fixture WRB Gleysol – horizonte glei dentro de 50 cm.

Usage

```
make_gleissolo_canonical()
```

Value

A `PedonRecord` populated with the canonical horizons and site metadata for this reference profile.

make_gleysol_canonical

Build the canonical Gleysol fixture

Description

Synthetic Gleysol from a high-water-table floodplain: A with low chroma but no explicit redox features (so gleyic test is anchored on Bg); Bg with diagnostic redoximorphic features (35% by volume) within the upper 50 cm. By construction:

- gleyic_properties: PASSES on Bg.
- argic, ferralic, mollic, cambic, plinthic, spodic, calcic, gypsic, salic: FAIL.

Usage

```
make_gleysol_canonical()
```

Value

A PedonRecord.

make_gypsisol_canonical

Build the canonical Gypsisol fixture

Description

Synthetic Gypsisol on gypsiferous parent material: shallow A; gypsum accumulation rising sharply in the By1 horizon (35% gypsum over 50 cm) – the diagnostic gypsic horizon. By construction:

- gypsic: PASSES on By1 and By2.
- calcic, salic: FAIL.
- argic, ferralic, mollic: FAIL.

Usage

```
make_gypsisol_canonical()
```

Value

A PedonRecord.

make_histosol_canonical *Build the canonical Histosol fixture*

Description

Synthetic boreal-mire Histosol: thick (50 cm) surface organic horizon with OC ~ 35%, low chroma, no exchangeable-base data reported (typical of histic profiles where laboratory chemistry on organic material is reported separately). By construction:

- histic_horizon: PASSES on Oa.
- Mineral horizons below; mollic / umbric NA (no BS reported).

Usage

```
make_histosol_canonical()
```

Value

A PedonRecord.

make_hscs2024_example *Example pedons for the Hungarian soil types*

Description

Small, plausible profiles written for the tests and examples, one per type of the Hungarian diagnostic soil classification (Micheli et al. 2024), with designations in the FAO (2006) notation the text adopts and the properties the key reads, including a total water-soluble salt content (soluble_salts_pct) and exchangeable cations, so that the types before each one are decided. They are not real profiles.

Usage

```
make_hscs2024_example(
  type = c("peat", "anthropogenic", "rocky_skeletal", "rock_influenced", "solonetz",
    "solonchak", "swelling_clay", "meadow", "carbonate", "steppe", "clay_illuvial",
      "sandy", "brown_earth", "deposit", "weakly_developed")
)
```

Arguments

type	One of "peat", "anthropogenic", "rocky_skeletal", "rock_influenced", "solonetz", "solonchak", "swelling_clay", "meadow", "carbonate", "steppe", "clay_illuvial", "sandy", "brown_earth", "deposit", "weakly_developed".

Value

A PedonRecord.

Examples

```
p <- make_hscs2024_example("meadow")
p$horizons[, c("designation", "top_cm", "bottom_cm")]
```

---

make_ka5_example *Example pedons for the types of the German KA5 soil systematics*

Description

Small, plausible profiles written for the tests and examples, one per type of Tab. 51 (56), with designations in the KA5 notation (declared in `site$designation_system`) and the properties soilKey's order of tests reads on the data. They are not real profiles. `near_miss = TRUE` gives a twin that misses the type's gate (a depth, a share, a record moved across it); `route = "data"` drops the designations and the declared notation, so that the horizons are read on the data.

Usage

```
make_ka5_example(
  type = c("KV", "KM", "KH", "HH", "HN", "JP", "JG", "JS", "JD", "IA", "IW", "UEA", "MR",
    "MC", "MO", "MK", "MD", "MH", "MN", "GH", "GM", "GN", "GG", "AO", "AT", "AQ", "AZ",
    "AB", "FF", "FS", "OO", "OL", "TC", "TT", "RN", "RQ", "RR", "RZ", "XX", "YK", "YE",
    "YY", "YU", "YO", "SG", "SH", "SS", "PP", "LF", "LL", "BB", "CR", "CF", "DD", "VV",
      "VW"),
  near_miss = FALSE,
  route = c("notation", "data"),
  subtype = NULL,
  variety = NULL
)
```

Arguments

type	The code of a type: "KV", "KM", "KH", "HH", "HN" (Moore); "JP", "JG", "JS", "JD", "IA", "IW" (subhydric); "UEA" (the Strand, the key's Ue-A), "MR", "MC", "MO", "MK", "MD", "MH", "MN", "GH", "GM", "GN", "GG", "AO", "AT", "AQ", "AZ", "AB" (semiterrestrial); "FF", "FS", "OO", "OL", "TC", "TT", "RN", "RQ", "RR", "RZ", "XX", "YK", "YE", "YY", "YU", "YO", "SG", "SH", "SS", "PP", "LF", "LL", "BB", "CR", "CF", "DD", "VV", "VW" (terrestrial).
near_miss	Return the twin that misses the type's gate.
route	"notation" (default) or "data".

subtype	The code of a subtype, as printed ("BBn", "GG-BB", "LF-YU", "BB-GG", "MRb", "KV-KM" ...; the 223 of pp. 202-262, the Strand's also as "UEAn", and the open "[GHu]" and "[KHm]"): the example profile of that subtype, written from one of its printed formulas with depths inside its windows (type and near_miss are then not read). The type examples key to the Norm subtype of their type (the Treposol's to an example subtype, the Moorgley's to the Niedermoorgley: they have no Norm).
variety	The code of a variety of Tab. 49 ("p", "g", "eu", "mo", "v" ...; the 38): an example profile of a subtype Tab. 50 offers it to, with the variety's horizon or record (type, near_miss and subtype are then not read).

Value

A PedonRecord.

Examples

```
classify_ka5(make_ka5_example("LL"), on_missing = "silent")$name
classify_ka5(make_ka5_example("LL", route = "data"), on_missing = "silent")$name
classify_ka5(make_ka5_example(subtype = "GG-BB"), on_missing = "silent")$name
classify_ka5(make_ka5_example(variety = "p"), on_missing = "silent")$name
```

---

make_kastanozem_canonical

Build the canonical Kastanozem fixture

Description

Synthetic continental-semiarid Kastanozem on loess-like substrate: mollic surface (chroma 3, value 3) – dark enough for mollic but not dark enough for Chernozem (chroma 3 > 2 in the upper 20 cm); secondary carbonates accumulating in the Bk. By construction:

- mollic: PASSES.
- kastanozem: PASSES.
- chernozem: FAILS.
- phaeozem: PASSES too (WRB 2022 Phaeozems allow carbonates); the key reaches the Kastanozems first.

Usage

```
make_kastanozem_canonical()
```

Value

A PedonRecord.

make_kidpr2008_example

Example pedons for the divisions of the Russian soil classification

Description

Small, plausible profiles written for the tests and examples, one per division of the Russian soil classification (Field Guide of 2008), with designations in the Russian notation (declared in `site$designation_system`) and a few properties. They are not real profiles.

Usage

```
make_kidpr2008_example(
  division = c("AA", "AB", "BA", "BB", "BC", "DA", "CA", "CB", "CC", "CD", "CE", "CF",
    "CG", "CH", "CI", "CJ", "CK", "CL", "CM", "CN", "CO", "CP", "CQ", "CR", "CS", "CT",
      "CU")
)
```

Arguments

`division`	The code of a division: `"AA"`, `"AB"` (trunk A), `"BA"`, `"BB"`, `"BC"` (trunk B), `"DA"` (trunk D, the division soilKey codes [DA]), `"CA"` to `"CU"` (trunk C).

Value

A `PedonRecord`.

Examples

```
classify_kidpr2008(make_kidpr2008_example("CI"), on_missing = "silent")$name
```

make_latossolo_canonical

Perfil canonico de Latossolo (SiBCS 5a ed., Cap 10)

Description

Reusa fixture WRB Ferralsol – B latossolico imediatamente abaixo de A, sem horizonte argilico acima.

Usage

```
make_latossolo_canonical()
```

Value

A `PedonRecord` populated with the canonical horizons and site metadata for this reference profile.

make_leptosol_canonical

Build the canonical Leptosol fixture

Description

Synthetic mountain-slope Leptosol on metamorphic rock: a thin A (10 cm) directly over continuous rock. By construction:

- leptic_features: PASSES (R at 10 cm <= 25).
- Other diagnostics fail on thickness, missing data, or absent diagnostic features.

Usage

```
make_leptosol_canonical()
```

Value

A PedonRecord.

make_lixisol_canonical

Build the canonical Lixisol fixture

Description

Synthetic Mediterranean / sub-tropical Lixisol on weathered calcareous parent material: argic horizon at Bt1 with low-activity clay (CEC/clay ~ 20) but high base saturation (BS ~ 70%) thanks to carbonate-buffered weathering. By construction:

- argic: PASSES on Bt1.
- lixisol: PASSES (CEC low, BS high).
- acrisol, alisol, luvisol: FAIL.

Usage

```
make_lixisol_canonical()
```

Value

A PedonRecord.

make_luvisol_canonical

Build the canonical Luvisol fixture

Description

Synthetic temperate-zone Luvisol on loess: clear textural differentiation, Bt with clay coatings, high base saturation, high- activity clay. By construction:

- argic: PASSES on horizon Bt1 (clay increase from E (18%) to Bt1 (35%) gives ratio 1.94 in the 15-40% band; thickness 25 cm; texture clay loam; no glossic features).
- ferralic: FAILS (CEC/clay ~ 45 cmol_c/kg clay in the Bt – well above the 16 cmol_c/kg threshold).
- mollic: FAILS (A horizon: moist value 4 > 3, thickness 10 cm < 20 cm).

Usage

```
make_luvisol_canonical()
```

Value

A PedonRecord.

make_luvissolo_canonical

Perfil canonico de Luvissolo (SiBCS 5a ed., Cap 11)

Description

Solo com B textural argila Ta + V alta. Tipico do semiarido com rocha basica.

Usage

```
make_luvissolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_neossolo_canonical

Perfil canonico de Neossolo Litolico (SiBCS 5a ed., Cap 12)

Description

Solo raso sobre rocha continua dura. Sem horizonte B diagnostico.

Usage

```
make_neossolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_nitisol_canonical

Build the canonical Nitisol fixture

Description

Synthetic East-African Nitisol on weathered basalt: clay-rich (>= 50%), Fe-rich (DCB ~ 6%), polyhedral structure with shiny ped surfaces. By construction nitic_horizon passes.

Usage

```
make_nitisol_canonical()
```

Value

A PedonRecord.

make_nitossolo_canonical

Perfil canonico de Nitossolo Vermelho (SiBCS 5a ed., Cap 13)

Description

Solo argiloso (>= 35% argila desde superficie) com B nitico (estrutura forte em blocos + cerosidade), gradiente textural baixo (B/A <= 1.5).

Usage

```
make_nitossolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_nzsc4_example

Example pedons for the New Zealand orders

Description

Small, plausible profiles written for the tests and examples, one per order of the New Zealand Soil Classification, with designations in the New Zealand notation and the properties the order key reads. Most key to the last (Orthic or default) group of their order. Their subgroup may be provisional where a measure the subgroups read is missing (sulfidic S, exchangeable Ca and Mg, the origin of the layers). They are not real profiles.

Usage

```
make_nzsc4_example(
  order = c("organic", "gley", "ultic", "podzol", "allophanic", "pumice", "melanic",
    "semiarid", "oxidic", "granular", "pallic", "brown", "anthropic", "recent", "raw")
)
```

Arguments

order	One of "organic", "gley", "ultic", "podzol", "allophanic", "pumice", "melanic", "semiarid", "oxidic", "granular", "pallic", "brown", "anthropic", "recent", "raw".

Value

A PedonRecord.

Examples

```
p <- make_nzsc4_example("pallic")
p$horizons[, c("designation", "top_cm", "bottom_cm")]
```

`make_organossolo_canonical`
Perfil canonico de Organossolo (SiBCS 5a ed., Cap 14)

Description

Solo organico saturado, com horizonte H histico >= 60 cm e SOC alto. Tipico de varzea / brejo.

Usage

```
make_organossolo_canonical()
```

Value

A `PedonRecord` populated with the canonical horizons and site metadata for this reference profile.

`make_phaeozem_canonical`
Build the canonical Phaeozem fixture

Description

Synthetic humid-temperate Phaeozem on non-calcareous loess: mollic (chroma 2, value 2-3) and high BS, but no secondary carbonates anywhere – typical of more leached / less-arid steppe-forest transition. By construction:

- `mollic`: PASSES.
- `phaeozem`: PASSES.
- `chernozem`, `kastanozem`: FAIL (no carbonates).

Usage

```
make_phaeozem_canonical()
```

Value

A `PedonRecord`.

`make_planosol_canonical` *Build the canonical Planosol fixture*

Description

Synthetic temperate Planosol with abrupt textural change: sandy E (clay 12%) overlies a clay-rich Bt (35%) at 25 cm with an abrupt boundary. By construction `planic_features` passes.

Usage

```
make_planosol_canonical()
```

Value

A `PedonRecord`.

`make_planossolo_canonical` *Perfil canonico de Planossolo (SiBCS 5a ed., Cap 15)*

Description

Solo com horizonte E sobrejacente a B planico (mudanca textural abrupta + cores neutras + cromas baixos).

Usage

```
make_planossolo_canonical()
```

Value

A `PedonRecord` populated with the canonical horizons and site metadata for this reference profile.

make_plinthosol_canonical

Build the canonical Plinthosol fixture

Description

Synthetic seasonally-saturated tropical Plinthosol: A horizon with typical Cerrado SOC; Btv with diagnostic plinthite (25% by volume over 60 cm); persistent plinthite at depth. By construction:

- plinthic: PASSES on Btv and Cv.
- argic, ferralic, mollic, spodic, calcic, gypsic, salic: FAIL.

Usage

```
make_plinthosol_canonical()
```

Value

A PedonRecord.

make_plintossolo_canonical

Perfil canonico de Plintossolo (SiBCS 5a ed., Cap 16)

Description

Reusa fixture WRB Plinthosol – horizonte plintico iniciando dentro de 40 cm. The WRB fixture writes its plinthic horizons "Btv" and "Cv", as KST and FAO do. In the Brazilian notation "v" marks vertic features and plinthite is "f", which for KST and FAO is frozen soil: a suffix read one way by one key is read another way by the next. This profile is classified by the three keys (it is one of the demonstration profiles of the app), so its horizons carry neither letter ("Bt", "C") and its plinthite is the measured one.

Usage

```
make_plintossolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_podzol_canonical *Build the canonical Podzol fixture*

Description

Synthetic boreal / temperate-coniferous Podzol: bleached E (low clay, low CEC), illuvial Bs with diagnostic Al/Fe oxalate accumulation, weathered C. By construction:

- spodic: PASSES on Bs (Al_ox + 0.5*Fe_ox = 0.6, pH 4.5, 40 cm thick).
- argic, ferralic, mollic, cambic, plinthic, calcic, gypsic, salic: FAIL.

Usage

```
make_podzol_canonical()
```

Details

E horizon Munsell is set to chroma 3 (rather than canonical 1-2 of a true albic) to keep gleyic_properties clearly negative under the conservative v0.2 criterion.

Value

A PedonRecord.

make_retisol_canonical *Build the canonical Retisol fixture*

Description

Synthetic temperate Retisol on loess over clay-rich substrate: bleached E with glossic tongues penetrating the underlying argic Bt. By construction retic_properties passes via the "glossic" designation pattern; argic also passes (this is correct – Retisols are argic + retic features, and the WRB key tests RT before AC/LX/AL/LV). The Bt is below 1.5 kg/dm3, so that the claric E has no stagnic properties (WRB 2022 3.2.14 criterion 4) and the profile does not key as a Stagnosol.

Usage

```
make_retisol_canonical()
```

Value

A PedonRecord.

make_rp2008_example *Example pedons for the references of the Referentiel pedologique 2008*

Description

Small, plausible profiles written for the tests and examples, one per reference (the 110 of the 34 GER), in the RP notation (`site$designation_system = "rp2008"`) with the properties their tests read. They are not real profiles.

Usage

```
make_rp2008_example(
  reference = c("alocrisols_typiques", "alocrisols_humiques", "vitrandosols",
   "silandosols_eutriques", "silandosols_dystriques", "silandosols_perhydriques",
   "aluandosols_hapliques", "aluandosols_perhydriques", "anthroposols_transformes",
   "anthroposols_artificiels", "anthroposols_reconstitues", "anthroposols_construits",
    "anthroposols_archeologiques", "arenosols", "brunisols_eutriques",
   "brunisols_dystriques", "rendosols", "calcosols", "dolomitosols", "rendisols",
    "calcisols", "magnesisols", "calcarisols", "chernosols_hapliques",

    "chernosols_typiques", "chernosols_melanoluviques", "colluviosols",
    "cryosols_histiques", "cryosols_mineraux", "ferrallitisols_meubles",
    "ferrallitisols_nodulaires", "ferrallitisols_petroxydiques",
    "ferrallitisols_duroxydiques", "oxydisols_meubles", "oxydisols_nodulaires",
   "oxydisols_petroxydiques", "ferruginosols_meubles", "ferruginosols_semiluviques",
   "ferruginosols_luviques", "ferruginosols_petroxydiques", "fersialsols_carbonates",
    "fersialsols_calciques", "fersialsols_insatures", "fersialsols_eluviques",

    "fluviosols_bruts", "fluviosols_juveniles", "fluviosols_typiques",
   "fluviosols_brunifies", "grisols_eluviques", "grisols_degrades", "grisols_hapliques",
    "gypsosols_hapliques", "gypsosols_petrogypsiques", "histosols_fibriques",
    "histosols_mesiques", "histosols_sapriques", "histosols_composites",
    "histosols_leptiques", "leptismectisols", "lithosols", "neoluvisols",
   "luvisols_typiques", "luvisols_degrades", "luvisols_derniques", "luvisols_tronques",
    "quasi_luvisols", "nitosols", "organosols_holorganiques",

    "organosols_calcaires", "organosols_satures", "organosols_insatures",
   "pelosols_typiques", "pelosols_brunifies", "pelosols_differencies", "peyrosols",
    "phaeosols_hapliques", "phaeosols_melanoluviques", "planosols_typiques",
    "planosols_distaux", "planosols_structuraux", "podzosols_ocriques",
    "podzosols_meubles", "podzosols_duriques", "podzosols_humiques",
    "podzosols_humo_duriques", "podzosols_eluviques", "podzosols_placiques",
   "post_podzosols", "rankosols", "reductisols_typiques", "reductisols_stagniques",

   "redoxisols", "regosols", "salisols_chloruro_sulfates", "salisols_carbonates",
    "sodisols_indifferencies", "sodisols_solonetziques", "sodisols_solodises",
   "sodisalisols", "salisodisols", "thalassosols_bruts", "thalassosols_juveniles",
```

```
    "thalassosols_polderises", "thiosols", "sulfatosols", "veracrisols",
    "lithovertisols", "topovertisols", "paravertisols_hapliques",
    "paravertisols_planosoliques")
)
```

Arguments

reference	The code of a reference (see `system.file("rules", "rp2008", "key.yaml", package = "soilKey")`): "alocrisols_typiques", "alocrisols_humiques", "vitrandosols", "silandosols_eutriques", "silandosols_dystriques", "silandosols_perhydriques", "aluandosols_hapliques", "aluandosols_perhydriques", "anthroposols_transformes", "anthroposols_artificiels", "anthroposols_reconstitues", "anthroposols_construits", "anthroposols_archeologiques", "arenosols", "brunisols_eutriques", "brunisols_dystriques", "rendosols", "calcosols", "dolomitosols", "rendisols", "calcisols", "magnesisols", "calcarisols", "chernosols_hapliques", "chernosols_typiques", "chernosols_melanoluviques", "colluviosols", "cryosols_histiques", "cryosols_mineraux", "ferrallitisols_meubles", "ferrallitisols_nodulaires", "ferrallitisols_petroxydiques", "ferrallitisols_duroxydiq "oxydisols_meubles", "oxydisols_nodulaires", "oxydisols_petroxydiques", "ferruginosols_meubles", "ferruginosols_semiluviques", "ferruginosols_luviques", "ferruginosols_petroxydiques", "fersialsols_carbonates", "fersialsols_calciques", "fersialsols_insatures", "fersialsols_eluviques", "fluviosols_bruts", "fluviosols_juveniles", "fluviosols_typiques", "fluviosols_brunifies", "grisols_eluviques", "grisols_degrades", "grisols_hapliques", "gypsosols_hapliques", "gypsosols_petrogypsiques", "histosols_fibriques", "histosols_mesiques", "histosols_sapriques", "histosols_composites", "histosols_leptiques", "leptismectisols", "lithosols", "neoluvisols", "luvisols_typiques", "luvisols_degrades", "luvisols_derniques", "luvisols_tronques", "quasi_luvisols", "nitosols", "organosols_holorganiques", "organosols_calcaires", "organosols_satures", "organosols_insatures", "pelosols_typiques", "pelosols_brunifies", "pelosols_differencies", "peyrosols", "phaeosols_hapliques", "phaeosols_melanoluviques" "planosols_typiques", "planosols_distaux", "planosols_structuraux", "podzosols_ocriques", "podzosols_meubles", "podzosols_duriques", "podzosols_humiques", "podzosols_humo_duriques", "podzosols_eluviques", "podzosols_placiques", "post_podzosols", "rankosols", "reductisols_typiques", "reductisols_stagniques", "redoxisols", "regosols", "salisols_chloruro_sulfates", "salisols_carbonates", "sodisols_indifferencies", "sodisols_solonetziques", "sodisols_solodises", "sodisalisols", "salisodisols", "thalassosols_bruts", "thalassosols_juveniles", "thalassosols_polderises", "thiosols", "sulfatosols", "veracrisols", "lithovertisols", "topovertisols", "paravertisols_hapliques", "paravertisols_planosoliq

Value

A PedonRecord.

Examples

```
classify_rp2008(make_rp2008_example("brunisols_eutriques"), on_missing = "silent")$name
```

make_scsj2017_example *Example pedons for the Japanese great groups*

Description

Small, plausible profiles written for the tests and examples, one per great group of the Soil Classification System of Japan (2017), with designations in the Japanese notation and the properties the key reads. They are not real profiles.

Usage

```
make_scsj2017_example(
 great_group = c("man_made", "organic", "andosol", "podzol", "fluvic", "red_yellow",
    "stagnic", "eutrosol", "brown_forest", "regosol")
)
```

Arguments

great_group	One of "man_made", "organic", "andosol", "podzol", "fluvic", "red_yellow", "stagnic", "eutrosol", "brown_forest", "regosol".

Value

A PedonRecord.

Examples

```
p <- make_scsj2017_example("podzol")
p$horizons[, c("designation", "top_cm", "bottom_cm")]
```

make_solonchak_canonical *Build the canonical Solonchak fixture*

Description

Synthetic Solonchak from a coastal-arid setting: surface salt accumulation gives the diagnostic salic horizon (EC 25 dS/m over 25 cm); EC declines but stays elevated in the Bz; non-saline C below. By construction:

- salic: PASSES on Az.
- gypsic, calcic: FAIL.
- argic, ferralic, mollic: FAIL.

Usage

```
make_solonchak_canonical()
```

Value

A PedonRecord.

`make_solonetz_canonical` *Build the canonical Solonetz fixture*

Description

Synthetic Solonetz on saline-sodic substrate: argic Btn with columnar structure and high exchangeable Na (ESP ~ 28%). By construction `natric_horizon` passes.

Usage

```
make_solonetz_canonical()
```

Value

A PedonRecord.

`make_stagnosol_canonical` *Build the canonical Stagnosol fixture*

Description

Synthetic Stagnosol: redoximorphic features in a perched layer (Bg, 15-50 cm; redox 25%) but the deeper subsoil is well-drained (BC redox 2%, C redox 0). Since v0.9.312 what tells stagnic from gleyic is the recorded perched water table and the Bg's matrix, which has no reductimorphic colour (WRB 2022 3.2.6 and 3.2.14, as Hugo reads them): `stagnic_properties` passes, `gleyic_properties` does not, and the grey Bg covers more than one third of 0-60 cm (the Stagnosols key, Ch 4, p. 108).

Usage

```
make_stagnosol_canonical()
```

Value

A PedonRecord.

make_synthetic_pedon_with_spectra

Build a synthetic PedonRecord with attached spectra (testing aid)

Description

Generates a small, deterministic PedonRecord with n_horizons horizons and a Vis-NIR spectral matrix (350:2500 nm). Useful for exercising fill_from_spectra in tests and vignettes.

Usage

```
make_synthetic_pedon_with_spectra(
  n_horizons = 5L,
  wavelengths = 350:2500,
  seed = 1L
)
```

Arguments

n_horizons	Integer number of horizons (default 5).
wavelengths	Integer vector of wavelengths (default 350:2500).
seed	Integer seed for the RNG used to generate the spectra.

Value

A PedonRecord with a $spectra$vnir matrix attached.

make_technosol_canonical

Build the canonical Technosol fixture

Description

Synthetic urban / industrial Technosol: surface horizon with 30% anthropogenic artefacts (brick, glass, slag, plastic). By construction technic_features passes.

Usage

```
make_technosol_canonical()
```

Value

A PedonRecord.

`make_umbrisol_canonical` *Build the canonical Umbrisol fixture*

Description

Synthetic humid-temperate Umbrisol on weathered acidic schist: deep organic-rich dark surface with low base saturation – the acid analogue of a Phaeozem. By construction `umbric_horizon` passes; `mollic` fails on BS < 50.

Usage

```
make_umbrisol_canonical()
```

Value

A `PedonRecord`.

`make_vertisol_canonical` *Build the canonical Vertisol fixture*

Description

Synthetic Vertisol from a smectite-rich plain: deep clay (50-55%) with strong slickensides in the Bss horizon. Surface chroma 4 (above the mollic cap) so that vertic_properties is the only v0.2 diagnostic that passes. By construction:

- `vertic_properties`: PASSES on Bss and BC.
- `argic`, `ferralic`, `mollic`, `cambic`, `plinthic`, `spodic`, `calcic`, `gypsic`, `salic`: FAIL.

Usage

```
make_vertisol_canonical()
```

Value

A `PedonRecord`.

make_vertissolo_canonical

Perfil canonico de Vertissolo (SiBCS 5a ed., Cap 17)

Description

Solo argiloso (>= 30% argila desde superficie) com horizonte vertico (slickensides + fendas + clay alto) iniciando dentro de 100 cm. Reusa structure / fixture do WRB Vertisol.

Usage

```
make_vertissolo_canonical()
```

Value

A PedonRecord populated with the canonical horizons and site metadata for this reference profile.

make_za1991_example

Example pedons for the soil forms of the South African classification

Description

Small synthetic profiles written for the tests and examples, one per soil form of the South African soil classification of 1991 (named in lower case: `"hutton"`, `"clovelly"` ...), each with one layer per diagnostic horizon or material of the form's sequence and the properties soilKey's tests read, and five special cases: `"lichtenburg"` (named by van Huyssteen 2020, outside the 1991 key: provisional), `"lithosol"` (an orthic A of 80 mm over unconsolidated material: no form, the land class rock and lithosols), `"undecided"` (an orthic A over a B whose hue is not recorded: Hutton or Clovelly undecided), `"buried"` (400 mm of aeolian sand over a Valsrivier: the buried soil keys) and `"truncated"` (a red apedal B at the surface over a dorbank: Garies, truncated). They are not real profiles.

Usage

```
make_za1991_example(form = "hutton", notation = FALSE, family = NULL)
```

Arguments

`form`	The form, in lower case, or a special case.
`notation`	Write the name of each horizon in `diagnostic_za` (the South African notation) besides the data (default `FALSE`).
`family`	Optional family code of the form (`"Hu1200"`, or its four digits `"1200"`); `NULL` (default) for the first family.

Details

Since v0.9.359 each form example also holds the classes of one of its families: the first of the form by default (Hu1100 for Hutton), or the family named in `family`, made by changing the layer each family criterion reads (its colour, clay, cations, structure, carbonates, consistence) or adding the layer it reads (rock under the organic O, a podzolic layer under the yellow-brown apedal B of Constantia).

Value

A PedonRecord.

Examples

```
classify_za1991(make_za1991_example("clovelly"), on_missing = "silent")$name
classify_za1991(make_za1991_example("hutton", family = "Hu2200"), on_missing = "silent")$name
```

`mineral_material` *Mineral material (WRB 2022 Ch 3.3.11): < 20% SOC AND < 35% volume artefacts containing >= 20% organic carbon. The complement of organic_material / organotechnic_material.*

Description

Mineral material (WRB 2022 Ch 3.3.11): < 20% SOC AND < 35% volume artefacts containing >= 20% organic carbon. The complement of organic_material / organotechnic_material.

Usage

```
mineral_material(pedon, max_oc = 20, max_organotechnic = 35)
```

Arguments

pedon	A PedonRecord.
max_oc	Numeric threshold or option (see Details).
max_organotechnic	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`mollic` *Mollic horizon (WRB 2022)*

Description

Tests whether any near-surface horizon meets the mollic horizon criteria. The mollic horizon is the diagnostic surface horizon of Chernozems, Phaeozems, Kastanozems, and several other RSGs; it indicates a thick, dark, base-rich, organic-matter-enriched topsoil formed under steppe or comparable vegetation.

Usage

```
mollic(
  pedon,
  min_thickness = 20,
  min_oc = 0.6,
  min_bs = 50,
  surface_top_cm = 5
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness in cm (default 20).
`min_oc`	Minimum SOC % (default 0.6).
`min_bs`	Minimum base saturation % (default 50).
`surface_top_cm`	Maximum top depth (cm) for a horizon to be considered "surface-related" (default 5). v0.1 uses this as a proxy for the WRB rule that mollic must form continuously from the soil surface (after mixing of upper 20 cm if required).

Details

Since v0.9.314 the criteria are read as WRB 2022 Ch 2.1 (p. 23) sets them for a horizon of several subhorizons: every criterion except the thickness in every layer, unless the text says otherwise. Criteria 1 (structure; structure not recorded does not block it), 2 (>= 0.6% SOC) and 3 (colour, items a-c; a plough layer takes the colour of the subhorizon below it) are read in each layer, and the horizon is the run of such layers from the mineral soil surface. Criterion 5, the base saturation, is the weighted average over the run (by pH where it is not recorded, Annex 2 Table 9.2), and criterion 6 its thickness: >= 20 cm (6b), or since v0.9.315 >= 10 cm directly over continuous rock, technic hard material or a cryic, petrocalcic, petroduric, petrogypsic or petroplinthic horizon (6a, p. 53; the umbric horizon names cryic, petroduric and petroplinthic, p. 70). A moist value with no dry value recorded reads the dry value as the moist value + 1; a layer with no colour recorded at all and >= 1.5% SOC is read as dark (Hugo, 2026-10-29). A dry colour alone leaves criterion 3 undecided unless its value is above 5. The colour subtest records where each layer's reading came from.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition. International Union of Soil Sciences, Vienna. Chapter 3 – Mollic horizon.

`mudanca_textural_abrupta`

Mudanca textural abrupta (SiBCS Cap 1, p 30-31)

Description

"Consiste em um consideravel aumento no teor de argila em pequena distancia na zona de transicao entre o horizonte A ou E e o horizonte subjacente B":

- A or E with less than 200 g/kg of clay: the B has at least twice as much;
- A or E with 200 g/kg or more: the B has 200 g/kg more or above ("de 300 para no minimo 500, ou de 220 para no minimo 420"),

"em uma distancia vertical <= 7,5 cm".

Usage

```
mudanca_textural_abrupta(pedon)
```

Arguments

pedon	A PedonRecord.

Details

Since v0.9.244 the test is that of the book: the first B horizon is compared with the horizon immediately above it, which has to be an A or an E; and the distance is read from the distinctness of the transition (`boundary_distinctness` of the horizon above: abrupt, under 2.5 cm, or clear, 2.5 to 7.5 cm). Where it is recorded as gradual or diffuse there is no abrupt change; where it is not recorded the two horizons are compared. The layer returned is the first B horizon. A BA is a B horizon here and a BC is not, as in item (h) of the B textural.

Until v0.9.243 the function was the WRB `abrupt_textural_difference`, which compares every pair of adjoining layers, in the order of the rows: a jump of clay between two B horizons, or from a B to a C, was an abrupt change. A profile with no horizon designation at all is still tested that way.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

mulmic_material — *Mulmic material (WRB 2022 Ch 3.3.12): mineral material developed from organic material; \>= 8% SOC, with low BD, structural / chroma criteria.*

Description

Mulmic material (WRB 2022 Ch 3.3.12): mineral material developed from organic material; \>= 8% SOC, with low BD, structural / chroma criteria.

Usage

```
mulmic_material(pedon, min_oc = 8, max_chroma = 2)
```

Arguments

pedon	A PedonRecord.
min_oc	Numeric threshold or option (see Details).
max_chroma	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

natric_horizon — *Natric horizon (WRB 2022)*

Description

Tests for the natric horizon: an argic horizon with diagnostic sodium accumulation. Diagnostic of Solonetz.

Usage

```
natric_horizon(pedon, min_esp = 15, min_pH_h2o = 7)
```

Arguments

pedon	A PedonRecord.
min_esp	Minimum ESP % (default 15).
min_pH_h2o	Minimum pH(H2O) for the ESP-only path (default 7.0; alkaline gate to exclude false-positive acidic Bt horizons).

Details

v0.9.310: WRB 2022 Ch 3.1.21 (p. 53-54) criterion 3 (columnar or prismatic structure, or blocky with tongues of the overlying layer) and criterion 4 (ESP $\geq$ 15 throughout the upper 40 cm, or Mg + Na > Ca + exchange acidity there with ESP $\geq$ 15 within 50 cm). Until then an ESP $\geq$ 15 in any argic layer sufficed; the USDA natric helpers keep that reading.

Value

A `DiagnosticResult`.

v0.9.76 designation + ESP-only inference (opt-in)

Field-described Solonetz profiles in NCSS / KSSL data routinely reach the natric ESP threshold (computed from `na_cmol` / `cec_cmol`) without satisfying the strict `argic()` clay-increase test, because surveyors record `Btk`-suffix designations (carbonates dominate the horizon designation choice) rather than `Btn`/`Bn` or `clay_pct` is missing.

With `options(soilKey.natric_designation_inference = TRUE)` the function accepts a layer as natric when the canonical argic test returns NA or FALSE AND *either*:

1. the designation matches `[A-Z][a-z0-9]*n` (an `n` master-letter modifier in the horizon name – e.g.\ `Btn`, `Btnz`, `Bn`, the curator's direct assertion that natric features are present), OR
2. ESP >= `min_esp` on a B-prefixed subsoil layer (`top_cm` > `20`) AND the layer's pH(H2O) >= 7 (alkaline – typical of true natric, excludes acidic Bt horizons that happen to read high Na from sea-spray).

Default is `FALSE` (canonical behaviour preserved).

References

IUSS Working Group WRB (2022), Chapter 3, Natric horizon.

`nitic_horizon` *Nitic horizon (WRB 2022)*

Description

Tests for a nitic horizon, WRB 2022 Chapter 3.1.22, criterion by criterion on each layer:

1. $\geq$ 30% clay (`min_clay`);
2. moderate to strong angular or subangular blocky structure breaking into polyhedral or flat-edged second-level structure, or polyhedral structure, with pressure faces (shiny surfaces) on $\geq$ 25% of the aggregate surfaces. Read from `structure_type`, `structure_grade`, `structure_type_secondary` and `pressure_faces`, "common" or more (since v0.9.317; until then from clay films, which WRB 2022 keeps apart from pressure faces);
3. $\geq$ 4% Fe-dith (`fe_dcb_pct`), $\geq$ 0.2% Fe-ox (`fe_ox_pct`) and Fe-ox / Fe-dith $\geq$ 0.05;
4. not part of a plinthic horizon (`plinthic`);

5. the layers that meet 1-4 are $\geq$ 30 cm thick together, as one horizon (`min_thickness`).

A criterion is `FALSE` when the data contradict it and `NA` when the data it needs are missing: a profile without Fe-ox or structure records gets `NA`, not a nitic horizon. With `options(soilKey.morphological_inference = TRUE)` missing structure and Fe data are read from a Bt designation with a CEC of 8-36 cmolc/kg clay and no albic E above (the v0.9.18 reading of legacy profiles), recorded in the evidence.

Usage

```
nitic_horizon(
  pedon,
  min_clay = 30,
  min_fe_dcb = 4,
  min_thickness = 30,
  min_fe_ox = 0.2,
  min_feox_fedith = 0.05,
  max_clay_drop_pct = NULL,
  max_decrease_depth = NULL
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_clay`	Minimum clay % (default 30).
`min_fe_dcb`	Minimum dithionite Fe % (default 4).
`min_thickness`	Minimum thickness in cm (default 30).
`min_fe_ox`	Minimum oxalate Fe % (default 0.2).
`min_feox_fedith`	Minimum Fe-ox / Fe-dith ratio (default 0.05).
`max_clay_drop_pct`, `max_decrease_depth`	Unused since v0.9.220, kept so that existing calls still run.

Details

Until v0.9.219 the structure and shiny-surface tests never vetoed, Fe-ox was not read, a plinthic horizon was not excluded, any ferralic horizon in the profile excluded a nitic one (WRB 2022 does not: Nitisols key out before Ferralsols), the Fe inference above ran by default, and a clay drop of more than 8 percentage points within 50 cm vetoed it. The clay rule of the Nitisol key is in `nitisol`.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 3.1.22, Nitic horizon.

nitisol *Nitisol RSG gate (WRB 2022)*

Description

The Nitisol entry of the WRB 2022 key (Chapter 4): "1. a nitic horizon starting $\leq$ 100 cm from the mineral soil surface; and 2. from the mineral soil surface to the nitic horizon, a clay content that is at least half of the weighted average clay content of the nitic horizon; and 3. no vertic horizon starting above or at the upper limit of the nitic horizon." Criterion 2 is read layer by layer; a vertic horizon soilKey cannot establish does not exclude.

Usage

```
nitisol(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 4, Nitisols.

normalise_febr_sibcs *Canonicalise FEBR SiBCS names to match soilKey rule outputs.*

Description

FEBR ships SiBCS labels in mixed legacy/modern form (`"Podzolicos"` for old name of Argissolos, singular vs plural, Portuguese accents). This helper folds them to the form produced by `run_sibcs_key()` so that benchmark accuracies can be computed without false negatives.

Usage

```
normalise_febr_sibcs(x, level = c("order", "subordem"))
```

Arguments

x	Character vector of FEBR SiBCS names.
level	One of `"order"` (default) or `"subordem"`.

Details

A label is read by its first word, with any punctuation around it dropped (`"ESPODOSSOLO."` is Espodossolos). Since v0.9.225 the result must be one of the 13 orders of SiBCS 5 after the legacy map; otherwise it is `NA`. Until then the first word came out whatever it was, so `"CLASSIFICACAO ? CAMBISSOLO ..."` became the class "Classificacaos", `"ALISSOLO CROMICO"` "Alissolos" and `"LATOSOL AMARELO"` "Latosois": labels no prediction can equal, which the benchmarks counted as scored profiles and as errors (631 of the 7,694 labels of the national BDsolos export, in 79 forms).

Since v0.9.226, when the first word gives no order the label is read once more as an old name: a leading "Solo" or "Solos" is set aside ("SOLO LITOLICO" is Neossolos, "SOLOS ORGANICOS" Organossolos), a hyphen separates words as a space does ("LATOSSOLO-VERMELHO", "SOLONETZ-SOLODIZADO"), and the names of `.SIBCS_LEGACY_TWO_WORD_MAP` ("TERRA ROXA ESTRUTURADA", "BRUNO NAO CALCICO") are looked up before those of `.SIBCS_LEGACY_ORDER_MAP`. A label the first word already places is read as before. At `level = "subordem"` an old name gives its order followed by the next word of the label, which is not a suborder of SiBCS 5 in general: the old names are for comparisons at the order level.

Since v0.9.229 the order of the name is not the last word when the label states attributes that SiBCS 5 places elsewhere: a Podzolico Vermelho-Amarelo or Vermelho-Escuro, or a label that begins with "Argissolo", that states "Eutrofico" and high-activity clay is Luvissolos, and a Cambissolo in the old form that states "Eutrofico", high-activity clay and "A chernozemico" is Chernossolos. See `.sibcs_stated_order()`.

Value

Character vector of normalised SiBCS names; `NA` for labels that are out-of-scope for the comparison: any label that is not an order of SiBCS 5 after the legacy maps (old names without an entry there, "Solos ..." followed by none of them, names of other systems, text that is not a class).

See Also

`normalise_febr_wrb`, `normalise_febr_usda`

`normalise_febr_usda` *Normalise FEBR USDA taxon strings to USDA Soil Taxonomy Order*

Description

FEBR ships USDA Soil Taxonomy labels at the subgroup or great-group granularity (e.g. "TYPIC HAPLUDULT", "ACRUSTOX"). The suffix of the final word encodes the Order: `...OX` -> Oxisols, `...ULT` -> Ultisols, `...EPT` -> Inceptisols, etc. This helper extracts the Order from the suffix so the benchmark can compare against `classify_usda()$rsg_or_order` at `level = "order"`.

Usage

```
normalise_febr_usda(x)
```

Arguments

x	Character vector of FEBR USDA names.

Value

Character vector of normalised Order names ("Oxisols", "Ultisols", "Inceptisols", ...).

---

normalise_febr_wrb *Normalise FEBR WRB taxon strings to RSG-only*

Description

FEBR ships WRB names with full qualifier strings, e.g. "HUMIC FERRALSOL", "HAPLIC ACRISOL (ALUMIC, HYPERDYSTRIC, ...)". The trailing word (before any qualifier parens) is the RSG. This helper extracts and normalises it to soilKey's plural Title Case form ("Ferralsols", "Acrisols"), matching `ClassificationResult$rsg_or_order`.

Usage

```
normalise_febr_wrb(x)
```

Arguments

x	Character vector of FEBR WRB names.

Value

Character vector of normalised RSG names.

---

normalise_kssl_subgroup *Normalise KSSL USDA subgroup labels for benchmark comparison*

Description

KSSL stores 'samp_taxsubgrp' in lower-case, space-separated form ("typic hapludalfs", "aquic argiudolls"). soilKey's 'classify_usda()' returns Title Case names ("Typic Hapludalfs"). The benchmark runner at 'level = "subgroup"' lowercases both sides and trims whitespace, but this helper makes the normalisation explicit when users want to compare KSSL labels against arbitrary classifier output. Idempotent.

Usage

```
normalise_kssl_subgroup(x)
```

Arguments

x	Character vector of KSSL subgroup names.

Value

Lowercase, single-space-separated vector.

---

npdb_dry_colour_model *Dry colour of A horizons from their moist colour*

Description

Fits, on the A horizons of a set of pedons that record both colours, a linear model of the dry Munsell value on the moist value and of the dry chroma on the moist chroma, and estimates its error by cross-validation grouped by pedon (the horizons of a pedon stay in one fold): the range of the out-of-fold errors that holds `level` of them (from its lower to its upper tail). The pedons are dealt to the folds in turn, in the order given, so the result does not depend on the random number generator. On the NPDB the error is close to one value unit, so the estimate decides the dry value of 5.5 of the chernozemic A for most dark A horizons but seldom the limits of 3.5 and 4.5 between the Black, Dark Brown and Brown Chernozems.

Usage

```
npdb_dry_colour_model(pedons, level = 0.8, folds = 10L)
```

Arguments

pedons	A list of PedonRecord.
level	Share of the cross-validated errors the range holds.
folds	Number of folds.

Value

A list of class `npdb_dry_colour_model`: `value` and `chroma` (each with `coef`, `band` = the error range, `rmse` out of fold, `n` horizons), `level`.

See Also

npdb_scenario

npdb_scenario *A scenario for the NPDB: estimated dry colour and soil climate*

Description

Adds to NPDB pedons what the records lack and the Canadian key reads for the Chernozems (and the Gray Brown Luvisols and some subgroups), each as a range, for benchmark_npdb:

- the dry value and chroma of the A horizons that record a moist colour and no dry one, from dry_colour (a npdb_dry_colour_model), with the range of its error;
- the mean annual soil temperature, as the range of npdb_soil_climate (from the air temperature to 4 degrees above it; kept as a range only where it crosses 0 or 8 degrees, the limits the key reads), and the moisture subclass by Thornthwaite, with the subclass by Hargreaves and every subclass in between where they differ.

The pedons hold the central estimate (the middle of the soil temperature range) in the usual fields (munsell_value_dry, munsell_chroma_dry, site$mean_annual_soil_temp_C, site$soil_moisture_subclass), and the Thornthwaite moisture index in site$npdb_cmi; recorded values are never replaced. The ranges are kept in site$npdb_scenario (a second call adds to them); a pedon with no site$id, or one not in climate, gets no climate. benchmark_npdb classifies such a pedon at every combination of the ends of its ranges and counts a class as decided only where they all give it; otherwise the result is provisional.

Usage

```
npdb_scenario(pedons, climate = NULL, dry_colour = NULL)
```

Arguments

pedons	A list of PedonRecord from load_npdb_pedons.
climate	A data frame from npdb_soil_climate (its id matching site$id), or NULL.
dry_colour	A npdb_dry_colour_model, or NULL.

Value

The pedons (copies), with the estimates.

Examples

```
## Not run:
peds <- load_npdb_pedons("path/to/npdb_v3_csv")
nm <- fetch_aafc_normals(vapply(peds, function(p) p$site$lat, 0),
                         vapply(peds, function(p) p$site$lon, 0),
                         id = names(peds))
sc <- npdb_scenario(peds, climate = npdb_soil_climate(nm),
                    dry_colour = npdb_dry_colour_model(peds))
benchmark_npdb(sc, level = "order", set_aside = character(0))

## End(Not run)
```

`npdb_soil_climate` *Soil climate of the Canadian system from climate normals*

Description

Estimates, from monthly climate normals (as returned by `fetch_aafc_normals`), what the Canadian System of Soil Classification reads as the soil climate (Ch 2, p. 16, criterion 7 of the chernozemic A; the soil temperature classes and moisture subclasses of its Tables 3 and 4, after the map Soil Climates of Canada, Clayton et al. 1977):

- the mean annual air temperature (MAAT, the mean of the twelve monthly means) and, for the mean annual soil temperature (MAST), the range from the MAAT to the MAAT plus `mast_offset` degrees: in Canada the soil under snow is warmer than the air;
- the climatic moisture index (CMI) of Table 4, the growing-season precipitation as a percentage of the water crops would use: here the precipitation over the potential evapotranspiration (PET) of the months of the growing season (April to October) whose mean temperature is above 5 degrees. The PET is computed in two ways: by Thornthwaite (monthly mean temperatures and day length), the central estimate, which matches the zones the book gives the Chernozem great groups on the NPDB (Brown subarid to semiarid, Dark Brown semiarid, Black and Dark Gray subhumid); and by Hargreaves and Samani (monthly minimum and maximum temperatures and extraterrestrial radiation), the other end of the range (on the NPDB mostly drier, sometimes wetter);
- the moisture subclass from each index (Table 4): perhumid above 84, humid 74-84, subhumid 59-73, semiarid 46-58, subarid 25-45, arid below 25.

Usage

```
npdb_soil_climate(normals, mast_offset = c(0, 4))
```

Arguments

`normals`	A data frame with `id`, `lat` and the columns of `fetch_aafc_normals`.
`mast_offset`	The range added to the MAAT for the MAST, in degrees (two numbers, read from the lower to the higher).

Value

A data frame: `id`, `maat`, `mast_lo`, `mast_hi`, `p_gs` (mm), `gs_months`, `pet_gs` and `cmi` and `moisture_subclass` (Thornthwaite), `pet_gs_hargreaves`, `cmi_hargreaves` and `moisture_subclass_hargreaves`.

References

Soil Classification Working Group (1998). The Canadian System of Soil Classification, 3rd ed., Tables 3 and 4. Clayton, J.S. et al. (1977). Soils of Canada. Thornthwaite, C.W. (1948). Geographical Review 38: 55-94. Hargreaves, G.H., Samani, Z.A. (1985). Applied Engineering in Agriculture 1: 96-99. Allen, R.G. et al. (1998). FAO Irrigation and Drainage Paper 56.

nsd_citation *Citation for data read from the New Zealand National Soils Database*

Description

Returns the acknowledgement for profiles of the National Soils Database (NSD) of Manaaki Whenua - Landcare Research, as exported by the National Soil Data Repository (NSDR) and read by `read_nsd_pedons`.

Usage

```
nsd_citation(accessed = NULL)
```

Arguments

`accessed`	Optional date the survey exports were downloaded.

Value

A character string.

Examples

```
nsd_citation()
```

ollama_is_running *Is the local Ollama HTTP API reachable?*

Description

Probes `http://127.0.0.1:11434/api/tags` (the standard Ollama endpoint) with a short HTTP HEAD-style GET. Returns `TRUE` only if the request returns HTTP 200 in under `timeout_s` seconds. Used by `vlm_pick_provider` for the `provider = "auto"` fallback chain. Override the URL via `options(soilKey.ollama_url = "http://host:port")`.

Usage

```
ollama_is_running(url = NULL, timeout_s = 1.5)
```

Arguments

`url`	Override URL to probe (default reads `getOption("soilKey.ollama_url", default = "http://127.0.0.1:11434/api/tags")`).
`timeout_s`	Request timeout in seconds (default 1.5).

Value

Logical scalar.

`organic_material` *Organic material (WRB 2022 Ch 3.3.13): \>= 20% SOC + recognisability criteria. v0.3.3: SOC threshold only.*

Description

Organic material (WRB 2022 Ch 3.3.13): \>= 20% SOC + recognisability criteria. v0.3.3: SOC threshold only.

Usage

```
organic_material(pedon, min_oc = 20)
```

Arguments

`pedon`	A `PedonRecord`.
`min_oc`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`organotechnic_material` *Organotechnic material (WRB 2022 Ch 3.3.14): \>= 35% volume of artefacts that themselves contain \>= 20% organic C. Soil itself has < 20% SOC.*

Description

Organotechnic material (WRB 2022 Ch 3.3.14): \>= 35% volume of artefacts that themselves contain \>= 20% organic C. Soil itself has < 20% SOC.

Usage

```
organotechnic_material(pedon, min_artefacts = 35, max_oc = 20)
```

Arguments

`pedon`	A `PedonRecord`.
`min_artefacts`	Numeric threshold or option (see Details).
`max_oc`	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`ornithogenic_material` *Ornithogenic material (WRB 2022 Ch 3.3.15): bird-influenced topsoil. Mehlich-3 P >= 750 mg/kg + designation pattern* `Aornit|Bornit`.

Description

Ornithogenic material (WRB 2022 Ch 3.3.15): bird-influenced topsoil. Mehlich-3 P >= 750 mg/kg + designation pattern `Aornit|Bornit`.

Usage

```
ornithogenic_material(pedon, min_p_mehlich3 = 750)
```

Arguments

`pedon`	A PedonRecord.
`min_p_mehlich3`	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`ossl_demo_sa` *Synthetic OSSL South America demo subset*

Description

A small, deterministic, OSSL-shaped artefact for use in vignettes, examples and tests when the real Open Soil Spectral Library data is not available (no network, sensitive deployment, CI). The object has the canonical `list(Xr, Yr, metadata)` shape consumed by predict_ossl_mbl / fill_from_spectra, so the in-package demo path is identical to the real-data path.

Usage

```
ossl_demo_sa
```

Format

A list with three elements:

`Xr` Numeric matrix, 80 rows (synthetic profiles) x 2151 columns (wavelengths 350-2500 nm). Reflectance values in [0.05, 0.85].

`Yr` Data frame, 80 rows x 9 columns (`clay_pct`, `sand_pct`, `silt_pct`, `cec_cmol`, `bs_pct`, `ph_h2o`, `oc_pct`, `fe_dcb_pct`, `caco3_pct`). Property ranges follow the OSSL global summary statistics.

`metadata` Named list with provenance information (`region`, `n_profiles`, `snapshot`, `seed`, `note`, ...).

Details

This is a **synthetic** placeholder: the spectra are generated from a tropical-soil baseline plus property-correlated absorption bands (1400 nm OH-water, 1900 nm clay-OH, 2200 nm Al-OH, 900 nm Fe-oxide) with deterministic noise. It is *not* a substitute for real OSSL measurements. For paper-grade work, populate a real OSSL artefact via:

```
ossl_lib <- download_ossl_subset(region = "south_america")
```

Re-build the demo with `source("data-raw/build_ossl_demo.R")`.

Source

Synthetic; built by `data-raw/build_ossl_demo.R` with seed 20260430. The OSSL property ranges that drove the simulation come from Sanderman, J. *et al.* (2024), *Open Soil Spectral Library*, `https://soilspectroscopy.org/`.

Examples

```
data(ossl_demo_sa)
dim(ossl_demo_sa$Xr)
#> [1]   80 2151
head(ossl_demo_sa$Yr)

## Not run:
# Use it as the ossl_library argument to predict_ossl_mbl():
pedon <- make_synthetic_pedon_with_spectra()
fill_from_spectra(pedon,
                  library      = "ossl",
                  method       = "mbl",
                  ossl_library = ossl_demo_sa)

## End(Not run)
```

`ossl_library_template` *Canonical schema for an 'ossl_library' object*

Description

`predict_ossl_mbl` and `predict_ossl_plsr_local` take an `ossl_library` argument that must be a list with two named elements:

Usage

```
ossl_library_template(
  wavelengths = 350:2500,
 properties = c("clay_pct", "sand_pct", "silt_pct", "cec_cmol", "bs_pct", "ph_h2o",
    "oc_pct", "fe_dcb_pct", "caco3_pct")
)
```

Arguments

`wavelengths`	Integer vector of wavelengths (default `350:2500` nm for Vis-NIR/SWIR).
`properties`	Character vector of property column names to seed the empty `Yr` data.frame with.

Details

- `Xr`: numeric matrix, rows = OSSL training spectra, columns = wavelengths. Must align (after preprocessing) with the column space used by the spectra you predict on.
- `Yr`: data.frame keyed by property name (e.g. `clay_pct`, `cec_cmol`), one row per training spectrum.

This function returns an empty template you can populate from a real OSSL extract (e.g. via the `ossl-import` Python package or the public S3 mirror at `https://storage.googleapis.com/soilspec4gg-public/`).

soilKey does **not** bundle OSSL data; until you populate this template with real values, all 'predict_ossl_*' calls fall back to the deterministic synthetic predictor (which prints a warning).

Value

A list with `Xr` (a 0-row matrix of the right column dimension) and `Yr` (an empty data.frame with the requested columns).

oxic_usda *Oxic horizon (USDA Soil Taxonomy)*

Description

The USDA oxic horizon is the diagnostic of Oxisols. Its central criteria match the WRB 2022 ferralic horizon closely enough that v0.2 simply delegates: every fixture that classifies as Oxisol via USDA also classifies as Ferralsol via WRB and vice-versa. The fine-grained differences (USDA's water-dispersible-clay test, the sand-fraction weatherable-mineral cut-offs) are tracked in the diagnostics.yaml for v0.8 refinement.

Usage

```
oxic_usda(pedon, ...)
```

Arguments

pedon	A PedonRecord.
...	Passed to ferralic.

Details

The CEC limit is inclusive here: KST 13 (Ch. 3, oxic horizon, criterion 6) asks for "an apparent CEC of 16 cmol(+) or less per kg clay", where WRB 2022's ferralic horizon asks for "< 16". Since v0.9.212 `ferralic()` follows WRB and this function passes `cec_inclusive = TRUE`, so a horizon at exactly 16 is oxic but not ferralic.

Value

A DiagnosticResult (with `name = "oxic_usda"`).

References

Soil Survey Staff (2014). *Keys to Soil Taxonomy*, 12th edition. USDA-NRCS, Washington DC. Chapter 3 – Diagnostic Horizons; oxic.

panpaic *Panpaic horizon (WRB 2022 Ch 3.1)*

Description

From Quechua *p'anpay* = "to bury". A buried diagnostic horizon (any horizon whose original surface was subsequently overlain by younger material). Used by the Panpaic qualifier and by the Cambisols / Anthrosols branches.

Usage

```
panpaic(pedon)
```

Arguments

pedon	A PedonRecord.

Details

Since v0.9.220 the four criteria of WRB 2022 Ch 3.1.23 on a buried surface horizon (an A designation with a b suffix, or below a lithic discontinuity, 2A): $\geq$ 0.2% SOC; SOC $\geq$ 25% (relative) and $\geq$ 0.2% (absolute) higher than in the overlying layer; a lithic discontinuity at its upper limit; $\geq$ 5 cm thick. Until v0.9.219 any designation containing a b passed, AB included.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

PedonRecord *PedonRecord: structured representation of a single pedon*

Description

PedonRecord: structured representation of a single pedon

PedonRecord: structured representation of a single pedon

Details

The central data carrier in soilKey. A PedonRecord bundles everything we know about one soil profile: site metadata, the horizons table (with a fixed canonical schema — see horizon_column_spec), optional Vis-NIR/MIR spectra, profile photographs, source documents, and a provenance log that records, per (horizon, attribute) pair, where each value came from (measured, extracted_vlm, predicted_spectra, inferred_prior, user_assumed).

All diagnostic functions (argic, ferralic, mollic, ...) consume a PedonRecord directly. The provenance log is what allows the final ClassificationResult to assign a meaningful evidence grade.

Horizon columns and site properties that soilKey does not recognise are kept but read by no key. PedonRecord$new() warns about them (with the closest recognised names) when it is called from outside the package; see unrecognised_attributes, horizon_attribute_spec and site_attribute_spec. options(soilKey.warn_unrecognised = FALSE) turns the warning off.

Value

An R6 object of class PedonRecord.

Public fields

`site` List. Site-level metadata: `lat`, `lon`, `crs` (default 4326), `date`, `country`, `elevation_m`, `slope_pct`, `aspect_deg`, `landform`, `parent_material`, `land_use`, `vegetation`, `drainage_class`, plus an arbitrary `id`.

`horizons` data.table with the canonical horizon schema.

`spectra` List with optional `vnir` matrix (rows = horizons, cols = wavelengths in nm), `mir` matrix, and `metadata` list.

`images` List of named lists describing profile photographs.

`documents` List of named lists describing source documents.

`provenance` data.table with columns `horizon_idx`, `attribute`, `source`, `confidence`, `notes`.

Methods

Public methods:

- `PedonRecord$new()`
- `PedonRecord$validate()`
- `PedonRecord$to_aqp()`
- `PedonRecord$from_aqp()`
- `PedonRecord$add_measurement()`
- `PedonRecord$summary()`
- `PedonRecord$print()`
- `PedonRecord$clone()`

Method `new()`**:** Construct a PedonRecord.

Usage:

```
PedonRecord$new(
  site = NULL,
  horizons = NULL,
  spectra = NULL,
  images = NULL,
  documents = NULL,
  provenance = NULL
)
```

Arguments:

`site` List of site-level metadata.

`horizons` data.frame/data.table of horizons.

`spectra` Optional list with `vnir`, `mir`, `metadata`.

`images` Optional list of image descriptors.

`documents` Optional list of document descriptors.

`provenance` Optional provenance data.table; if NULL, an empty one is created.

Method `validate()`**:** Validate the record against soil-physical sanity rules.

Checks: top < bottom for every horizon; no overlapping depths; clay+silt+sand sum to 100 ± 2 where all three are reported; pH values plausible (1..12); CEC >= sum of exchangeable bases (Ca, Mg, K, Na); Munsell value/chroma in plausible ranges; coarse fragments percent in [0, 100]; OC geographic ranges. Returns a list with `valid`, `errors`, `warnings`, `n_horizons`.

Usage:

```
PedonRecord$validate(strict = FALSE, verbose = TRUE)
```

Arguments:

`strict` If `TRUE`, throws on errors instead of returning.

`verbose` If `TRUE`, prints messages via cli.

Returns: Invisibly, a list summarising the validation outcome.

Method `to_aqp()`**:** Coerce to an aqp `SoilProfileCollection`.

Usage:

```
PedonRecord$to_aqp()
```

Returns: A `SoilProfileCollection`. Requires the aqp package.

Method `from_aqp()`**:** Populate this record from an aqp `SoilProfileCollection`.

Usage:

```
PedonRecord$from_aqp(spc, top_col = "top_cm", bottom_col = "bottom_cm")
```

Arguments:

`spc` A `SoilProfileCollection`.

`top_col` Name of the top-depth column in `spc` (mapped to `top_cm`).

`bottom_col` Name of the bottom-depth column (mapped to `bottom_cm`).

Returns: Invisibly self (mutated in place).

Method `add_measurement()`**:** Add a measurement (or extracted/predicted value) and record its provenance.

Usage:

```
PedonRecord$add_measurement(
  horizon_idx,
  attribute,
  value,
  source = "measured",
  confidence = 1,
  notes = NA_character_,
  overwrite = FALSE
)
```

Arguments:

`horizon_idx` Integer horizon index (1-based).

`attribute` Name of the horizon column to set.

`value` New value for that cell.

`source` One of "measured", "extracted_vlm", "predicted_spectra", "inferred_prior", "user_assumed".

`confidence` Numeric in [0, 1].

`notes` Optional free-text note.

`overwrite` If `FALSE` (default) and the cell already has a value from a more authoritative source, leave it alone. If `TRUE`, overwrite.

Returns: Invisibly self.

Method `summary()`**:** Compact summary list (for serialization or testing).

Usage:

```
PedonRecord$summary(...)
```

Arguments:

`...` Ignored (S3 summary signature compatibility).

Method `print()`**:** Pretty-print the record.

Usage:

```
PedonRecord$print(...)
```

Arguments:

`...` Ignored (S3 print signature compatibility).

Method `clone()`**:** The objects of this class are cloneable with this method.

Usage:

```
PedonRecord$clone(deep = FALSE)
```

Arguments:

`deep` Whether to make a deep clone.

Examples

```
# The canonical fixtures return ready-built PedonRecords:
pedon <- make_ferralsol_canonical()
pedon$site$id
nrow(pedon$horizons)
```

`pedons_from_spectral_table` *Build PedonRecords with attached Vis-NIR/MIR spectra from a table*

Description

Groups a reflectance + metadata table by profile and returns one `PedonRecord` per profile, with each profile's sample rows stacked into `$spectra$vnir` (rows = horizons, cols = wavelengths) and the lab attributes / depths written to the horizons. Taxonomic labels are stored in `$site` (`reference_wrb` / `reference_sibcs` / `reference_st`). These pedons are the query objects for `classify_*(gapfill = list(method = "spectra", ossl_library = <lib>))`.

Usage

```
pedons_from_spectral_table(
  reflectance,
  metadata,
  id_col = "id",
  profile_col = NULL,
  wavelengths = NULL,
```

```
  resample_to = NULL,
  property_map = NULL,
  label_map = NULL,
  normalize = c("auto", "none", "percent"),
  keep_properties = FALSE,
  verbose = TRUE
)
```

Arguments

`reflectance`	Reflectance data: a matrix / data.frame with rows = samples and columns named by wavelength (nm); OR a long data.frame with `id_col`, `wavelength_nm`, `reflectance`; OR a path to a CSV in either form.
`metadata`	A data.frame with one row per sample carrying `id_col` plus lab attributes and optional taxonomic labels and `lat`/ `lon`. Rows are aligned to `reflectance` by `id_col`.
`id_col`	Sample identifier column shared by both tables (default `"id"`).
`profile_col`	Column grouping samples into profiles (default `id_col`: one profile per sample, e.g. a topsoil library).
`wavelengths`	Optional explicit wavelength vector (nm) when the reflectance columns are not wavelength-named.
`resample_to`	Optional target wavelength grid (nm) to linearly resample every spectrum onto (e.g. `350:2500`); default keeps the native grid.
`property_map, label_map`	Optional named lists overriding the alias auto-detection, e.g. `property_map = list(clay_pct = "ARGILA")`.
`normalize`	One of `"auto"` (divide by 100 when values look like percent), `"percent"`, or `"none"`.
`keep_properties`	If `TRUE`, also write the mapped lab attributes to the horizons (default `FALSE` – a field pedon usually has only the scan, which is the scenario the spectral fill targets).
`verbose`	Print a one-line summary (default `TRUE`).

Value

A list of `PedonRecord` objects.

See Also

`read_spectral_library`, `benchmark_spectral_fill`

pedon_json_schema *JSON Schema for a soilKey PedonRecord*

Description

Returns a Draft-2020-12 JSON Schema describing the canonical `PedonRecord` structure: a `site` object with site-level metadata plus a `horizons` array where each element matches the canonical horizon schema documented by `horizon_column_spec`.

Usage

```
pedon_json_schema(as = c("list", "json"), pretty = TRUE)
```

Arguments

`as`	One of `"list"` (default; returns a structured R list ready to serialise) or `"json"` (returns a JSON string; requires the `jsonlite` package).
`pretty`	Logical, only used for `as = "json"`.

Value

A list (default) or a JSON string.

Examples

```
## Not run:
schema <- pedon_json_schema()
names(schema)
#> [1] "$schema"     "$id"         "title"       "type"        "required"   "properties"

# Validate a JSON profile against the schema:
if (requireNamespace("jsonvalidate", quietly = TRUE)) {
  schema_json <- pedon_json_schema(as = "json")
  jsonvalidate::json_validate('{"site":{...},"horizons":[...]}',
                                  schema_json, engine = "ajv")
}

## End(Not run)
```

pedon_to_spc *Convert a soilKey PedonRecord to an aqp SoilProfileCollection*

Description

The mapping respects aqp's expected column conventions and sets the metadata required by `getArgillicBounds()`, `getCambicBounds()`, and `mollicEpipedon()`:

Usage

```
pedon_to_spc(pedon)
```

Arguments

pedon	A PedonRecord.

Details

- `id` from `pedon$site$id`
- `top` / `bottom` from `top_cm` / `bottom_cm`
- `name` (designation) from `designation`
- `texcl` (texture class) derived via `texture_class_from_pct`
- `clay`, `silt`, `sand` from `clay_pct` / `silt_pct` / `sand_pct`
- `m_hue`, `m_value`, `m_chroma`, `d_value`, `d_chroma` from `munsell_*_moist` and `munsell_*_dry`

Internal use; the soilKey diagnostics call this on the fly when `engine = "aqp"`. Direct use is supported for users who want to plug additional aqp algorithms (`slab`, `slice`, `glom`) into a soilKey workflow.

Value

A `aqp::SoilProfileCollection` with one site (the pedon) and one row per horizon.

petrocalcic *Petrocalcic horizon (WRB 2022)*

Description

A continuously cemented variant of the calcic horizon. Same chemistry (CaCO3 \>= 15%) plus moderate-or-greater cementation in at least 50% of the layer.

Usage

```
petrocalcic(pedon, min_thickness = 10, min_caco3_pct = 15)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_caco3_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

petroduric *Petroduric horizon (WRB 2022): cemented duric.*

Description

Petroduric horizon (WRB 2022): cemented duric.

Usage

```
petroduric(pedon, min_thickness = 10, min_duripan_pct = 10)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_duripan_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

petrogypsic *Petrogypsic horizon (WRB 2022): cemented gypsic.*

Description

Petrogypsic horizon (WRB 2022): cemented gypsic.

Usage

```
petrogypsic(pedon, min_thickness = 10, min_gypsum_pct = 5)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_gypsum_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

petroplinthic *Petroplinthic horizon (WRB 2022)*

Description

Continuous or fractured layer cemented by Fe (WRB 2022 Ch 3.1.27, p. 59). Since v0.9.312 a layer qualifies when its Fe cement is recorded (an iron pan, pan_type ferricrete or ironpan, or plinthite \>= min_plinthite_pct) and its cementation class is at least strongly cemented (criterion 4; until v0.9.311 moderately was enough), over min_thickness cm of adjoining layers (criterion 7). A layer whose Fe cement or cementation is not recorded is undecided.

Usage

```
petroplinthic(pedon, min_thickness = 10, min_plinthite_pct = 15)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_plinthite_pct	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

phaeozem *Phaeozem RSG diagnostic (WRB 2022)*

Description

Tests whether a profile satisfies the Phaeozem RSG criteria (WRB 2022 Ch 4, p. 113): a mollic horizon AND a base saturation (1 M NH4OAc, pH 7) \>= 50% throughout to 100 cm from the mineral soil surface or to a limiting layer, whichever is shallower. Carbonates are allowed: the Chernozems and Kastanozems are keyed out before. Where BS is not recorded, pH is read as in WRB 2022 Annex 2, Table 9.2. A profile not described down to that depth is undecided.

Usage

```
phaeozem(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 4, Phaeozems (p. 113).

---

pick_engine *Choose the best diagnostic engine for a single pedon*

Description

Per-pedon heuristic: returns "aqp" if the pedon's horizon table has the morphological richness that makes aqp's canonical NRCS dispatch reliable, otherwise returns "soilkey" (the more permissive hand-coded path).

Usage

```
pick_engine(pedon, min_score = 3L)
```

Arguments

pedon	A PedonRecord.
min_score	Integer (1-5). Minimum completeness score for "aqp" engine to fire (default 3).

Value

Character: "aqp" or "soilkey".

Heuristic

We score each pedon on a 0-5 morphology-completeness scale; aqp fires when score \>= min_score (default 3). The five axes:

1. **Designation present** (any layer has a non-blank designation, e.g. "A1", "Bt2", "Bw").
2. **Texture quantitative** (any layer has both clay_pct and sand_pct populated).
3. **Munsell complete** (any layer has all three of munsell_hue_moist, munsell_value_moist, munsell_chroma_moist populated).
4. **Structure recorded** (any layer has a non-blank structure_grade).
5. **Clay films / argic evidence** (any layer has a non-blank clay_films_amount or designation pattern matching Bt).

Why this matters

On BDsolos RJ (data-rich), the heuristic recommends aqp for ~99 canonical thresholds). On LUCAS topsoil-only (data-sparse), it recommends aqp for ~0 clay-films / designation axes are unfilled. Calling classify_*(pedon) routed through the heuristic gives the correct engine per pedon, recovering both the BDsolos RJ lift AND the LUCAS robustness.

See Also

argic, cambic.

---

pick_engine_batch *Per-pedon batch engine recommendation*

Description

Vectorised version of pick_engine returning the recommended engine for each pedon in a list.

Usage

```
pick_engine_batch(pedons, min_score = 3L)
```

Arguments

pedons	A list of PedonRecord objects.
min_score	Integer; forwarded to pick_engine.

Value

Character vector of length(pedons) with values "aqp" or "soilkey".

`pisoplinthic` *Pisoplinthic horizon (WRB 2022)*

Description

WRB 2022 Ch 3.1.28 (p. 60): \>= 40% of the volume (whole soil) yellowish, reddish or blackish concretions or nodules over 2 mm, at least moderately cemented, not part of a petroplinthic horizon, \>= 15 cm thick. Since v0.9.312 the concretions are read from `ferric_nodules_pct` (Fe/Mn nodules or concretions; the ISIS reader stores the hard ones over 2 mm), adjoining sublayers counting together; a layer where they are not recorded is undecided. Until v0.9.311 the test was plinthite \>= 15% with a "pisolitic" structure, which is not the WRB 2022 definition.

Usage

```
pisoplinthic(
  pedon,
  min_thickness = 15,
  min_plinthite_pct = 15,
  min_nodules_pct = 40
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Numeric threshold or option (see Details).
`min_plinthite_pct`	Not used since v0.9.312; kept so that existing calls still work.
`min_nodules_pct`	Minimum volume % of concretions and nodules (default 40).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

pi_to_confidence *Map a 95% prediction interval to a [0, 1] confidence score*

Description

Tightens confidence as the prediction interval narrows relative to the predicted value: `confidence = 1 - (PI95_width / |value|) / 4`, floored at 0 and capped at 1. When `value` is near zero we fall back to an absolute-width heuristic so we never blow up.

Usage

```
pi_to_confidence(pi95_low, pi95_high, value = NULL)
```

Arguments

`pi95_low`	Lower 2.5% quantile of the prediction.
`pi95_high`	Upper 97.5% quantile of the prediction.
`value`	Optional point prediction. When supplied, normalisation is by `|value|`; otherwise by `|midpoint|`.

Details

Properties of the mapping:

- Zero-width interval -> confidence = 1.
- Interval whose width equals `|value| * 4` -> confidence = 0.
- NA value or NA bounds -> confidence = 0.5 (neutral).

Value

Numeric in `[0, 1]`.

plaggic *Plaggic horizon (WRB 2022): sod-derived topsoil >= 20 cm with low BD AND independent evidence of human input.*

Description

v0.9.2.C tightening: the v0.3.3 implementation accepted ANY thick, low-BD, OC-rich A horizon, which over-fired across natural mollic / umbric / chernic surfaces. The diagnostic now requires, in addition to the OC + BD + thickness baseline, at least one independent anthropogenic-input marker:

- `p_mehlich3_mg_kg >= 50` (sustained sod / manure additions concentrate Mehlich-3 P in the topsoil), OR

- `artefacts_pct > 0` (any human artefact volume fraction is sufficient as a presence signal), OR
- designation pattern `Apl` / `Aplg` / `Apk` / explicit "plagg".

Without one of those markers the diagnostic returns FALSE even when OC + BD + thickness pass. This mirrors the v0.9.1 `qual_plaggic` gate but enforces the rule at the diagnostic level so any caller (SiBCS, USDA, future modules) inherits the protection.

Usage

```
plaggic(
  pedon,
  min_thickness = 20,
  max_bd = 1.5,
  min_oc = 0.6,
  min_p_mehlich3 = 100
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Numeric threshold or option (see Details).
`max_bd`	Numeric threshold or option (see Details).
`min_oc`	Numeric threshold or option (see Details).
`min_p_mehlich3`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`planic_features` *Planic features (WRB 2022)*

Description

Tests whether the profile shows an abrupt textural change between adjacent horizons (clay-doubling within 7.5 cm vertical distance, typically at the E/Bt boundary). Diagnostic of Planosols.

Usage

```
planic_features(pedon, min_ratio = 2, require_abrupt_boundary = TRUE)
```

Arguments

pedon	A PedonRecord.
min_ratio	Minimum clay ratio (default 2.0).
require_abrupt_boundary	If TRUE (default), the upper horizon must have boundary_distinctness matching "abrupt".

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 5, Planosols.

---

planosol *Planosol RSG gate (WRB 2022 Ch 4, p 107)*

Description

WRB 2022: an abrupt textural difference \<= 75 cm from the mineral soil surface and, within 5 cm directly above or below it, (1) stagnic properties in which the area of reductimorphic plus oximorphic features is \>= 50% and (2) reducing conditions in some parts of the volume with the reductimorphic features.

Usage

```
planosol(pedon, strict = NULL)
```

Arguments

pedon	A PedonRecord.
strict	Logical or NULL; recorded in the evidence. Since v0.9.312 it changes nothing (the fallback it disabled is gone).

Details

The two layers that meet at the abrupt textural difference stand for the 5 cm above and below it. Stagnic properties are those of stagnic_properties; a claric layer directly above the abrupt textural difference is stagnic (Ch 3.2.14, criterion 4) and its whole area counts as reductimorphic. Since v0.9.312 the abrupt textural difference has its transition criterion (boundary_distinctness), the depth limit of 75 cm is applied, and the v0.2 planic_features fallback (clay doubling at an abrupt boundary, with no redox evidence), which is not in WRB 2022, is gone.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`plinthic` *Plinthic horizon (WRB 2022)*

Description

Tests whether any horizon meets the plinthic horizon criteria (WRB 2022 Ch 3.1.30, p. 62), the diagnostic of Plinthosols with the pisoplinthic and petroplinthic horizons. The depth limit belongs to the Plinthosols key (`plinthosol()`), not to the horizon.

Usage

```
plinthic(pedon, min_thickness = 15, min_plinthite_pct = 15)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness in cm (default 15).
`min_plinthite_pct`	Minimum volume % plinthite (default 15).

Details

Criteria read (v0.9.312):

- 1: `plinthite_pct` \>= 15 (the oximorphic features inside former aggregates of criterion 1, as `plinthite_pct` records them); a layer where it is not recorded is undecided;
- 2: Fe-dithionite \>= 2.5% (2.a) when `fe_dcb_pct` is measured; below it the layer is undecided, since 2.b (10% Fe in the oximorphic features) and 2.c (hardening after drying and wetting) are not recorded; not measured, the recorded plinthite is taken for 2.c;
- 3: Feox/Fedith < 0.1 when both are measured;
- 4: not part of a petroplinthic or pisoplinthic horizon;
- 5: \>= 15 cm thick; adjoining sublayers (Btv1 over Btv2) count together.

The result is `NA` when undecided layers could make the horizon.

Value

A `DiagnosticResult`.

v0.9.72 designation morphological inference (opt-in)

Field-described Brazilian Plintossolos profiles (e.g.\ the Embrapa Redape curated dataset) routinely encode plinthite via the designation suffix `f` in the master letter sequence (e.g.\ `Btf`, `2Btf`, `Cf`) – the curator's direct assertion that plinthite is present – without recording `plinthite_pct` as a numeric volume percent.

With `options(soilKey.plinthic_designation_inference = TRUE)` the function accepts a layer as plinthic when:

1. the canonical `plinthite_pct` test is `NA` for that layer, AND
2. the designation matches `[A-Z]+[A-Za-z]*f[0-9]?` (a `f` master-letter modifier in any sub-position).

Default is `FALSE` (canonical behaviour preserved).

References

IUSS Working Group WRB (2022), Chapter 3.1.30, Plinthic horizon.

`posterior_classify` *Bayesian posterior classifier (optional)*

Description

Combines a deterministic `ClassificationResult` with a spatial prior. The deterministic key remains authoritative – this function reports only an alternative probabilistic view useful for downstream uncertainty quantification.

Usage

```
posterior_classify(result, prior, epsilon = 0.001)
```

Arguments

`result`	A `ClassificationResult` from `classify_wrb2022`.
`prior`	A spatial-prior data.table (as returned by `spatial_prior`).
`epsilon`	Small smoothing constant added to all prior entries before normalising, so RSGs unseen by the prior do not receive zero posterior.

Details

Posterior is computed under the simple model:

$$P(rsg|site, evidence) \propto L(rsg|evidence) \times P(rsg|site)$$

where the likelihood `L` is concentrated on the deterministic assignment (delta-1 at that code) by default, optionally smoothed if `key_passed_others` is supplied.

Value

A `data.table` with columns `rsg_code`, `prior`, `likelihood`, `posterior`.

`predict.soilKey_pls_model` *Predict from a soilKey_pls_model*

Description

S3 method that applies a trained PLSR model from `train_pls_from_ossl` to a (pre-processed) numeric matrix and returns predictions plus a 95 built from the cross-validated training RMSE.

Usage

```
## S3 method for class 'soilKey_pls_model'
predict(object, X, ...)
```

Arguments

`object`	A `soilKey_pls_model` object.
`X`	A pre-processed numeric matrix (rows = samples, columns = wavelengths). Must have the same column count used at training time.
`...`	Reserved.

Value

A data.frame with columns `value`, `pi95_low`, `pi95_high`, one row per sample.

`predict_from_spectra` *Predict soil properties from spectra*

Description

Ergonomic, named entry point for the OSSL-backed predictive pipeline. Accepts either a `PedonRecord` or a numeric spectra matrix, applies the same preprocessing used at training time (recorded on each model), and returns predictions in the canonical long-form schema.

Usage

```
predict_from_spectra(
  pedon_or_spectra,
  models = NULL,
  properties = NULL,
  overwrite = FALSE,
  verbose = TRUE,
  ...
)
```

Arguments

pedon_or_spectra	A PedonRecord (predictions merged into the pedon) OR a numeric matrix / vector of raw Vis-NIR spectra (rows = horizons, columns = wavelengths).
models	A named list of soilKey_pls_model objects (output of train_pls_from_ossl). Required.
properties	Character vector of property names to predict. Defaults to all properties in models.
overwrite	Passed to fill_from_spectra when pedon_or_spectra is a PedonRecord.
verbose	Verbosity passed downstream.
...	Ignored (reserved for future backends).

Details

When pedon_or_spectra is a PedonRecord, this function delegates to fill_from_spectra with method = "pretrained" and the predictions are written back to the pedon (with source = "predicted_spectra" provenance). When pedon_or_spectra is a numeric matrix or vector, this function returns the prediction data.table directly without touching any pedon.

Value

Either the mutated PedonRecord (invisibly) or a data.table with columns horizon_idx, property, value, pi95_low, pi95_high, n_neighbors.

Examples

```
## Not run:
lib <- download_ossl_subset(region = "south_america")
models <- train_pls_from_ossl(lib,
                              properties = c("clay_pct", "ph_h2o"))
predict_from_spectra(my_pedon, models = models)

## End(Not run)
```

---

predict_lab_from_spectra

Predict CIE Lab from Vis-NIR reflectance spectra

Description

Convenience wrapper: predict_xyz_from_spectra followed by the standard CIE Lab transform under D65 / 2-degree observer.

Usage

```
predict_lab_from_spectra(spectra, wavelengths)
```

Arguments

spectra	Reflectance values, in 0..1 or 0..100. A numeric vector (one sample), a numeric matrix (rows = samples, cols = wavelengths) or a data.frame.
wavelengths	Numeric vector of the wavelengths (in nm) corresponding to the columns of spectra. Must cover at least 400-700 nm; values outside 380-780 are ignored.

Value

A data.frame with columns L, a, b.

---

predict_munsell_from_spectra

Predict Munsell hue / value / chroma from Vis-NIR reflectance spectra

---

Description

Combines predict_xyz_from_spectra with the Munsell renotation interpolation in **munsellinterpol** (CRAN, GPL). Returns hue (e.g. "7.5YR"), value (0..10) and chroma (0..20) per sample, plus the soilKey fields

Usage

```
predict_munsell_from_spectra(spectra, wavelengths, round_chip = TRUE)
```

Arguments

spectra	Reflectance values, in 0..1 or 0..100. A numeric vector (one sample), a numeric matrix (rows = samples, cols = wavelengths) or a data.frame.
wavelengths	Numeric vector of the wavelengths (in nm) corresponding to the columns of spectra. Must cover at least 400-700 nm; values outside 380-780 are ignored.
round_chip	If TRUE (default), snaps the predicted HVC to the nearest standard Munsell chip grid via munsellinterpol::roundHVC(). FALSE returns continuous HVC (useful for further numeric work).

Details

The Munsell renotation is defined under *Illuminant C*, while the colorimetry here is computed under D65, so the conversion adapts D65 -> C. It calls munsellinterpol::XYZtoMunsell(XYZ, white =) (munsellinterpol >= 3.4-0), which performs that chromatic adaptation internally, and falls back to the numerically identical XYZ -> CIELAB(D65) -> LabToMunsell() route on older versions. Feeding D65 chromaticities straight to xyYtoMunsell() (with no white) would bias every colour toward green-yellow (a perfect neutral would return Chroma ~ 0.65 rather than 0); this routine avoids that. The D65 reference white is derived from the same bundled CIE table the colorimetry integrates against (so a constant-reflectance spectrum maps to an exact neutral, and a perfect reflecting diffuser to Munsell value 10), and the conversion is vectorised over all rows of spectra at once. At zero Chroma the Munsell hue is undefined, so a neutral is reported with hue "N"

in both the rounded and the continuous (round_chip = FALSE) notation. munsell_hue_moist, munsell_value_moist, munsell_chroma_moist ready to write into a PedonRecord via the pedon's add_measurement method (see also fill_munsell_from_spectra).

This is the v0.9.47 unblock for the v0.9.35 Argissolo Vermelho / Amarelo / Vermelho-Amarelo color-confusion case: when a user has Vis-NIR spectra (which Embrapa's BDsolos / FEBR do not include but the OSSL does), the Munsell hue can be recovered physically without waiting for the surveyor's morphological description.

Value

A data.frame with columns munsell_hue_moist, munsell_value_moist, munsell_chroma_moist, munsell_string (e.g. "7.5YR 4/6"), X, Y, Z, one row per sample.

Examples

```
## Not run:
# White reflector across the visible: should map to a near-neutral
# high-value Munsell color.
wl <- seq(380, 780, by = 5)
R  <- rep(0.9, length(wl))
predict_munsell_from_spectra(R, wavelengths = wl)

## End(Not run)
```

---

predict_ossl_mbl *Memory-based learning prediction against the OSSL library*

Description

Predicts a set of soil properties from pre-processed Vis-NIR or MIR spectra using *memory-based learning* (MBL) – the recommended OSSL workflow for heterogeneous libraries. Defaults follow the literature (Ramirez-Lopez et al., 2013): k = 100 neighbours, PLS-score dissimilarity, local PLS regression with 5 components, internal leave-one-out validation.

Usage

```
predict_ossl_mbl(
  X,
  properties,
  region = "global",
  k = 100L,
  ossl_library = NULL,
  ...
)
```

Arguments

X	A pre-processed numeric matrix (rows = horizons, columns = wavelengths).
properties	Character vector of OSSL-supported property names.
region	One of "global", "south_america", "north_america", "europe", "africa".
k	Integer number of neighbours.
ossl_library	Optional list with the OSSL training spectra (Xr) and reference values (Yr, a data.frame keyed by properties). When NULL, the synthetic path is used.
...	Additional arguments forwarded to resemble::mbl.

Details

If resemble::mbl is installed and an ossl_library artefact is supplied (a list with elements Xr, Yr) the function delegates to resemble::mbl(); otherwise it returns a deterministic synthetic prediction conditioned on the input spectra so that downstream code, tests and vignettes run without external dependencies. The fallback is annotated via the notes attribute on the returned data.table.

Value

A data.table with columns horizon_idx, property, value, pi95_low, pi95_high, n_neighbors. The "backend" attribute records which path was taken ("resemble" or "synthetic").

References

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J. A. M., & Scholten, T. (2013). The spectrum-based learner: A new local approach for modeling soil Vis-NIR spectra of complex datasets. *Geoderma*, 195–196, 268–279.

---

predict_ossl_plsr_local

Local PLSR prediction against the OSSL library

Description

Selects the k nearest neighbours to each test spectrum in the OSSL training set and fits a local PLS regression. Like predict_ossl_mbl, this function dispatches to resemble::mbl (with a local_algorithm = "pls" setting) when the dependency is available; otherwise it falls back to the synthetic predictor.

Usage

```
predict_ossl_plsr_local(
  X,
  properties,
  region = "global",
  k = 100L,
  ossl_library = NULL,
  ...
)
```

Arguments

X	A pre-processed numeric matrix (rows = horizons, columns = wavelengths).
properties	Character vector of OSSL-supported property names.
region	One of "global", "south_america", "north_america", "europe", "africa".
k	Integer number of neighbours.
ossl_library	Optional list with the OSSL training spectra (Xr) and reference values (Yr, a data.frame keyed by properties). When NULL, the synthetic path is used.
...	Additional arguments forwarded to resemble::mbl.

Value

A data.table with the same schema as predict_ossl_mbl.

---

predict_ossl_pretrained

Pre-trained OSSL prediction

Description

Applies the OSSL-distributed pre-trained PLSR / Cubist models for a set of soil properties to pre-processed spectra. Pre-trained models are loaded from ossl_models, a named list of property models that each must implement a predict(model, X) interface returning a data.frame with columns value, pi95_low, pi95_high. When ossl_models is NULL, the synthetic predictor is used.

Usage

```
predict_ossl_pretrained(
  X,
  properties,
  region = "global",
  ossl_models = NULL,
  ...
)
```

Arguments

X	A pre-processed numeric matrix (rows = horizons, columns = wavelengths).
properties	Character vector of OSSL-supported property names.
region	One of "global", "south_america", "north_america", "europe", "africa".
ossl_models	Optional named list of pre-trained models, keyed by property name.
...	Reserved.

Value

A data.table with columns horizon_idx, property, value, pi95_low, pi95_high, n_neighbors. n_neighbors is NA_integer_ for pre-trained models. The "backend" attribute records which path was taken.

predict_xyz_from_spectra *Predict CIE XYZ tristimulus values from Vis-NIR reflectance spectra*

Description

Numerically integrates user reflectance against the CIE 1931 2-degree Standard Observer color-matching functions, weighted by the D65 illuminant. Returns the tristimulus values X, Y, Z on the standard scale where $Y = 100$ for a perfect diffuse white.

Usage

```
predict_xyz_from_spectra(spectra, wavelengths)
```

Arguments

spectra	Reflectance values, in 0..1 or 0..100. A numeric vector (one sample), a numeric matrix (rows = samples, cols = wavelengths) or a data.frame.
wavelengths	Numeric vector of the wavelengths (in nm) corresponding to the columns of spectra. Must cover at least 400-700 nm; values outside 380-780 are ignored.

Value

A data.frame with columns X, Y, Z, one row per sample.

See Also

predict_lab_from_spectra, predict_munsell_from_spectra.

preprocess_spectra *Pre-process Vis-NIR or MIR spectra*

Description

Applies a chosen pre-processing pipeline to a numeric matrix of soil spectra. Rows are samples (typically horizons) and columns are wavelengths. Returns a numeric matrix; SG-based methods shorten the spectrum by w - 1 columns at the edges (default w = 5 so two columns are dropped from each side).

Usage

```
preprocess_spectra(X, method = c("snv+sg1", "snv", "sg1"), w = 5L, p = 2L)
```

Arguments

X	Numeric matrix or data.frame of spectra (rows = samples, columns = wavelengths). Wavelengths should be evenly spaced.
method	One of "snv", "sg1", "snv+sg1". Default "snv+sg1".
w	Window size for the SG filter. Must be odd; default 5.
p	Polynomial order for the SG filter. Default 2.

Details

Supported method values:

"snv" Standard Normal Variate. Each row is centered on its own mean and divided by its own standard deviation.

"sg1" Savitzky-Golay 1st derivative with a window of five wavelengths and a quadratic polynomial.

"snv+sg1" SNV followed by SG1 (default; the standard pipeline used by OSSL pretrained models for Vis-NIR).

If prospectr is available, we use prospectr::standardNormalVariate and prospectr::savitzkyGolay (Rcpp implementation, faster and supports arbitrary window/polynomial). The native fallback uses the classical 5-point first-derivative coefficients (-2, -1, 0, 1, 2) / 10, which is the closed-form Savitzky-Golay solution for window 5 / polynomial 2 / derivative 1.

Value

A numeric matrix. Column names (wavelengths) are preserved where possible; SG trimming drops (w - 1) / 2 columns from each edge.

References

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772–777.

Stevens, A., & Ramirez-Lopez, L. (2024). *prospectr*: Misc. functions for processing and sample selection of spectroscopic data. R package version 0.2.7.

Examples

```
set.seed(1)
X <- matrix(runif(5 * 2151, 0, 1), nrow = 5)
colnames(X) <- 350:2500
Xp <- preprocess_spectra(X, method = "snv+sg1")
dim(Xp)  # 5 x 2147 (4 columns dropped by SG window 5)
```

pretic *Pretic horizon (WRB 2022): "Amazonian Dark Earth" (terra preta de indio) horizon – thick anthropogenic surface with high P, SOC, and incorporated charcoal / pottery.*

Description

Pretic horizon (WRB 2022): "Amazonian Dark Earth" (terra preta de indio) horizon – thick anthropogenic surface with high P, SOC, and incorporated charcoal / pottery.

Usage

```
pretic(pedon, min_thickness = 20, min_oc = 1.5, min_p_mehlich3 = 30)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_oc	Numeric threshold or option (see Details).
min_p_mehlich3	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

print.soilKey_pls_model *Print method for soilKey_pls_model*

Description

Print method for soilKey_pls_model

Usage

```
## S3 method for class 'soilKey_pls_model'
print(x, ...)
```

Arguments

x	A soilKey_pls_model object.
...	Reserved.

Value

The object, invisibly.

prior_consistency_check

Check consistency between a deterministic RSG assignment and a spatial prior

Description

Returns a list describing whether the assigned RSG is plausible under the given prior. The deterministic classification is never overridden – this is purely a sanity-check signal.

Usage

```
prior_consistency_check(rsg_code, prior, threshold = 0.01)
```

Arguments

rsg_code	Two-letter RSG code (e.g. "FR"). Either the rsg_or_order from a ClassificationResult (in which case it must be the RSG name; we try to translate via the trace) or the bare code from a key trace entry.
prior	A spatial-prior data.table from spatial_prior.
threshold	Probability below which an assignment is flagged inconsistent (default 0.01).

Value

A list with elements:

- consistent: TRUE / FALSE / NA.
- p: probability of the assigned RSG in the prior (or NA_real_ if not found).
- threshold: the threshold used.
- status: a short status string – "consistent", "inconsistent", or "no_data".
- note: human-readable explanation.
- top_prior: data.table with the top three classes from the prior (for messages).

protocalcic_properties

Protocalcic properties (WRB 2022 Ch 3.2.8)

Description

Accumulations of secondary carbonates, visible when moist, that occupy \>= 5% of the exposed area with masses, nodules, concretions or filaments, or cover \>= 10% of the surfaces of aggregates, biopore walls or the undersides of coarse fragments (WRB 2022 3.2.8, p. 78). The test reads the recorded abundance of visible secondary carbonates, `secondary_carbonates_pct` (or the visual `carbonate_soft_pct` plus `carbonate_hard_pct`). Laboratory CaCO3 is not evidence of secondary carbonates: with no abundance recorded, a layer with no CaCO3 (`caco3_pct` 0, or no effervescence) has no protocalcic properties and any other layer is undecided. A `k` suffix with no abundance is undecided too, unless `options(soilKey.morphological_inference = TRUE)`, where it is accepted as morphological evidence.

Usage

```
protocalcic_properties(pedon, min_secondary_pct = 5)
```

Arguments

`pedon`	A `PedonRecord`.
`min_secondary_pct`	Minimum visible secondary carbonates, % of the exposed area (default 5).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence; the per-layer status (TRUE / FALSE / NA) is in `evidence$layer_status`.

`protogypsic_properties` *Protogypsic properties (WRB 2022 Ch 3.2.9): visible secondary gypsum \>= 1% but below the gypsic gate.*

Description

Protogypsic properties (WRB 2022 Ch 3.2.9): visible secondary gypsum \>= 1% but below the gypsic gate.

Usage

```
protogypsic_properties(pedon, min_caso4_pct = 1, max_caso4_pct = 5)
```

Arguments

`pedon`	A `PedonRecord`.
`min_caso4_pct`	Numeric threshold or option (see Details).
`max_caso4_pct`	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`protovertic` *Protovertic horizon (WRB 2022 Ch 3.1)*

Description

A weakly developed vertic horizon – the swelling/shrinking machinery is present but does not reach the full vertic intensity (cracks too narrow, or slickensides only "few", or thickness too small). Used by the Protovertic qualifier; relevant for soils that would be Vertisols if the cracks/slickensides were a notch stronger.

Usage

```
protovertic(pedon, min_clay = 30, min_thickness = 15)
```

Arguments

`pedon`	A PedonRecord.
`min_clay`	Numeric threshold or option (see Details).
`min_thickness`	Numeric threshold or option (see Details).

Details

v0.3.5 detection: clay \>= 30% AND any positive vertic evidence (slickensides at \>= "few" OR cracks_width_cm \>= 0.2 OR a wedge/lenticular structure_type) AND thickness \>= 15 cm. The positive cases that pass the strict vertic_horizon test are explicitly excluded so the two diagnostics partition the vertic-spectrum cleanly.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

read_febr_pedons *Load FEBR datasets as a list of PedonRecord objects*

Description

Wraps `febr::readFEBR()` (CRAN package, FEBR v1.9.9+ recommended) and adapts the returned `camada` (layer) + `observacao` tables to the soilKey schema. Auto-detects Munsell columns across the ~6 distinct conventions found in the 200 FEBR datasets that carry color data, parses PT-BR Munsell strings (`"2,5YR 3/6"`) and converts FEBR's standard units to soilKey conventions.

Usage

```
read_febr_pedons(
  dataset_codes = c("ctb0039"),
  febr_repo = NULL,
  min_munsell_coverage = 0,
  verbose = TRUE
)
```

Arguments

`dataset_codes`	Character vector of FEBR dataset IDs (e.g. `c("ctb0032", "ctb0562")`). Pass `"all"` to download every Munsell-bearing dataset; this is heavy (network calls per dataset). Default: a small curated sample for development.
`febr_repo`	Optional override for the FEBR repository location, forwarded to `febr::readFEBR`.
`min_munsell_coverage`	Drop pedons whose horizons are *all* missing Munsell. Default 0 (keep all); set to 0.5 to keep only pedons with at least 50 horizons having a Munsell hue.
`verbose`	If `TRUE` (default), prints per-dataset join statistics.

Details

Iron (v0.9.221), from the FEBR dictionary codes (g/kg): dithionite Fe (`ferro_ditionito_*`, or `fe2o3_ditionito_*` converted to the element) goes to `fe_dcb_pct`; oxalate Fe (`ferro_oxalato_*` / `fe2o3_oxalato_*`) to `fe_ox_pct`; total Fe of the sulfuric attack, as the oxide (`fe2o3_sulfurico_*`, or `ferro_sulfurico_*` converted), to `fe2o3_sulfuric_pct`; all in %. Until v0.9.220 the first `fe2o3_*` column of any method filled `fe_dcb_pct`, unconverted.

Per the May 2026 scan, ~80 `febr_index_munsell` to get the curated list of Munsell-bearing dataset IDs.

Value

A list of `PedonRecord` objects with `site$id` = FEBR `observacao_id`, `site$reference_sibcs` = the surveyor's classification when available, and one horizon per FEBR `camada` row.

See Also

febr_index_munsell, load_bdsolos_csv.

Examples

```
## Not run:
# Single dataset (35 perfis, 100% Munsell coverage)
pedons <- read_febr_pedons("ctb0039")

# Multiple datasets
pedons <- read_febr_pedons(c("ctb0032", "ctb0562", "ctb0568"))

# All Munsell-bearing datasets (slow; 200 datasets, ~36k horizons)
all_pedons <- read_febr_pedons("all")

## End(Not run)
```

---

read_isis *Read profiles of the ISRIC World Soil Reference Collection (ISIS)*

Description

Reads profiles live from ISRIC's ISIS GraphQL service and returns them as PedonRecords, ready for classify_all or benchmark_isis. Nothing is written to disk. Each pedon carries the names ISRIC gives it, the licence and the citation (isis_citation); see isis_profile_to_pedon for what the translation reads.

Usage

```
read_isis(
  profile_codes = NULL,
  country = NULL,
  max_profiles = 100L,
  absent_unless_recorded = TRUE,
  timeout = 60,
  progress = NULL
)
```

Arguments

profile_codes	Profile codes, e.g. c("KE009", "ZW005").
country	ISO 3166 two-letter country code, e.g. "KE", read when profile_codes is NULL.
max_profiles	The most profiles to read (default 100).
absent_unless_recorded	Passed to isis_profile_to_pedon.
timeout	Seconds to wait for each request (default 60).
progress	Optional function(done, total).

Value

A named list of PedonRecords, one per profile read. On a failed request the profiles read until then are returned with a warning.

Examples

```
## Not run:
p <- read_isis(c("KE009", "ZW005"))
classify_all(p$KE009)
p$KE009$site$reference_wrb_full

## End(Not run)
```

---

read_nsd_pedons *Read New Zealand National Soils Database survey exports*

Description

Reads the survey exports of the New Zealand National Soils Database (NSD, Manaaki Whenua - Landcare Research), as the National Soil Data Repository (NSDR) serves them (one zip of relational CSV tables per survey), into PedonRecord objects. The tables are read in memory from the zips (or from unzipped folders); nothing is written to disk.

Usage

```
read_nsd_pedons(
  path,
  absent_unless_recorded = TRUE,
  old_notation = c("translate", "asis"),
  classified_only = FALSE,
  accessed = NULL,
  verbose = FALSE
)
```

Arguments

path	One or more NSD survey zips, folders holding the unzipped CSV tables of a survey, or folders of either.
absent_unless_recorded	If TRUE (default), a described horizon (one with a matrix colour) with no mottle record, in a survey that records mottles, has none, unless its designation marks gleying.
old_notation	"translate" (default) or "asis": how designations in the older New Zealand notation are read (see Details).
classified_only	If TRUE, only profiles with an NZSC class.
accessed	Optional date the exports were downloaded, kept in the site and the citation.
verbose	Print the number of profiles read per survey.

Details

Each pedon carries the field description (designations, depths, boundary, moist Munsell colour, field texture, strength, failure, stickiness, plasticity, cementation, mottles, structure, clay coats and the colour of coatings, pans, roots) and the laboratory results of its samples (carbon, nitrogen, pH in water, KCl, CaCl2 and NaF, P retention, CEC and exchangeable cations, base saturation, KCl-extractable Al, oxalate, dithionite and pyrophosphate Fe and Al, particle size, bulk density, volcanic glass in the sand, stones, CaCO3, EC of a 1:5 extract, water at 1500 kPa, COLE), with the NSD table and column of every value in the provenance. Codes are decoded through the survey's `cl_concept` table. The site declares `designation_system = "nz_legacy"` (v0.9.395): the surveys are written in the New Zealand notation, many of them in its older form (A11, A12, B21, B22, B3 beside Bw), so the NZSC key reads them as New Zealand designations while the older-notation reading of the WRB and SiBCS diagnostics (A2 = E, B3 = BC) is judged from each profile's own designations. The site holds the surveyor's classes: `reference_nzsc` (the NZSC code, e.g. `"MOT"`, recoded to the 4th edition where the 4th edition renamed it), `reference_nzsc_name`, the order, group and subgroup codes (`reference_nzsc_order`, `_group`, `_subgroup`), the record as written (`reference_nzsc_raw`), `reference_usda` (the Soil Taxonomy order of the recorded name), `reference_usda_grtgroup`, `reference_usda_subgroup`, the NZ Genetic class and series; and the drainage class, the parent material (the soil note "PARENT ROCK"), the landform, whether the land is drained, the mean annual air temperature, latitude and longitude (from NZTM2000), the licence and the citation (`nsd_citation`). The header of `R/nsd.R` lists every reading.

Designations: the NSD gives many horizons a designation in the current notation (`designation_fao`), which is read (an FAO H master as the New Zealand O); the others keep the surveyor's designation, often in the older New Zealand notation (Taylor and Pohlen 1962). With `old_notation = "translate"`, (B) is read as Bw, the suffixes fe, ca and cs as s, k and y, an upper-case G suffix as g, and digits written between the master and its suffixes are moved to the end (B2g as Bg2). A2 is read as E, with its subdivisions and suffixes (A21 as E1, A2g as Eg). The old master G (a gley horizon) is read as Bg when a B horizon lies below it in the profile, otherwise as Cg (GC as BCg or Cg); the old master D (the substratum) as C, its prefixes and suffixes kept (BD as BC, DG as Cg) except its old suffixes r and n, which are not read (Dr and Dn as C; the NSD's own conversions give them as Cu, C, Bt or Cm). A cemented sesquioxidic B written Bms is read as Bfm (a placic horizon) when 10 mm or less thick and as Bsm (an ortstein-pan) when thicker, as the NZSC Diagnostics define them. The provenance keeps the recorded designation and what it was read as.

Structure: the older ped term nut (not a shape of Milne et al. 1995) is granular or polyhedral, one of the two (NZ Soil Description Handbook 2024, Note 12.1). It is kept in `structure_type_one_of` with `structure_type` empty: `classify_nzsc4` reads a criterion under each shape, and the other keys read the shape as not recorded. The older strength words moderately weak, moderately firm and moderately strong (of the Taylor and Pohlen 1962 scale) are not read.

Parent material: the soil note PARENT ROCK is also the `layer_origin` of each horizon whose origin is not recorded (the NSD exports record none per horizon). A note naming a sequence of materials ("loess over outwash gravels") is split by the profile's lithologic groups (2C, IIC) when its parts are as many; otherwise it stays the site's `parent_material` only.

The NSD records no field reactive-aluminium (NaF) test, Childs' test, sensitivity, fluidity, packing or Ksat, which the NZSC key reads; the laboratory pH in NaF is kept in `ph_naf`.

Value

A named list of `PedonRecord` objects (names are the NSD site identifiers).

References

Manaaki Whenua - Landcare Research (2020). National Soils Database (NSD). doi:10.26060/95m4cz25. Milne, J.D.G., Clayden, B., Singleton, P.L. and Wilson, A.D. (1995). Soil Description Handbook. Manaaki Whenua Press.

See Also

benchmark_nsd, classify_nzsc4.

Examples

```
## Not run:
peds <- read_nsd_pedons("path/to/nsd/surveys")
b <- benchmark_nsd(peds, level = "order")
b$accuracy

## End(Not run)
```

---

read_pedon_csv *Read a horizon spreadsheet (CSV/TSV) into a PedonRecord*

Description

The everyday entry point for anyone who has a soil profile in a spreadsheet: one row per horizon, one column per attribute, using soilKey's canonical column names (`top_cm`, `bottom_cm`, `designation`, `clay_pct`, `ph_h2o`, `munsell_hue_moist`, ...). The full list is `names(horizon_column_spec())`; the quickest start is the bundled template `system.file("extdata", "perfil_exemplo.csv", package = "soilKey")`.

Usage

```
read_pedon_csv(file, site = NULL, sep = "auto")
```

Arguments

`file`	Path to a `.csv` (comma) or `.tsv` (tab) file, one row per horizon.
`site`	Optional named list of site metadata (see PedonRecord). Defaults to `list(id = <file base name>)`.
`sep`	Field separator. `"auto"` (default) uses a tab for `.tsv` files and a comma otherwise.

Details

Only recognised columns are used (any extras are carried through untouched), so a messy export still works: soilKey uses the columns it understands, and recognises names that differ only in capitalisation or separators and common aliases (`Clay`, `SOC`, `CEC`, ...). Columns and site properties that it does not recognise are named in a warning, with the closest recognised names (see `unrecognised_attributes`). Site metadata (id, lat/lon, soil moisture / temperature regime) is optional and passed via `site`; without it the profile still classifies, just with less specificity where site data would have refined the name.

Value

A `PedonRecord`.

See Also

`classify_csv` to go straight from a file to the three classifications; `classify_all`; `PedonRecord`.

Examples

```
f <- system.file("extdata", "perfil_exemplo.csv", package = "soilKey")
pedon <- read_pedon_csv(f)
classify_all(pedon)$summary
```

---

read_sali_pedons *Read Queensland SALI sites as pedons*

Description

Turns the files saved by `fetch_sali` into `PedonRecord` objects for `classify_asc3` and the other keys: the site layer (`layer/layer101_off*.json`, or a data frame saved as `layer/layer101.rds`), the site reports (`pdf/<PROJECT>_<SITE>.pdf`) and the laboratory samples of the SALI soil chemistry API (`lab/<PROJECT>_<SITE>_samples.json`). A site is read when both its report and its laboratory file are on disk; sites whose report was over the download cap, whose report has no morphology table, or whose label names no ASC order (the layer's "Unclassified" sites) are left out.

Usage

```
read_sali_pedons(
  path,
  sites = NULL,
  cores = 1L,
  max_n = NULL,
  seed = NULL,
  accessed = NULL,
  verbose = TRUE
)
```

Arguments

path	The folder fetch_sali wrote.
sites	Optional sites to read (a data frame with project_code and site_id, or "<PROJECT>_<SITE>").
cores	Number of processes (parallel::mclapply; always 1 on Windows).
max_n	Optional maximum number of sites read (a random sample).
seed	Optional seed for that sample.
accessed	Optional date the data were downloaded, for the citation.
verbose	Print a summary.

Details

The morphology exists only in the report (a PDF), read from the coordinates of its words (**pdftools**). Each row of its Profile morphology table is translated into the coded horizon of the NatSoil API (Field Handbook codes) and read by the NatSoil reader's own conventions (see load_natsoil_pedons), the laboratory samples too, so both Australian panels are read alike. These readings (the SALI pilot of 2026-10-30, kept by Hugo) apply:

- Observation 1 of the site is read (a report may hold several).
- The surveyors' label of the layer is read by position (order, suborder, great group, subgroup) and each name matched, case-insensitive, to the 3rd edition (inst/rules/asc3); a name the 3rd edition does not have stays NA. "No Available Class" is ZZ and "Class Undetermined" YY. A suborder with no great groups (Hypergypsic, Lutic, Stratic Rudosols) has its subgroup in the great-group column, as NatSoil has it. The edition the report cites is kept in site$sali_asc_edition, without changing the codes. The report's confidence phrases are the Field Handbook's 1-4; other phrases (a conversion from another classification, a rapid check of key attributes) are "C".
- An empty cell is not recorded; "No mottles", "No segregations" and "No coarse fragments" are recorded absent. A colour without its moisture status is moist. Mottles of biological mixing, mechanical mixing and substrate influence are not mottles; jarosite mottles are.
- Textures: "Loam, fine sandy" is a loam; sandy and silty clays are clays of unspecified grade; heavy or light before a grade lighter than a clay is the + or - qualifier. The first structure listed is read; "Blocky" alone names no type. "Not recorded (historical)" is NA. A strength without its moisture is not read.
- Field pH, by depth: Raupach and Tucker first, then an electrode in situ, then a 1:5 suspension by probe (the pH after peroxide and the field EC are not used); it is read only where the laboratory has no pH, as in NatSoil. The effervescence field test gives the horizon's effervescence.
- Laboratory samples (not linked to horizons in SALI) are filed under the horizon holding their middle, then read by NatSoil's overlap rule; bulked samples count; results flagged invalid are dropped; "<x" is x / 2 with NatSoil's limit rule. A trailing letter of a method variant is dropped (6B2b is 6B2) and 2Z2_Clay is 2Z2_C.
- Tidal landform spellings are NatSoil's (inter-tidal is intertidal); no tidal flooding is inferred, nor the soil water status; the depth to free water and the substrate are kept in the site (sali_depth_to_free_water, sali_substrate) but not used.

And Hugo's decisions of 2026-10-30:

- MIR predictions (method codes ending in _MC2, _MC91, _MQ01, _MQ02 and the like) are used only where nothing was measured for that attribute in that horizon (a field pH the horizon reads counts as measured), and the provenance marks them as predictions (source `"predicted_spectra"`, a note naming the MIR codes; two calibrations predicting the same method are averaged). MIR methods the NatSoil reader has no code for are read by the property their name gives: 15C2_CEC as the CEC, 6B4 as organic carbon and 15J2 as the effective CEC.
- The lab table printed in the report is used only when the API holds no samples for the site's observation.

The site keeps the surveyors' classification (`reference_asc_order`, `reference_asc_suborder`, `reference_asc_great_group`, `reference_asc_subgroup`, `reference_asc_confidence`), as `benchmark_sali` and `benchmark_natsoil` read it, the label and the report's ASC name (`sali_asc_label`, `sali_pdf_asc`), names left unmatched (`sali_unmapped`), the drainage, landform, surface condition, microrelief and surface coarse fragments, coordinates (GDA94), the licence and the citation (`sali_citation`), where the laboratory values came from (`sali_lab_source`: `"api"`, `"report table"` or `"none"`) and how many values are MIR predictions (`sali_n_predicted`). The provenance has one row per horizon for its morphology and one per laboratory value, which names the source (the API or the report's table) and the method; the report's notes are kept there too.

Value

A named list of `PedonRecord` objects (names `SALI_<PROJECT>_<SITE>`), with an attribute `"sali_status"`: a data frame of the sites considered and what became of each.

References

Queensland Government. Soil and Land Information (SALI). National Committee on Soil and Terrain (2009). Australian Soil and Land Survey Field Handbook, 3rd edition. CSIRO Publishing.

See Also

`fetch_sali`, `benchmark_sali`, `sali_citation`, `load_natsoil_pedons`.

---

read_spectral_library *Read a Vis-NIR / MIR reflectance + lab table into an OSSL-shaped library*

Description

Turns an arbitrary spectral dataset (e.g. a Brazilian Vis-NIR/MIR library) into the canonical `list(Xr, Yr, metadata)` object consumed by `fill_from_spectra` and `classify_by_spectral_neighbours`. Column names are mapped to the package's canonical attributes (clay_pct, sand_pct, ..., and the taxonomic label columns `wrb_rsg` / `sibcs_ordem` / `usda_order`) via a built-in alias table (including Portuguese headers such as *argila* / *silte* / *carbono*) or an explicit `property_map` / `label_map`.

Usage

```
read_spectral_library(
  reflectance,
  metadata,
  id_col = "id",
  wavelengths = NULL,
  resample_to = NULL,
  property_map = NULL,
  label_map = NULL,
  normalize = c("auto", "none", "percent"),
  verbose = TRUE
)
```

Arguments

reflectance	Reflectance data: a matrix / data.frame with rows = samples and columns named by wavelength (nm); OR a long data.frame with id_col, wavelength_nm, reflectance; OR a path to a CSV in either form.
metadata	A data.frame with one row per sample carrying id_col plus lab attributes and optional taxonomic labels and lat/ lon. Rows are aligned to reflectance by id_col.
id_col	Sample identifier column shared by both tables (default "id").
wavelengths	Optional explicit wavelength vector (nm) when the reflectance columns are not wavelength-named.
resample_to	Optional target wavelength grid (nm) to linearly resample every spectrum onto (e.g. 350:2500); default keeps the native grid.
property_map, label_map	Optional named lists overriding the alias auto-detection, e.g. property_map = list(clay_pct = "ARGILA").
normalize	One of "auto" (divide by 100 when values look like percent), "percent", or "none".
verbose	Print a one-line summary (default TRUE).

Value

A list with Xr (numeric reflectance matrix), Yr (data frame of mapped properties + labels + lat/lon), and metadata (provenance). Ready to pass as ossl_library=.

See Also

pedons_from_spectral_table, benchmark_spectral_fill, fill_from_spectra

`read_wosis_layers_graphql` *Read one WoSIS profile's layers live from ISRIC*

Description

Fetches the layers of a single profile from ISRIC's WoSIS GraphQL endpoint and returns them in the soilKey horizon schema. Nothing is written to disk.

Usage

```
read_wosis_layers_graphql(
  profile_id,
  timeout = 30,
  max_layers = 40L,
  page_size = 10L,
  progress = NULL
)
```

Arguments

`profile_id`	A WoSIS `profile_id`, as returned by `read_wosis_profiles_graphql`.
`timeout`	Seconds to wait for each request (default 30).
`max_layers`	Largest profile, in layers, that is read (default 40).
`page_size`	Layers per request (default 10).
`progress`	Optional `function(done, total)` called after each page, for a progress bar.

Details

Units are converted from ISRIC's own catalogue (organic carbon, total nitrogen and carbonate equivalent arrive in g/kg and become %), and only properties whose method matches the soilKey column are mapped: volumetric coarse fragments, saturated-paste EC, gravimetric water retention and CEC at pH 7. WoSIS serves no exchangeable bases, base saturation or Fe/Al oxides, so those columns stay empty and the classifier reports them as missing.

Value

A data.frame of horizons in the soilKey schema, top to bottom, with attributes `"provenance"` (one row per value: horizon, column, WoSIS code, unit and method), `"licence"` (the most restrictive licence among the layers), `"dataset"`, `"n_layers"` and `"depth_shift_cm"` (non-zero when depths were shifted so an organic surface layer starts at 0 cm). Zero rows, with a `"wosis_error"` attribute, when ISRIC cannot be reached, the profile has no layers, or it has more than `max_layers`.

Load on ISRIC

The endpoint is a free public service whose database cancels any statement that runs longer than about 30 seconds, and the cost of a query grows with the number of layers times the number of properties. The layers are therefore counted first with a cheap query, then read in pages of `page_size` in a stable order. Profiles recorded as many fine depth increments (some have 90 one-centimetre layers) exceed `max_layers` and are not read unless you raise it deliberately.

See Also

`wosis_profile_to_pedon`, `wosis_citation`.

Examples

```
# Needs network access; skipped offline and in batch checks.
online <- interactive() && requireNamespace("httr", quietly = TRUE) &&
  !inherits(try(httr::HEAD("https://graphql.isric.org", httr::timeout(5)),
                silent = TRUE), "try-error")
if (online) {
  p <- read_wosis_profiles_graphql(country = "Argentina", n_max = 1)
  if (nrow(p)) {
    h <- read_wosis_layers_graphql(p$profile_id[1])
    h[, c("top_cm", "bottom_cm", "designation", "clay_pct", "oc_pct")]
  }
}
```

---

read_wosis_profiles_graphql

Query WoSIS profiles live from ISRIC

Description

Reads soil profiles straight from ISRIC's WoSIS GraphQL endpoint. Nothing is cached to disk: soilKey distributes no soil observations, and each call is made by you, to ISRIC, under ISRIC's terms.

Usage

```
read_wosis_profiles_graphql(
  country = NULL,
  wrb_rsg = NULL,
  n_max = 25L,
  licence_filter = c("permissive", "any"),
  timeout = 30,
  max_scan = 300L
)
```

Arguments

country	Optional country name, e.g. "Brazil".
wrb_rsg	Optional WRB Reference Soil Group, e.g. "Ferralsols".
n_max	Maximum profiles to return (default 25, at most 100).
licence_filter	"permissive" (default) or "any".
timeout	Seconds to wait for each request (default 30).
max_scan	Most profiles examined per call (default 300).

Value

A data.frame, one row per profile, with profile_id, profile_code, country, lat, lon, wrb_rsg, usda_order, dataset, licence, licence_short, n_layers and depth_cm. Zero rows if nothing matched. Never throws on a network failure: it returns zero rows with a "wosis_error" attribute, since an unreachable server is an expected condition. Attributes "wosis_excluded" and "wosis_excluded_licences" report what the licence filter withheld, and "wosis_no_layers" how many profiles had no layers at all.

Licence

WoSIS is licensed *per profile* as each data provider specified, and about half of it carries a Non-Commercial restriction. licence_filter = "permissive" (the default) returns only profiles free of that restriction (CC BY, or public domain); "any" returns everything. Either way the licence and the source dataset travel with every profile, because CC BY requires attribution when you reuse the data. Cite the data with wosis_citation.

Load on ISRIC

The endpoint accepts at most 100 profiles per request, so profiles are read in pages of 100, in a stable order, and at most max_scan are examined per call. The number of layers and the depth of each listed profile come from one light query on the layers table, which lets a caller tell a full profile from a single topsoil sample before loading it.

See Also

wosis_profile_to_pedon to turn a row into a PedonRecord with its horizons.

Examples

```
# Needs network access; skipped offline and in batch checks.
online <- interactive() && requireNamespace("httr", quietly = TRUE) &&
  !inherits(try(httr::HEAD("https://graphql.isric.org", httr::timeout(5)),
                silent = TRUE), "try-error")
if (online) {
  p <- read_wosis_profiles_graphql(country = "Argentina", n_max = 5)
  if (nrow(p)) p[, c("profile_code", "wrb_rsg", "n_layers", "licence_short")]
}
```

reducing_conditions — *Reducing conditions (WRB 2022 Ch 3.2.10) – per-pedon test wrapping* `test_reducing_conditions`.

Description

Reducing conditions (WRB 2022 Ch 3.2.10) – per-pedon test wrapping `test_reducing_conditions`.

Usage

```
reducing_conditions(pedon, min_redox_pct = 5)
```

Arguments

`pedon`	A `PedonRecord`.
`min_redox_pct`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

report — *Render a soilKey classification report*

Description

Produces a pedologist-facing report from one or more `ClassificationResult` objects, optionally including the source `PedonRecord`. The HTML output is fully self-contained (single file, inline CSS); the PDF output goes through `rmarkdown::render()` and therefore requires a working LaTeX install (or one of the alternative engines accepted by `rmarkdown`).

Usage

```
report(
  x,
  file,
  format = c("auto", "html", "pdf"),
  pedon = NULL,
  title = NULL,
  include_family = FALSE,
  specifiers = FALSE,
  lang = c("en", "pt"),
  systems = c("wrb2022", "sibcs", "usda"),
  ...
)
```

Arguments

x	A ClassificationResult, a list of ClassificationResults, or a PedonRecord (in which case all three keys are run automatically).
file	Output path. The format is inferred from the extension (.html or .pdf) unless format is given explicitly.
format	One of "auto", "html", "pdf".
pedon	Optional PedonRecord; when provided, its horizons table and provenance log are included.
title	Optional report title.
include_family	When x is a PedonRecord (so the three keys are run here), passes through to classify_usda to append the USDA family (5th category) to the subgroup. Default FALSE keeps the output byte-identical to earlier versions.
specifiers	When x is a PedonRecord, passes through to classify_wrb2022 to attach WRB depth specifiers (Epi-/Endo-/...) to depth-anchored qualifiers. Default FALSE. Both flags are ignored when x is already a (list of) ClassificationResult.
lang	Report language; "en" (default) or "pt" (Brazilian Portuguese).
systems	When x is a PedonRecord, the systems to run: any registered ids or aliases (see soilkey_systems). The default runs WRB 2022, SiBCS and USDA, as reports always have; add a national system (e.g. "cssc3") to include its result, with the class at each level, the classes left undecided before the assigned one and, for Hungary, the subtype and variety properties. Ignored when x is already a (list of) ClassificationResult.
...	Passed to method-specific renderers.

Details

This is an S3 generic with methods for ClassificationResult, list, and PedonRecord. Most users call report() directly with a list of three results (list(classify_wrb2022(p), classify_sibcs(p), classify_usda(p))) to get a cross-system one-pager.

Value

The output path, invisibly.

Examples

```
pedon <- make_ferralsol_canonical()
out <- file.path(tempdir(), "soilkey_report.html")
report(pedon, file = out, pedon = pedon)
file.exists(out)
```

`report_html` *Render a soilKey classification report as self-contained HTML*

Description

See `report` for the generic. This function writes a single-file HTML report with inline CSS (no external network requests, no 'htmltools' dependency) so it can be emailed or archived as-is.

Usage

```
report_html(
  x,
  file,
  pedon = NULL,
  title = NULL,
  include_family = FALSE,
  specifiers = FALSE,
  lang = c("en", "pt"),
  systems = c("wrb2022", "sibcs", "usda"),
  ...
)
```

Arguments

`x`	A `ClassificationResult`, list of results, or `PedonRecord`.
`file`	Output `.html` path.
`pedon`	Optional `PedonRecord`.
`title`	Report title.
`include_family, specifiers`	Passed through to the keys when `x` is a `PedonRecord`; see `report`.
`lang`	Report language; `"en"` (default) or `"pt"` (Brazilian Portuguese).
`systems`	The systems to run when `x` is a `PedonRecord`; see `report`.
`...`	Currently unused.

Value

The output path, invisibly.

`report_pdf` *Render a soilKey classification report as PDF*

Description

See `report` for the generic dispatcher. This function assembles a temporary '.Rmd' file with the same content as `report_html` (site, cross-system summary, classification cards, horizons, provenance) and renders it via `rmarkdown::render()`.

Usage

```
report_pdf(
  x,
  file,
  pedon = NULL,
  title = NULL,
  include_family = FALSE,
  specifiers = FALSE,
  lang = c("en", "pt"),
  systems = c("wrb2022", "sibcs", "usda"),
  ...
)
```

Arguments

`x`	A `ClassificationResult`, list of results, or `PedonRecord`.
`file`	Output `.pdf` path.
`pedon`	Optional `PedonRecord`.
`title`	Report title.
`include_family, specifiers`	Passed through to the keys when `x` is a `PedonRecord`; see `report`.
`lang`	Report language, `"en"` (default) or `"pt"` (Brazilian Portuguese).
`systems`	The systems to run when `x` is a `PedonRecord`; see `report`.
`...`	Passed to `rmarkdown::render()`.

Value

The output path, invisibly.

`report_to_qgis` *Export a classification result + pedon to a QGIS GeoPackage*

Description

Writes a single GeoPackage (`.gpkg`) that QGIS reads natively, containing one POINT layer (the profile location with all classification metadata as attributes) plus two attribute-only tables (the horizons schema and the provenance log). Lets a pedologist overlay the soilKey result on a soil-survey base map or join it with field-campaign vector data without writing R or SQL.

Usage

```
report_to_qgis(
  pedon,
  classifications,
  file,
  report_html = NULL,
  overwrite = TRUE
)
```

Arguments

`pedon`	A `PedonRecord`.
`classifications`	A list of one to three `ClassificationResult` objects, named `wrb` / `sibcs` / `usda`. Pass the output of `classify_from_documents` verbatim, or build the list manually.
`file`	Output path (`.gpkg`). Created with parents.
`report_html`	Optional path to a sibling HTML report (rendered via `report_html`) – stored in the `report_html` attribute of `pedon_point` so QGIS users can launch the report from the feature pop-up.
`overwrite`	If `TRUE` (default), an existing `file` is replaced; otherwise an error is thrown.

Value

The output `file` path, invisibly. Side-effect: writes a multi-layer GeoPackage.

Geometry handling

The point geometry uses the pedon's site CRS (`pedon$site$crs`, default EPSG:4326). When the site has no coordinates, the function still writes the two attribute tables but skips the point layer and emits a warning.

Layer schema

`pedon_point` site_id, country, year, lat, lon, crs, wrb_name, wrb_rsg, wrb_grade, wrb_principal, wrb_supplementary, sibcs_name, sibcs_ordem, sibcs_grade, usda_name, usda_order, usda_grade, n_horizons, report_html (relative path), generated_at.

`horizons_table` site_id, horizon_idx, top_cm, bottom_cm, designation, plus the canonical `horizon_column_spec()` attributes when present.

`provenance_log` site_id, horizon_idx, attribute, source, confidence, notes.

See Also

`report` for HTML / PDF reports; `classify_from_documents` for the high-level one-liner that produces compatible `classifications`.

Examples

```
## Not run:
pedon <- make_ferralsol_canonical()
results <- list(
  wrb   = classify_wrb2022(pedon, on_missing = "silent"),
  sibcs = classify_sibcs(pedon, include_familia = TRUE),
  usda  = classify_usda(pedon)
)
report_to_qgis(pedon, results,
               file        = "perfil_042.gpkg",
               report_html = "perfil_042.html")
# In QGIS: Layer -> Add Layer -> Add Vector Layer -> perfil_042.gpkg

## End(Not run)
```

`resolve_wrb_qualifiers` *Resolve WRB 2022 qualifiers for a Reference Soil Group*

Description

Walks the RSG's lists of principal and supplementary qualifiers (WRB 2022 Chapter 4, in `inst/rules/wrb2022/qualifiers` and tests each qualifier against the pedon, following the rules for naming soils of Chapter 2.2: in a slash group (`"Rhodic/Xanthic"`) only the first qualifier that applies is used ("If no specifier is used, only the first applicable qualifier can be used", Ch 2.1, p. 24-25); with `specifiers = TRUE` the texture qualifiers of the group (`"Arenic/Clayic/Loamic/Siltic"`) that apply at different depths are all used, each with its specifier of Ch 2.3.1 rule 2 (p. 27), from the top to the bottom of the profile ("Anoloamic, Endoclayic", p. 26; since v0.9.315); a qualifier made redundant by another is left out (Eutric when Calcaric or Dolomitic applies); Haplic applies only where the RSG lists it and no other principal qualifier does.

Usage

```
resolve_wrb_qualifiers(pedon, rsg_code, rules = NULL, specifiers = FALSE)
```

Arguments

pedon	A PedonRecord.
rsg_code	Two-letter RSG code (e.g. "FR" for Ferralsols).
rules	Optional pre-loaded rules list (saves I/O when many RSGs are tested).
specifiers	If TRUE, auto-attach WRB Ch 5 depth specifiers (Epi-/Endo-/Bathy-/Amphi-/Panto-/Kato-) to depth-anchored qualifiers based on the feature's actual depth. Default FALSE leaves names byte-identical to earlier versions.

Details

Both vectors come back in the order of the name. Principal qualifiers are written right to left, "the uppermost qualifier in the list is placed closest to the name of the RSG", so principal reads the matched list backwards. Supplementary qualifiers open with the texture qualifiers (top to bottom of the profile when specifiers make several apply), then follow the alphabetical order of the qualifier, not of the subqualifier. Until v0.9.217 both kept the order of the old lists.

Value

A list with principal and supplementary (character vectors, in the order they are written in the name), trace and trace_supplementary.

---

retic_properties *Retic properties (WRB 2022)*

Description

WRB 2022 Chapter 3.2.11: claric material interfingering into an argic or natric horizon from its upper limit. soilKey reads criteria 1 and 6 from the data, the finer-textured parts belonging to an argic or natric horizon (argic, natric_horizon) and the interfingering recorded at its upper limit, and stands in for criteria 2-5, 7 and 8 (claric coarser parts, colour and clay contrasts, width, 10-90% of the sections, not in a plough layer), which no column holds, by what the describer recorded on the uppermost argic or natric layer or the layer directly above it: a designation matching pattern; an FAO designation of a layer with distinct parts of an E and a B horizon (E/B, 2E/Bt, B/E, since v0.9.312); or (since v0.9.312) an E or claric layer whose lower boundary on the argic or natric horizon is irregular or broken (boundary_topography), the uneven limit that 3.2.1 names as an example of retic properties.

Usage

```
retic_properties(pedon, pattern = "glossic|retic|albeluvic")
```

Arguments

pedon	A PedonRecord.
pattern	Regex (default "glossic|retic|albeluvic").

Details

FALSE without an argic or natric horizon, or when designations or boundaries are recorded and none shows the interfingering; NA when the argic or natric horizon cannot be established or there are neither designations nor boundary topographies. Until v0.9.219 the designation alone decided, anywhere in the profile.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 3.2.11, Retic properties.

---

rp2008_crosswalk *Correspondence tables of the Referentiel pedologique 2008*

Description

The correspondence tables of the Referentiel pedologique 2008 (AFES, 2009) as data frames: codes and names as the book prints them, with the printed page, its doubts ("?") kept as a flag and its slips noted. They are notes for display and benchmarks: classify_rp2008 and classify_cpcs1967 never read them.

Usage

```
rp2008_crosswalk(
  table = c("wrb2006", "wrb2006_horizons", "wrb2006_qualifiers", "wrb2006_faux_amis",
    "cpcs_horizons", "cpcs_classes", "cpcs_new_in_rp", "legacy_rp1995")
)
```

Arguments

table	Which table (see Details).

Details

"wrb2006" Annex 5 (pp. 379-390): one row per reference of the 110 (two for the brunisols eutriques, which the annex splits by their saturation, and one more for the lithosols stricts), with its WRB 2006 name. The WRB 2006 is the edition the book cites, not soilKey's WRB 2022.

"wrb2006_horizons" Annex 5: the reference horizons the tables put beside a WRB 2006 diagnostic horizon (pp. 380, 383, 384).

"wrb2006_qualifiers" Annex 5: the RP qualifiers and their WRB 2006 qualifiers, "most often imperfect" (pp. 379, 380, 385, 391-392); scope is the group the book restricts a row to.

"wrb2006_faux_amis" Annex 5: the WRB 2006 qualifiers spelled like an RP qualifier but meaning something else (p. 392).

`"cpcs_horizons"` Annex 6 (p. 393): one row per CPCS 1967 horizon symbol and RP code; `intertropical` for the codes of the intertropical domain, `slip` for the codes that are not RP codes, `after_1967` for the symbols not in the CPCS 1967.

`"cpcs_classes"` Annex 7 (pp. 394-398): one row per CPCS category (keyed by its `cpcs1967` code) and RP entry: the entry's name, the reference codes or the group (GER) it names, its qualifiers, `double` for a double attachment, `none` where the book prints "???".

`"cpcs_new_in_rp"` Annex 7: the RP categories the book calls completely new with respect to the CPCS (p. 398).

`"legacy_rp1995"` The references of the RP of 1995 removed in 2008 (p. VII) and the names of 2008 the book gives for such solums; `reading` where the destination is soilKey's reading. Used by `rp2008_from_label`.

Value

A data frame, one row per pair of the table.

See Also

`rp2008_from_label`, `classify_rp2008`.

Examples

```
head(rp2008_crosswalk("wrb2006"))
rp2008_crosswalk("cpcs_horizons")[1:6, ]
subset(rp2008_crosswalk("cpcs_classes"), cpcs == "V/111")
```

---

rp2008_from_label *Read French soil labels as references of the Referentiel pedologique 2008*

Description

Reads labels of French datasets (a reference, a type with its qualifiers, a double attachment written with a hyphen) as the references of the RP 2008, for benchmarks. A label of the RP of 1995 whose reference was removed in 2008 (p. VII: the four old brunisols, the pseudo-luvisols, the reductisols dupliques ...) is read through `rp2008_crosswalk("legacy_rp1995")`; a reference of 1995 that keeps its name in 2008 is read as the 2008 one. Case, accents, the singular or plural and the qualifiers after the name do not matter: the longest name the label starts with decides.

Usage

```
rp2008_from_label(labels)
```

Arguments

`labels`	Character vector of labels.

Value

A data frame, one row per label: `label`; `references` (the codes, comma-separated: several when the label leaves the choice open, any of them); `ger` (the group of references); `source` (`"rp2008"` a reference of 2008, `"rp1995"` a label of 1995 removed in 2008, `"group"` a group named alone, `NA` not read); `double_with` (the codes of the second part of a double attachment, as in `luvisol typique-redoxisol`); `note`.

See Also

`rp2008_crosswalk`.

Examples

```
rp2008_from_label(c("Brunisol eutrique resature, limoneux",
                    "BRUNISOL SATURE", "luvisol typique-redoxisol",
                    "Sol brun lessive"))
```

---

run_all_benchmarks *Run the full soilKey benchmark suite and (optionally) write a report*

Description

Auto-detects which reference datasets are available locally, runs each via `benchmark_unified`, adds the offline canonical sanity row and the AfSP sample when present, and returns a tidy accuracy summary. When `report_path` is given, a consolidated Markdown report is written.

Usage

```
run_all_benchmarks(
  datasets = "auto",
  paths = NULL,
  max_n = 300L,
  level = "order",
  report_path = NULL,
  verbose = TRUE
)
```

Arguments

`datasets`	`"auto"` (default) detects available datasets; otherwise any subset of `c("bdsolos", "febr", "kssl", "lucas_esdb", "redape")`, the literal `"canonical"` (only the fixture sanity row), or `"all"` (every dataset regardless of availability – absent ones are skipped).
`paths`	Named list of dataset paths (see `benchmark_unified`). `NULL` uses the package defaults (override the root via `options(soilKey.benchmark_root = "...")`).
`max_n`	Cap on pedons per dataset (keeps the run fast). Default 300.

level	Comparison level forwarded where supported (currently the suite reports at "order" / top level).
report_path	File to write the Markdown report to, TRUE to auto-name one under inst/benchmarks/reports/, or NULL (default) for no file.
verbose	Print progress.

Value

Invisibly, a list with summary (data.frame: dataset, system, n_compared, accuracy), per_system (pooled), raw (full benchmark_unified output), weak (zero-recall classes) and config.

See Also

benchmark_unified, benchmark_redape.

Examples

```
## Not run:
res <- run_all_benchmarks(max_n = 250,
                          report_path = TRUE)
res$summary

## End(Not run)
```

---

run_classify_app *Launch the soilKey interactive classification Shiny app*

Description

Opens a local Shiny app ("Pro") that drives the soilKey pipeline from a browser – no R code required: build a pedon from a canonical fixture, a CSV upload, or an interactive horizon editor; classify under WRB 2022 / SiBCS 5 / USDA ST 13 with the full key trace; run VLM photo extraction, OSSL spectral gap-fill, the SoilGrids spatial prior, an interactive **leaflet** map that queries the class prior at a clicked point, and a Monte-Carlo robustness analysis; and download a cross-system HTML or PDF report. The interface is bilingual (English / Portuguese; see lang).

Usage

```
run_classify_app(
  ui = c("pro", "classic"),
  lang = c("en", "pt"),
  port = NULL,
  launch.browser = TRUE,
  ...
)
```

Arguments

ui	Kept for back-compatibility. "pro" (default) launches the professional multi-tab app. "classic" – the original single-page uploader – was **retired in v0.9.117**; passing it now emits a deprecation warning and launches the Pro app instead.
lang	Initial interface language: "en" (default) or "pt" (Brazilian Portuguese). Can also be switched live from the app's navbar.
port	Port for the local server. Default lets Shiny choose.
launch.browser	Whether to open the app in the default browser (default TRUE).
...	Additional arguments passed to runApp.

Details

Needs the optional packages **bslib**, **shinyWidgets**, **plotly** and **leaflet** (all in Suggests); the function raises a clear, copy-pasteable error if any are missing.

Value

Invisibly the value returned by shiny::runApp().

Examples

```
## Not run:
run_classify_app()              # professional multi-tab app (English)
run_classify_app(lang = "pt")   # interface em portugues

## End(Not run)
```

---

run_demo *Launch the soilKey Shiny demo (one-screen GUI)*

Description

Opens a Shiny app that lets a non-coder pick one of the 31 canonical profiles or upload a small horizons CSV, click **Classify**, and read the WRB / SiBCS / USDA names plus the deterministic key trace and the evidence grade. Useful for live demos, classroom teaching, and for pedologists who want to verify the package on a profile they already know without writing R code.

Usage

```
run_demo(...)
```

Arguments

...	Forwarded to shiny::runApp() (e.g. port = 4321, launch.browser = FALSE, host = "0.0.0.0").

Details

Requires the shiny package. The taxonomic key is still deterministic: no VLM is invoked from the GUI.

Value

Invisibly, the value returned by shiny::runApp().

Examples

```
## Not run:
  soilKey::run_demo()

## End(Not run)
```

---

run_sibcs_grande_grupo

Resolve o grande grupo (3o nivel) de um pedon classificado em uma subordem SiBCS

Description

v0.7.3: itera os Grandes Grupos da subordem em ordem canonica via o engine generico run_taxa_list;
a primeira test-block que passa captura o perfil. Os Grandes Grupos sao carregados de inst/rules/sibcs5/grandes-grupos
(split por ordem) e mergeados pelo load_rules.

Usage

```
run_sibcs_grande_grupo(pedon, subordem_code, rules = NULL)
```

Arguments

pedon	A PedonRecord.
subordem_code	Codigo da subordem (e.g. "OJ" para Organossolos Tiomorficos).
rules	Lista de regras carregada via load_rules.

Details

Quando a subordem nao tem bloco de Grandes Grupos definido (ainda nao wirado para todas as ordens), retorna list(assigned = NULL, trace = list()) – comportamento nao-fatal que permite classify_sibcs parar no 2o nivel sem erro.

Value

Lista com assigned (entrada YAML do Grande Grupo ou NULL) e trace.

run_sibcs_key *Roda a chave SiBCS 5a edicao sobre um pedon*

Description

Roda a chave SiBCS 5a edicao sobre um pedon

Usage

```
run_sibcs_key(pedon, rules = NULL)
```

Arguments

pedon	A PedonRecord.
rules	Conjunto de regras pre-carregado; se NULL, le inst/rules/sibcs5/key.yaml.

Value

Lista com assigned (entrada YAML da ordem atribuida) e trace.

run_sibcs_subgrupo *Resolve o subgrupo (4o nivel) de um pedon classificado em um Grande Grupo SiBCS*

Description

v0.7.3.B: itera os Subgrupos do Grande Grupo em ordem canonica via o engine generico run_taxa_list;
a primeira test-block que passa captura o perfil. Os Subgrupos sao carregados de inst/rules/sibcs5/subgrupos/<ordem>.
(split por ordem) e mergeados pelo load_rules.

Usage

```
run_sibcs_subgrupo(pedon, gg_code, rules = NULL)
```

Arguments

pedon	A PedonRecord.
gg_code	Codigo do Grande Grupo (e.g. "OJF" para Organossolos Tiomorficos Fibricos).
rules	Lista de regras carregada via load_rules.

Details

Em contraste com o 3o nivel (Grandes Grupos de Organossolos), Subgrupos de Cap 14 SEMPRE tem catch-all tests:{default:true} como ultima entrada de cada lista (subgrupo "tipico"), entao a classificacao sempre desce ao 4o nivel quando o GG foi resolvido.

Value

Lista com `assigned` (entrada YAML do Subgrupo ou `NULL`) e `trace`.

`run_sibcs_subordem` *Resolve a subordem de um pedon ja classificado em uma ordem SiBCS*

Description

Itera as subordens da ordem em ordem canonica via o engine generico `run_taxa_list`; a primeira cuja test-block passa captura o perfil. Se nenhuma passar, retorna a ultima subordem (catch-all `tests:{default:true}`).

Usage

```
run_sibcs_subordem(pedon, ordem_code, rules = NULL)
```

Arguments

`pedon`	A `PedonRecord`.
`ordem_code`	Codigo de uma letra da ordem (e.g. "L" para Latossolos).
`rules`	Lista de regras carregada via `load_rules`.

Value

Lista com `assigned` (entrada YAML da subordem ou `NULL` se a ordem nao tiver bloco) e `trace`.

`run_taxa_list` *Iterate a flat taxa list and evaluate tests in canonical order*

Description

Internal iterator extracted from `run_taxonomic_key` so nested categorical levels (subordens, grandes grupos, subgrupos, familias) can be iterated directly, without going through the `rules[[level_key]]` indirection that only makes sense at the top level.

Usage

```
run_taxa_list(pedon, taxa)
```

Arguments

`pedon`	A `PedonRecord`.
`taxa`	A list of taxon entries; each entry must have `code`, `name`, and `tests` fields, where `tests` is a block parseable by `evaluate_rsg_tests`.

Details

Behavioural note: when `taxa` is empty or `NULL`, returns `list(assigned = NULL, trace = list())` – a sub-level lookup with no canonical entries is non-fatal. The top-level `run_taxonomic_key` keeps the stricter "missing list is an error" semantics by guarding before calling this helper.

Value

A list with `assigned` (the entry of the assigned taxon, or NULL when `taxa` was empty) and `trace`.

`run_taxonomic_key` *Run a taxonomic key (system-agnostic engine)*

Description

Iterates over the taxa list at `rules[[level_key]]` in canonical order; the first taxon whose tests pass is assigned. `evaluate_rsg_tests` is reused as the per-taxon evaluator regardless of system – the test combinator semantics (`all_of` / `any_of` / `default` / `not_implemented_v01`) are the same in all three systems.

Usage

```
run_taxonomic_key(pedon, rules, level_key)
```

Arguments

`pedon`	A `PedonRecord`.
`rules`	A parsed rule set (output of `load_rules`).
`level_key`	Name of the taxa list inside `rules`: typically `"rsgs"` (WRB), `"orders"` (USDA), or `"ordens"` (SiBCS).

Details

Used at the TOP level (RSG / Order / Ordem). For nested categorical levels (subordens, grandes grupos, subgrupos, familias) iterate the flat taxa list directly via `run_taxa_list`.

Value

A list with `assigned` (the YAML entry of the assigned taxon) and `trace` (one entry per taxon tested).

run_usda_great_group *Run the USDA Great Group key for a given Suborder*

Description

Run the USDA Great Group key for a given Suborder

Usage

```
run_usda_great_group(pedon, suborder_code, rules = NULL)
```

Arguments

pedon	A PedonRecord.
suborder_code	The Suborder code (e.g. "AA" for Histels).
rules	Optional pre-loaded rule set.

Value

A list with assigned and trace; assigned is NULL if the Suborder has no great-groups YAML.

run_usda_key *Run the USDA Soil Taxonomy Order key over a pedon*

Description

Run the USDA Soil Taxonomy Order key over a pedon

Usage

```
run_usda_key(pedon, rules = NULL)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set; if NULL, reads inst/rules/usda/key.yaml.

Value

A list with assigned (the YAML entry of the assigned Order) and trace.

run_usda_subgroup *Run the USDA Subgroup key for a given Great Group*

Description

Run the USDA Subgroup key for a given Great Group

Usage

```
run_usda_subgroup(pedon, great_group_code, rules = NULL)
```

Arguments

pedon	A PedonRecord.
great_group_code	The Great Group code (e.g. "AAA" for Folistels).
rules	Optional pre-loaded rule set.

Value

A list with assigned and trace; assigned is NULL if the Great Group has no subgroups YAML.

run_usda_suborder *Run the USDA Suborder key for a given Order*

Description

Run the USDA Suborder key for a given Order

Usage

```
run_usda_suborder(pedon, order_code, rules = NULL)
```

Arguments

pedon	A PedonRecord.
order_code	The Order code (e.g. "GE" for Gelisols).
rules	Optional pre-loaded rule set.

Value

A list with assigned and trace; assigned is NULL if the Order has no suborders YAML.

run_wrb_key *Run the WRB 2022 key over a pedon*

Description

Iterates over the RSGs in canonical key order; the first RSG whose tests pass is assigned. RSGs whose tests return NA (stubbed diagnostics or insufficient data) are skipped and recorded in the trace.

Usage

```
run_wrb_key(pedon, rules = NULL)
```

Arguments

pedon	A PedonRecord.
rules	Optional pre-loaded rule set; if NULL, reads inst/rules/wrb2022/key.yaml.

Value

A list with assigned (the YAML entry for the assigned RSG) and trace (one entry per RSG tested, in order).

salic *Salic horizon (WRB 2022)*

Description

Tests whether any horizon meets the salic horizon criteria. The salic horizon is a horizon of soluble-salt accumulation, diagnostic for Solonchaks.

Usage

```
salic(
  pedon,
  min_thickness = 15,
  min_ec_dS_m = 15,
  alkaline_min_ec_dS_m = 8,
  alkaline_min_pH = 8.5,
  min_product = 450,
  alkaline_min_product = 240
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Minimum thickness in cm (default 15).
min_ec_dS_m	Primary EC threshold (default 15 dS/m at 25C).
alkaline_min_ec_dS_m	Alkaline-path EC threshold (default 8 dS/m, used when pH(H2O) \>= alkaline_min_pH).
alkaline_min_pH	Required pH(H2O) for alkaline path (default 8.5).
min_product	Primary path product (EC * thickness in dS/m * cm) threshold (default 450 per WRB 2022).
alkaline_min_product	Alkaline-path product threshold (default 240).

Details

Sub-tests called:

- test_ec_concentration – EC \>= 15 dS/m (primary) OR (EC \>= 8 dS/m AND pH(H2O) \>= 8.5) (alkaline).
- test_minimum_thickness – thickness \>= 15 cm.
- test_salic_product – EC * thickness product \>= 450 (primary) or \>= 240 (alkaline) per qualifying layer.

v0.3.1: alkaline-path and product test added (WRB 2022 Ch 3.1.20, p. 49). Earlier versions only enforced the primary EC + thickness gate.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition. International Union of Soil Sciences, Vienna. Chapter 3.1.20 – Salic horizon (p. 49).

---

sali_citation *Citation for data read from Queensland's SALI*

Description

Returns the attribution for the soil site data of Queensland's Soil and Land Information system (SALI) read by fetch_sali and read_sali_pedons: the site layer of Queensland Spatial, the site listing reports (SALI4320) and the SALI soil chemistry API. The data are licensed under CC BY 4.0; the State makes no representations or warranties about them.

Usage

```
sali_citation(accessed = NULL)
```

Arguments

accessed	Optional date the data were downloaded (its year is the year of the attribution; the current year when NULL).

Value

A character string.

Examples

```
sali_citation()
```

---

saprico *Material organico saprico (SiBCS Cap 14)*

Description

Material organico altamente decomposto: < 17% de fibras esfregadas OU indice de von Post H7-H10. Discrimina Organossolos Sapricos no 3o nivel categorico.

Usage

```
saprico(pedon)
```

Arguments

pedon	A PedonRecord.

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 14 (Organossolos), pp 224-226.

save_ossl_models — *Save / load trained OSSL-backed PLSR models*

Description

Thin wrappers around `saveRDS` / `readRDS` that also verify the deserialised object's shape. The on-disk file carries the soilKey version, training time and preprocess label as attributes; `load_ossl_models` preserves them.

Usage

```
save_ossl_models(models, path)

load_ossl_models(path)
```

Arguments

`models`	Output of `train_pls_from_ossl`.
`path`	File path. Use `.rds` or `.RData` as the suffix (saveRDS is used regardless).

Value

`save_ossl_models()` returns `path` invisibly. `load_ossl_models()` returns the model list.

shrink_swell_cracks — *Shrink-swell cracks (WRB 2022 Ch 3.2.12) – per-pedon test wrapping* `test_shrink_swell_cracks`.

Description

Shrink-swell cracks (WRB 2022 Ch 3.2.12) – per-pedon test wrapping `test_shrink_swell_cracks`.

Usage

```
shrink_swell_cracks(pedon, min_width_cm = 0.5)
```

Arguments

`pedon`	A `PedonRecord`.
`min_width_cm`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

sideralic_properties *Sideralic properties (WRB 2022 Ch 3.2.13)*

Description

Mineral material with a relatively low CEC. WRB 2022 (3.2.13) requires BOTH:

1. one or both of: clay >= 8% AND CEC/clay < 24 cmol_c/kg clay; OR bulk CEC < 2 cmol_c/kg soil;
2. evidence of soil formation as defined in criterion 3 of the cambic horizon (`test_cambic_soil_formation`).

Both must be met by the SAME layer. Criterion 2 was added in v0.9.127 (previously only criterion 1 was enforced); where the soil-formation evidence cannot be assessed (no Munsell/clay/Fe/carbonate adjacency data) the result is `NA` rather than a false positive.

Usage

```
sideralic_properties(pedon, max_cec_per_clay = 24, max_bulk_cec = 2)
```

Arguments

`pedon`	A `PedonRecord`.
`max_cec_per_clay`	Numeric threshold or option (see Details).
`max_bulk_cec`	Numeric threshold or option (see Details).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

site_attribute_spec *Metadata of the site properties*

Description

One row per site property that soilKey recognises (the named elements of `PedonRecord$site`), with the section it belongs to, its type, unit and description, the values a categorical property takes where the keys compare it with a fixed set, and the classification keys that read it. The properties of the last section, *reference classes and provenance*, are stored by the loaders and reports and read by no key. The metadata come from `inst/schemas/site-attributes.yaml`; the attribute reference `inst/ATTRIBUTES.md` lists them after the horizon attributes.

Usage

```
site_attribute_spec()
```

Value

A `data.frame` with the columns `name`, `type`, `section`, `unit`, `description`, and `allowed` and `systems` (comma-separated, empty when none).

See Also

`horizon_attribute_spec`, `unrecognised_attributes`.

Examples

```
s <- site_attribute_spec()
s[s$name %in% c("drainage_class", "soil_moisture_regime"),
  c("name", "type", "description")]
```

---

soilgrids_usda_lut *SoilGrids -> USDA Soil Order lookup table (placeholder)*

Description

Reserved for the future SoilGrids USDA layer. Currently returns the 12 USDA Order codes mapped to integers 1..12.

Usage

```
soilgrids_usda_lut()
```

Value

Named character vector.

---

soilgrids_wrb_lut *SoilGrids -> WRB code lookup table*

Description

Maps the integer raster values used by the SoilGrids 2.0 "MostProbable WRB" layer to soilKey's two-letter RSG codes (the codes used in `inst/rules/wrb2022/key.yaml`).

Usage

```
soilgrids_wrb_lut()
```

Details

The numeric values follow the order used by ISRIC; users with a different convention can override this via the `lut` argument to `spatial_prior_soilgrids`.

Value

Named character vector: names are integer-as-character ("1", "2", ...), values are RSG codes.

---

soilkey_systems *Classification systems known to soilKey*

Description

Lists the classification systems registered in the package, one row per system, in the order classify_all runs them. Each system is a folder inst/rules/<id>/ holding its key and a system.yaml entry. The id is what load_rules and classify_all accept; the aliases are accepted too.

Usage

```
soilkey_systems()
```

Value

A data.frame with one row per system and the columns id, name (English display name), edition, status, classifier (the function that classifies a PedonRecord in that system), levels and aliases (both comma-separated), and default (whether classify_all(systems = "all") runs the system).

See Also

classify_all, load_rules.

Examples

```
soilkey_systems()
```

---

soil_classes_at_location *Likely soil classes at a geographic location (spatial classification aid)*

Description

Returns a ranked list of the soil Reference Soil Groups (or SiBCS ordens, or USDA orders) most likely to occur at the given point, based on a global or regional dominant-soil raster (SoilGrids 2.0 by default). This is the **before-you-have-a-pedon helper**: a pedologist arriving in the field can call it with the GPS coordinates of the planned profile pit and see which classes are expected, plus what attributes typically distinguish them.

Usage

```
soil_classes_at_location(
  lat,
  lon,
  system = c("wrb2022", "sibcs", "usda"),
  buffer_m = 1000,
  source_url = NULL,
  top_n = 5,
  verbose = TRUE
)
```

Arguments

lat, lon	Numeric WGS-84 coordinates.
system	Classification system. One of "wrb2022" (default), "sibcs", "usda".
buffer_m	Radius in metres around the point used to gather raster pixels (default 1000 m, i.e. roughly 4 SoilGrids pixels).
source_url	Path / URL of the dominant-soil raster.
top_n	Keep the top N classes by probability (default 5).
verbose	Emit a cli summary.

Details

This function does **not** classify a profile. The deterministic key in classify_wrb2022 / classify_sibcs / classify_usda remains the only thing that assigns a class from horizon data. The output here is purely informational – a "shopping list" of what to confirm.

Value

A list as described under **Output**.

Data source

For real use, point source_url at a regional SoilGrids "MostProbable WRB" GeoTIFF / COG (one of the cuts at https://files.isric.org/soilgrids/latest/data/wrb/). For tests, options(soilKey.test_raster = "/tmp/syn.tif") is honoured. When no source is given, the function emits a cli_alert_warning() and returns an empty result – it does **not** pretend to know.

Output

A list with three elements:

distribution A data.table with columns rsg_code, rsg_name, probability, sorted by descending probability.

typical_attributes A data.table keyed by rsg_code with the canonical attribute ranges that distinguish each class (clay range, CEC range, BS range, etc.). The values come from the WRB 2022 / SiBCS 5 / KST 13ed canonical thresholds, NOT from the raster.

site The site list passed in, plus the buffer radius and the source URL.

See Also

spatial_prior_soilgrids for the post-classification consistency check.

Examples

```
## Not run:
# Mata Atlântica, Rio de Janeiro state.
res <- soil_classes_at_location(
  lat        = -22.7,
  lon        = -43.7,
  system     = "wrb2022",
  source_url = "https://files.isric.org/soilgrids/latest/data/wrb/MostProbable.vrt"
)
res$distribution         # ranked list of likely RSGs
res$typical_attributes   # canonical thresholds per RSG to confirm

## End(Not run)
```

---

soil_organic_carbon *Soil organic carbon (WRB 2022 Ch 3.3.16): organic C that does NOT belong to artefacts. v0.3.3: any layer with oc_pct >= 0.1 and artefacts_industrial_pct < 35.*

Description

Soil organic carbon (WRB 2022 Ch 3.3.16): organic C that does NOT belong to artefacts. v0.3.3: any layer with oc_pct >= 0.1 and artefacts_industrial_pct < 35.

Usage

```
soil_organic_carbon(pedon, min_oc = 0.1, max_artefacts = 35)
```

Arguments

pedon	A PedonRecord.
min_oc	Numeric threshold or option (see Details).
max_artefacts	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

solimovic_material	*Solimovic material (WRB 2022 Ch 3.3.17): hetero genous mass-movement material on slopes / footslopes (formerly "colluvic"). v0.3.3: detects via* `rock_origin == "colluvial"` *OR* `layer_origin == "solimovic"`.

Description

Solimovic material (WRB 2022 Ch 3.3.17): hetero genous mass-movement material on slopes / footslopes (formerly "colluvic"). v0.3.3: detects via `rock_origin == "colluvial"` OR `layer_origin == "solimovic"`.

Usage

```
solimovic_material(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

sombric	*Sombric horizon (WRB 2022): subsurface accumulation of humus that qualified neither as spodic nor as a true mollic-like horizon (low-base-saturation cool tropical highlands). v0.3.3 detects via designation pattern + OC criteria (BS < 50, OC > 0.6, depth > 25 cm).*

Description

Sombric horizon (WRB 2022): subsurface accumulation of humus that qualified neither as spodic nor as a true mollic-like horizon (low-base-saturation cool tropical highlands). v0.3.3 detects via designation pattern + OC criteria (BS < 50, OC > 0.6, depth > 25 cm).

Usage

```
sombric(
  pedon,
  min_thickness = 15,
  min_oc = 0.6,
  max_bs = 50,
  min_top_cm = 25,
  min_oc_increase = 0.1
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
min_oc	Numeric threshold or option (see Details).
max_bs	Numeric threshold or option (see Details).
min_top_cm	Numeric threshold or option (see Details).
min_oc_increase	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

spatial_prior *Spatial prior over RSGs (or Orders) at a pedon's location*

Description

Top-level dispatcher. Reads a categorical raster of soil classes (SoilGrids globally, Embrapa for Brazil), buffers the pedon's coordinates, tallies pixel classes within the buffer, and returns the empirical class frequency as a probability distribution.

Usage

```
spatial_prior(
  pedon,
  source = c("soilgrids", "embrapa"),
  system = c("wrb2022", "usda"),
  ...
)
```

Arguments

pedon	A PedonRecord with non-NULL site$lat / site$lon.
source	Backend to query: "soilgrids" (default) or "embrapa".
system	Classification system: "wrb2022" (default) or "usda". Embrapa source forces "sibcs5" internally regardless of this argument.
...	Passed through to the backend (spatial_prior_soilgrids or spatial_prior_embrapa).

Details

The prior is intentionally separate from the deterministic key. Pass the returned data.table to classify_wrb2022 via the prior argument; the result will then carry a prior_check entry (consistent / inconsistent / not_run).

Value

A `data.table` with columns `rsg_code` (character) and `probability` (numeric, summing to 1). Empty if the buffer extracts no valid pixels – callers should check `nrow()`.

`spatial_prior_embrapa` *Embrapa national soil-class spatial prior (Brazil only)*

Description

v0.5 stub. Reads a user-provided categorical raster of SiBCS orders / suborders, buffers the pedon's site, tallies pixel classes, and returns a probability distribution over SiBCS codes (or, with a user-provided LUT, over WRB equivalents).

Usage

```
spatial_prior_embrapa(
  pedon,
  raster_path = NULL,
  buffer_m = 3750,
  lut = NULL,
  n_classes_top = 10,
  ...
)
```

Arguments

`pedon`	A `PedonRecord`.
`raster_path`	Required. Path to a local categorical raster (GeoTIFF) of Embrapa SiBCS classes. There is no built-in file in v0.5 – download the polygon map from `https://www.embrapa.br/solos/sibcs` and rasterise it.
`buffer_m`	Buffer radius in metres (default 3750, i.e. ~15-cell neighbourhood at 250 m resolution).
`lut`	Optional named character vector mapping raster integer values to soil-class codes. If NULL, raster categories are used as-is (terra::levels).
`n_classes_top`	Keep only the top N classes (default 10).
`...`	Reserved.

Details

Unlike SoilGrids, Embrapa does not publish per-pixel probabilities, so the empirical frequency over a neighbourhood window (default 15 x 15 cells = ~3.75 km radius at 250 m resolution) is used as an approximation.

Value

A `data.table` with columns `rsg_code`, `probability`.

spatial_prior_soilgrids *SoilGrids spatial prior*

Description

Reads a categorical raster of dominant Reference Soil Groups around the pedon's site, buffers the point in metric coordinates, extracts all pixel values within the buffer, and returns the empirical class frequency as a probability distribution over RSG codes.

Usage

```
spatial_prior_soilgrids(
  pedon,
  system = c("wrb2022", "usda"),
  buffer_m = 250,
  source_url = NULL,
  n_classes_top = 10,
  lut = NULL,
  ...
)
```

Arguments

pedon	A PedonRecord with non-NULL site$lat and site$lon.
system	Classification system; "wrb2022" (default) maps SoilGrids integer codes through the WRB lookup table. "usda" is reserved for a future SoilGrids-USDA layer.
buffer_m	Buffer radius in metres around the point (default 250 m, i.e. one SoilGrids pixel).
source_url	Optional. A path or URL accepted by terra::rast. If NULL, falls back to getOption("soilKey.test_raster").
n_classes_top	Keep only the top N classes by frequency (default 10). Set to Inf to keep all.
lut	Optional named integer vector mapping raster values to RSG codes. Default is soilgrids_wrb_lut; pass a custom one if your raster uses different codes.
...	Reserved for future use.

Value

A data.table with columns rsg_code, probability.

Data source

For real use, pass source_url pointing at a SoilGrids "MostProbable WRB" GeoTIFF / COG, e.g. one of the regional cuts published at https://files.isric.org/soilgrids/latest/data/wrb/. For tests, set options(soilKey.test_raster = "/path/to/syn.tif") to point at a local synthetic raster – this avoids network access in CI.

Coordinate handling

We use `sf::st_transform` when sf is available; otherwise we fall back to `terra::project` on a single-point SpatVector. The buffer is constructed in metric (UTM) coordinates so `buffer_m` is in metres regardless of the pedon CRS. The raster itself is queried in its native CRS via terra's automatic reprojection.

See Also

`spatial_prior`, `soilgrids_wrb_lut`.

`spodic` *Spodic horizon (WRB 2022)*

Description

Tests whether any horizon meets the spodic horizon criteria. The spodic horizon is an illuvial horizon with active Al + Fe oxalate- extractable material plus organic matter; diagnostic of Podzols.

Usage

```
spodic(
  pedon,
  min_thickness = 2.5,
  min_alfe = 0.5,
  max_ph = 5.9,
  min_oc_in_b = 0.5,
  engine = NULL
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness in cm (default 2.5).
`min_alfe`	Minimum (Al_ox + 0.5 * Fe_ox) percent (default 0.5).
`max_ph`	Maximum ph_h2o (default 5.9).
`min_oc_in_b`	Minimum OC % in the candidate Bh / Bs layer for the v0.9.19 morphological inference path when Al / Fe oxalate are missing (default 0.5).
`engine`	One of `"soilkey"` (default; strict v0.9.19 morphological path requires Bh / Bs / Bhs designation + albic E above) or `"aqp"` (relaxed v0.9.84 path: any B* below E* with OC translocation peak). When `NULL`, reads `getOption("soilKey.diagnostic_engine")`.

Details

Sub-tests:

- `test_spodic_aluminum_iron` – (Al_ox + 0.5*Fe_ox) >= 0.5%
- `test_ph_below` – ph_h2o <= 5.9
- `test_minimum_thickness` – thickness >= 2.5 cm

v0.2 limitations: the WRB color criterion (hue 5YR or yellower with chroma <= 5, or specific dark colors) is not enforced. The (Al_ox + Fe_ox)/clay >= 0.05 alternative ratio test is not yet wired. Both deferred to v0.3.

Value

A `DiagnosticResult`.

v0.9.84 engine="aqp" relaxation

KSSL+NASIS Spodosols routinely use generic "B1" / "B2" / "Bw" designations rather than the specific Bh / Bs / Bhs that the v0.9.19 morphological-inference path requires. Of 14 KSSL+NASIS Podzol references, only 1 / 14 passes spodic via the v0.9.19 path; 7 / 14 have BOTH an E-designated albic-eligible horizon above AND an OC peak in a B horizon below (the canonical Podzol illuviation signature) but use generic B / Bw designations and so fail strict morph.

When `engine = "aqp"` (read from `getOption("soilKey.diagnostic_engine", "soilkey")` when `engine` is `NULL`) AND Al / Fe oxalate is unmeasured AND the v0.9.19 strict path did not fire, accept any `B*` designation below an `E*`-designated horizon when:

- `ph_h2o` <= `max_ph` in the B horizon, AND
- `oc_pct` >= `min_oc_in_b` in the B horizon, AND
- OC in the B is greater than the maximum OC in any horizon above (the translocation signature).

Default `engine` is `"soilkey"` – canonical behaviour bit-for-bit preserved.

References

IUSS Working Group WRB (2022), Chapter 3, Spodic horizon.

`stagnic_properties` *Stagnic properties (WRB 2022)*

Description

Tests for stagnic properties (WRB 2022 Ch 3.2.14, p. 81) in a layer starting within the upper `max_top_cm` (default 100).

Usage

```
stagnic_properties(
  pedon,
  max_top_cm = 100,
  min_redox_pct = 5,
  decay_factor = 3
)
```

Arguments

pedon	A PedonRecord.
max_top_cm	Maximum top depth (cm) of candidate layers (default 100).
min_redox_pct	The oximorphic features are more than this % of the layer (default 5).
decay_factor	Not used since v0.9.312 (the decay test was removed); kept so that existing calls still work.

Details

Since v0.9.312 the layer reading is that of gleyic_properties (Hugo's reading for descriptions that do not say where the mottles sit): a layer with more than 5% oximorphic features that are not gleyic (criteria 1-3: the matrix has no reductimorphic colour, the layer does not lie directly on a layer with such a matrix, and the site records no groundwater within 150 cm that is not perched, or a claric layer lies directly above it); a layer with claric colours in 95% or more of its area above an abrupt textural difference or a layer of 1.5 kg/dm3 or more (criterion 4), or above a layer with stagnic oximorphic features (criterion 5). Mottles with neither colour nor water evidence leave the layer NA. Until v0.9.311 a layer with 5% redox features passed when the features decayed with depth, which is not in WRB 2022.

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 3.2.14, Stagnic properties.

---

st_features_canonical *USDA Soil Taxonomy diagnostic features canonical table*

Description

Convenience wrapper for canonical_reference("ST_features"). Returns an 84-row data.frame with one row per diagnostic feature (epipedon / subsurface horizon / property / material) and columns: group, name, chapter, page, description, criteria. The criteria column is a list-column; each element holds the parsed criteria text per feature.

Usage

```
st_features_canonical(prefer_pkg = TRUE)
```

Arguments

prefer_pkg	Retained for compatibility and no longer used. Until v0.9.203 FALSE selected a copy bundled with soilKey; there is no bundled copy any more, so the data always come from SoilTaxonomy.

Value

The canonical Soil Taxonomy diagnostic-features reference (a list / data.frame).

---

subgrupo_planossolo_espessos

Subgrupo "espessos" de Planossolos (B planico profundo, > 100 cm)

Description

Discrimina os Subgrupos espessos de Planossolos (Cap 15: SNs Espessos, SNo Espessos, SXs Espessos, SXal Espessos, SXd Espessos, SXe Espessos): B planico cujo topo ocorre entre min_top_cm (exclusivo) e max_top_cm (inclusivo).

Usage

```
subgrupo_planossolo_espessos(pedon, min_top_cm = 100, max_top_cm = 200)
```

Arguments

pedon	A PedonRecord.
min_top_cm	Profundidade minima exclusiva do topo do B planico (default 100; passa se top > 100).
max_top_cm	Profundidade maxima inclusiva (default 200).

Details

Implementacao: identifica B planico via B_planico, captura o topo (mais raso) das camadas que passam, e testa se cai em (min_top_cm, max_top_cm].

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 15 (Planossolos), pp 251-260.

subgrupo_planossolo_mesicos

Subgrupo "mesicos" de Planossolos (B planico topo em [50, 100] cm)

Description

Discrimina os Subgrupos mesicos de Planossolos (Cap 15: SNs Mesicos, SNo Mesicos, SXs Mesicos, SXal Mesicos, SXd Mesicos, SXe Mesicos): B planico cujo topo ocorre entre `min_top_cm` (inclusivo) e `max_top_cm` (inclusivo).

Usage

```
subgrupo_planossolo_mesicos(pedon, min_top_cm = 50, max_top_cm = 100)
```

Arguments

`pedon`	A `PedonRecord`.
`min_top_cm`	Profundidade minima inclusiva (default 50).
`max_top_cm`	Profundidade maxima inclusiva (default 100).

Value

`DiagnosticResult`.

References

Embrapa (2018), SiBCS 5a ed., Cap 15 (Planossolos).

subgrupo_plintossolo_endico_concrecionario

Subgrupo "endico" de Plintossolos Concrecionarios (topo de horizonte concrecionario >= 40 cm)

Description

Discrimina o Subgrupo FFcoEn (Plintossolos Petricos Concrecionarios endicos): horizonte concrecionario cujo topo ocorre a >= `min_top_cm` cm.

Usage

```
subgrupo_plintossolo_endico_concrecionario(pedon, min_top_cm = 40)
```

Arguments

`pedon`	A `PedonRecord`.
`min_top_cm`	Profundidade minima inclusiva (default 40).

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 16, p 264.

`subgrupo_plintossolo_endico_litoplintico`
Subgrupo "endico" de Plintossolos Litoplinticos (topo de horizonte litoplintico >= 40 cm)

Description

Discrimina o Subgrupo FFlpEn (Plintossolos Petricos Litoplinticos endicos): horizonte litoplintico cujo topo ocorre a >= `min_top_cm` cm.

Usage

```
subgrupo_plintossolo_endico_litoplintico(pedon, min_top_cm = 40)
```

Arguments

`pedon`	A PedonRecord.
`min_top_cm`	Profundidade minima inclusiva (default 40).

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 16, p 264.

subgrupo_plintossolo_espessos

Subgrupo "espessos" de Plintossolos (horizonte plintico topo > 100 cm)

Description

Discrimina os Subgrupos espessos de Plintossolos Argiluvicos (FT*Es) e Haplicos (FXacEs, FXdEs, FXeEs): horizonte plintico cujo topo ocorre entre `min_top_cm` (exclusivo) e `max_top_cm` (inclusivo).

Usage

```
subgrupo_plintossolo_espessos(pedon, min_top_cm = 100, max_top_cm = 200)
```

Arguments

pedon	A PedonRecord.
min_top_cm	Profundidade minima exclusiva (default 100).
max_top_cm	Profundidade maxima inclusiva (default 200).

Value

DiagnosticResult.

References

Embrapa (2018), SiBCS 5a ed., Cap 16 (Plintossolos), pp 261-272.

takyric_properties

Takyric properties (WRB 2022 Ch 3.2.15) – per-pedon test wrapping `test_takyric_surface`.

Description

Takyric properties (WRB 2022 Ch 3.2.15) – per-pedon test wrapping `test_takyric_surface`.

Usage

```
takyric_properties(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

technic_features *Technic features (WRB 2022)*

Description

Tests for any of three WRB 2022 alternative qualifying conditions for Technosols:

1. Artefacts >= artefacts_min_pct (default 20%) by volume within the upper max_top_cm (default 100 cm).
2. A continuous geomembrane (geomembrane_present == TRUE) within the upper 100 cm.
3. Technic hard material (concrete, asphalt, mine spoil) with technic_hardmaterial_pct >= hardmaterial_min_pct (default 95%) at the surface (top_cm <= hardmaterial_max_top_cm, default 5).

Either path qualifies.

Usage

```
technic_features(
  pedon,
  artefacts_min_pct = 20,
  max_top_cm = 100,
  hardmaterial_min_pct = 95,
  hardmaterial_max_top_cm = 5
)
```

Arguments

pedon	A PedonRecord.
artefacts_min_pct	Minimum artefact percent (default 20).
max_top_cm	Maximum top depth (cm) for the artefact and geomembrane paths (default 100).
hardmaterial_min_pct	Minimum hard-material coverage (%) for the technic-hard-material path (default 95).
hardmaterial_max_top_cm	Surface depth window (cm) for the technic-hard-material path (default 5).

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 5, Technosols.

`technic_hard_material` *Technic hard material (WRB 2022 Ch 3.3.18): consolidated human-made material (asphalt, concrete, worked stones).*

Description

Technic hard material (WRB 2022 Ch 3.3.18): consolidated human-made material (asphalt, concrete, worked stones).

Usage

```
technic_hard_material(pedon)
```

Arguments

pedon	A PedonRecord.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`tephric_material` *Tephric material (WRB 2022 Ch 3.3.19): \>= 30% volcanic glass in 0.02-2 mm fraction AND no andic / vitric properties.*

Description

Tephric material (WRB 2022 Ch 3.3.19): \>= 30% volcanic glass in 0.02-2 mm fraction AND no andic / vitric properties.

Usage

```
tephric_material(pedon, min_glass = 30)
```

Arguments

pedon	A PedonRecord.
min_glass	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

terric — *Terric horizon (WRB 2022): topsoil thickened by long-term application of mineral material (sediment / sand additions). v0.3.3: thickness >= 20 cm + designation Au / Apc.*

Description

Terric horizon (WRB 2022): topsoil thickened by long-term application of mineral material (sediment / sand additions). v0.3.3: thickness >= 20 cm + designation Au / Apc.

Usage

```
terric(pedon, min_thickness = 20)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

texture_class_from_pct — *NRCS texture-class shorthand from clay / silt / sand percent*

Description

aqp's getArgillicBounds() requires an NRCS texture class column (e.g. "SCL", "C", "CL", "FS"). soilKey horizons only carry the percent fractions; this helper derives the class from the standard USDA texture triangle.

Usage

```
texture_class_from_pct(clay, silt, sand)
```

Arguments

clay	Numeric vector of clay percent (0-100).
silt	Numeric vector of silt percent.
sand	Numeric vector of sand percent. (clay + silt + sand should sum to ~100; mild deviations are tolerated.)

Details

Returns the standard NRCS abbreviation:

COS	Coarse sand
S	Sand
FS	Fine sand
VFS	Very fine sand
LS	Loamy sand
LFS	Loamy fine sand
SL	Sandy loam
FSL	Fine sandy loam
L	Loam
SIL	Silt loam
SI	Silt
SCL	Sandy clay loam
CL	Clay loam
SICL	Silty clay loam
SC	Sandy clay
SIC	Silty clay
C	Clay

Implementation follows the canonical USDA texture triangle; vector- ised over the input. NA in / NA out.

Value

Character vector of NRCS texture class abbreviations.

---

thionic

Thionic horizon (WRB 2022): post-oxidation acid sulfate horizon. Requires sulfidic_s_pct >= 0.01 AND pH(H2O) <= 4.

Description

Thionic horizon (WRB 2022): post-oxidation acid sulfate horizon. Requires sulfidic_s_pct >= 0.01 AND pH(H2O) <= 4.

Usage

```
thionic(
  pedon,
  min_thickness = 15,
  max_pH = 4,
  min_sulfidic_s = 0.01,
  adjoining = FALSE
)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).
max_pH	Numeric threshold or option (see Details).
min_sulfidic_s	Numeric threshold or option (see Details).
adjoining	FALSE (default): the thickness is asked of each layer. TRUE: of adjoining layers taken together, as the SiBCS horizonte sulfurico reads it since v0.9.246.

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

---

train_pls_from_ossl *Train pre-trained PLSR models from an OSSL library*

Description

Iterates over properties and fits one PLSR model per target against the OSSL spectra in ossl_library$Xr, with internal cross-validation to pick the optimal number of components per property. The returned list is a drop-in replacement for the ossl_models argument of predict_ossl_pretrained and fill_from_spectra.

Usage

```
train_pls_from_ossl(
  ossl_library,
  properties = c("clay_pct", "sand_pct", "silt_pct", "cec_cmol", "ph_h2o", "oc_pct"),
  ncomp_max = 20L,
  validation = c("CV", "LOO", "none"),
  segments = 10L,
  preprocess = "snv+sg1",
  min_n = 50L,
  verbose = TRUE
)
```

Arguments

ossl_library	A list with two named elements: Xr (numeric matrix of training spectra) and Yr (data.frame keyed by property name, one row per training spectrum). See ossl_library_template.
properties	Character vector of column names in ossl_library$Yr to train models for. Defaults to the six core soil properties exposed by OSSL.
ncomp_max	Integer. Upper bound on the number of PLS components to consider during cross-validation. Defaults to 20.

validation	One of "CV" (default, k-fold), "LOO" (leave-one-out, slow), "none" (uses ncomp_max components without selection).
segments	Number of CV segments when validation = "CV". Default 10.
preprocess	Pre-processing label passed to preprocess_spectra. Stored on the trained models so predict_from_spectra can reapply it.
min_n	Minimum number of valid training samples (after dropping rows with non-finite y or X). Properties below this threshold are skipped with a warning. Default 50.
verbose	If TRUE (default), prints a per-property summary on completion.

Details

Spectra are pre-processed inside the function (default "snv+sg1"); the same preprocessing is used downstream by predict_from_spectra so the user does not have to remember which transform was applied at training time.

Value

A named list of soilKey_pls_model objects, one per successfully trained property. Carries trained_at, soilKey_version and preprocess attributes for provenance.

Examples

```
## Not run:
lib <- download_ossl_subset(region = "south_america")
models <- train_pls_from_ossl(lib,
                              properties = c("clay_pct", "ph_h2o"))
result <- predict_from_spectra(my_pedon, models = models)

## End(Not run)
```

---

tsitelic *Tsitelic horizon (WRB 2022 Ch 3.1)*

Description

From Georgian *tsiteli* = red. A red colour-defined horizon formed on weathered basalt or similar Fe-rich parent material in Caucasian / Mediterranean settings. Used by the Cambisols key (Ch 4 p 123, criterion 4) and by the Tsitelic qualifier.

Usage

```
tsitelic(pedon, min_thickness = 10)
```

Arguments

pedon	A PedonRecord.
min_thickness	Numeric threshold or option (see Details).

Details

Diagnostic criteria (v0.3.5 simplification):

- Munsell hue \<= 2.5YR (i.e. 2.5YR, 10R, 7.5R, 5R, 2.5R) AND value \<= 4 (moist) AND chroma \>= 4 (moist);
- evidence of soil formation (cambic-style criterion 3) proxied by clay \>= 8% AND `structure_grade` not "single grain" / "massive";
- thickness \>= 10 cm.

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

`umbric_horizon` *Umbric horizon (WRB 2022)*

Description

Tests for the umbric horizon – a thick, dark, organic-rich surface horizon like mollic, but with low base saturation (< 50%). Diagnostic of Umbrisols.

Usage

```
umbric_horizon(
  pedon,
  min_thickness = 20,
  min_oc = 0.6,
  max_bs = 50,
  surface_top_cm = 5
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_thickness`	Minimum thickness (cm; default 20).
`min_oc`	Minimum SOC % (default 0.6).
`max_bs`	Maximum base saturation % (default 50; profile must be BELOW this).
`surface_top_cm`	Maximum depth (cm) below the mineral soil surface at which the stack may start (default 5).

Details

Implementation reuses the mollic sub-tests on the contiguous stack of dark layers from the mineral soil surface (organic layers on top are not part of it), and tests the weighted average base saturation of the stack against `max_bs` (WRB 2022 3.1.39, pp. 69-70).

Value

A DiagnosticResult.

References

IUSS Working Group WRB (2022), Chapter 3.1.39, Umbric horizon (pp. 69-70).

`unrecognised_attributes` *Attributes that soilKey does not recognise*

Description

Lists the horizon columns and site properties that no classification key reads: a horizon column that is neither a canonical name of horizon_column_spec nor one that ensure_horizon_schema renames onto it (a difference of capitalisation or separators, or a common alias such as SOC), and a site property that is not in site_attribute_spec. Such attributes are kept in a PedonRecord but have no effect on a classification. PedonRecord$new() warns about them when it is called from outside the package, and so does read_pedon_csv; options(soilKey.warn_unrecognised = FALSE) turns the warning off.

Usage

```
unrecognised_attributes(x, site = NULL)
```

Arguments

x	A PedonRecord, or a data.frame of horizons.
site	A list of site properties, when x is a data.frame; ignored for a PedonRecord, whose own site is read.

Value

A data.frame with one row per attribute that is not recognised and the columns where ("horizons" or "site"), name, suggestion (the closest recognised names, comma-separated, empty when none) and note: empty, or "site property" for a horizon column that is a site property, "horizon attribute" for a site property that is a horizon column, or "already given" when the column it would be renamed to is also present. No rows when everything is recognised.

See Also

horizon_attribute_spec, site_attribute_spec, inst/ATTRIBUTES.md.

Examples

```
h <- data.frame(top_cm = c(0, 20), bottom_cm = c(20, 60),
                clay = c(30, 45), organic_carbon_gkg = c(11, 4),
                soil_moisture_regime = "ustic")
unrecognised_attributes(h, site = list(id = "p1", drainge_class = "well"))
```

usda_to_wrb_rsg *USDA Soil Taxonomy <-> WRB Reference Soil Group correlation table*

Description

Returns the single most-common WRB RSG for a given USDA Order + optional Suborder. Based on IUSS WRB (2022) Annex 6.

Usage

```
usda_to_wrb_rsg(usda_order, usda_suborder = NULL)
```

Arguments

usda_order	Character vector of USDA Order names. Case- insensitive; trailing 's' stripped (e.g.\ both "Mollisols" and "Mollisol" accepted).
usda_suborder	Optional character vector of USDA Suborder names (case-insensitive) used to refine the mapping. Same length as usda_order or recycled.

Value

Character vector of WRB Reference Soil Group names (singular, no plural 's'). NA for unrecognised inputs.

Caveat

This is a "best-guess" cross-walk for benchmark validation only. Real-world correlation requires per-pedon evaluation of WRB diagnostic horizons. Use this function to derive a reasonable *expected* WRB classification from a USDA-classified pedon (e.g.\ from KSSL/NASIS) so that classify_wrb2022() can be validated against an external taxonomy on the same profiles.

References

IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition, Annex 6. International Union of Soil Sciences, Vienna.

Examples

```
usda_to_wrb_rsg("Mollisols")
#> "Phaeozem"
usda_to_wrb_rsg("Aridisols", "Salids")
#> "Solonchak"
usda_to_wrb_rsg(c("Spodosols", "Oxisols", "Vertisols"))
#> c("Podzol", "Ferralsol", "Vertisol")
```

validate_horizon_geometry *Validate horizon depth geometry*

Description

A pure, side-effect-free check of a horizon table's depth geometry, independent of any PedonRecord. The Pro app's Pedon builder calls it to give immediate feedback while horizons are edited, and it is a handy guard before constructing a profile from an untrusted CSV.

Usage

```
validate_horizon_geometry(horizons)
```

Arguments

horizons	A data frame with at least numeric top_cm and bottom_cm columns (and optionally a designation column).

Details

It reports two severities:

errors (these make a sane classification impossible) a missing or non-numeric top_cm/bottom_cm; a negative depth; a horizon whose top_cm >= bottom_cm (inverted or zero thickness); two horizons whose depths overlap.

warnings (allowed, but worth surfacing) the shallowest horizon not starting at the surface (0 cm); a gap between consecutive horizons; horizons entered out of increasing-depth order; a duplicated horizon designation.

This complements PedonRecord$validate(), which additionally checks chemistry (texture sums, pH, CEC vs bases, Munsell ranges); use that for a built record and this for a raw table.

Value

A list with valid (logical; TRUE when there are no errors), errors and warnings (character vectors of human-readable English messages), and details – a named list of the offending row indices (or values) per check, so a caller can compose its own (e.g. localised) messages.

Examples

```
h <- data.frame(top_cm = c(0, 20, 55), bottom_cm = c(20, 55, 90),
                designation = c("A", "AB", "Bt"))
validate_horizon_geometry(h)$valid          # TRUE

bad <- data.frame(top_cm = c(0, 40), bottom_cm = c(50, 30))  # overlap+inverted
validate_horizon_geometry(bad)$errors
```

validate_pedon_json *Validate a PedonRecord against the JSON schema*

Description

Convenience wrapper that converts a PedonRecord (or a compatible list) to JSON and validates it via jsonvalidate::json_validate against the canonical schema returned by pedon_json_schema.

Usage

```
validate_pedon_json(x)
```

Arguments

x	A PedonRecord or a list with the same shape.

Details

Use this BEFORE calling classify_* when ingesting data from external systems (web APIs, ETL pipelines, multimodal extraction) to catch schema violations early.

Value

A logical scalar (TRUE when valid). Validation errors appear as the errors attribute when FALSE.

Examples

```
## Not run:
p <- make_ferralsol_canonical()
validate_pedon_json(p)
#> [1] TRUE

## End(Not run)
```

vertic_horizon *Vertic horizon (WRB 2022 Ch 3.1.40)*

Description

A clay-rich subsurface horizon with slickensides or wedge-shaped aggregates, formed by shrinking and swelling. Used by Vertisols.

Usage

```
vertic_horizon(
  pedon,
  min_clay = 30,
  min_thickness = 25,
  min_le_cm = 6,
  le_max_depth_cm = 100,
  min_crack_width_cm = 0.5,
  reading = c("wrb2022", "usda")
)
```

Arguments

pedon	A PedonRecord.
min_clay	Numeric threshold or option (see Details).
min_thickness	Numeric threshold or option (see Details).
min_le_cm	Minimum LE sum (cm) for the COLE-based path of reading = "usda" (default 6, per KST 13ed Ch 16).
le_max_depth_cm	Depth window (cm) for the COLE-based path of reading = "usda" (default 100).
min_crack_width_cm	Minimum shrink-swell crack width (cm). Defaults to 0.5 (WRB/USDA); the SiBCS horizonte_vertico wrapper passes 1.0 per Embrapa (2018) Cap 2 p.73.
reading	"wrb2022" (default): the criteria of WRB 2022 Ch 3.1.40. "usda": the reading of the USDA Vertisols and the SiBCS horizonte vertico, with the KST 13 COLE path (see Details).

Details

Since v0.9.311 the default reading (reading = "wrb2022") follows the diagnostic criteria of WRB 2022 Ch 3.1.40 (p. 70). Mineral material with:

1. \>= 30% clay;
2. wedge-shaped aggregates (a structure_type that names them) or slickensides on \>= 10% of the aggregate surfaces (slickensides "common" or more);
3. shrink-swell cracks: cracks_width_cm \>= 0.5, or, where the width is not recorded, slickensides "common" or more or an ss designation as evidence of them;
4. a thickness of \>= 25 cm, of adjoining layers that meet 1 to 3 taken together.

Each criterion is read per layer as met, not met or unrecorded. The horizon is NA only when layers that might meet them (a criterion unrecorded) could make up the thickness. Organic layers and continuous rock are not part of it. WRB gives the COLE (\>= 0.06) as additional information only, so this reading does not use cole_value.

reading = "usda" keeps the reading that the USDA Vertisols and the SiBCS horizonte vertico are built on, as it was before v0.9.311: slickensides "common" or more and cracks \>= min_crack_width_cm

in one layer of `min_thickness`, OR the KST 13 linear-extensibility alternative: summed `cole_value * thickness` over the upper `le_max_depth_cm` \>= `min_le_cm` (KST 13ed Ch 16, p. 343).

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

v0.9.72 designation morphological inference (opt-in)

Field-described Brazilian Vertissolos profiles (e.g.\ the Embrapa Redape curated dataset) encode vertic morphology via a `v` master-letter modifier in the horizon designation (`Bv`, `Bvk1`, `Cv`, `Cvz`) without recording `slickensides` class or `shrink_swell_cracks_cm` as numeric inputs. With `options(soilKey.vertic_designation_inference = TRUE)` the function accepts a layer as vertic when the criteria of the reading fail or are NA AND the layer has `clay_pct` >= `min_clay` AND its designation matches a `v` master-letter modifier. Default is `FALSE`.

`vertic_properties` *Vertic properties (WRB 2022)*

Description

Tests whether any horizon shows vertic properties – shrink-swell clay behaviour evidenced by slickensides, wedge-shaped peds, and deep cracks. Diagnostic for Vertisols.

Usage

```
vertic_properties(
  pedon,
  min_clay = 30,
  min_thickness = 25,
  slickenside_levels = c("common", "many", "continuous")
)
```

Arguments

`pedon`	A `PedonRecord`.
`min_clay`	Minimum clay percent (default 30, per WRB 2022).
`min_thickness`	Minimum thickness (cm) of the vertic layer (default 25 per WRB 2022 Ch 3.2.x).
`slickenside_levels`	Vector of `slickensides` values accepted as evidence (default `c("common", "many","continuous")`).

Details

Sub-tests:

- `test_clay_above` – clay >= 30%
- `test_slickensides_present` – slickensides at or above the "common" level
- `test_minimum_thickness` – combined vertic layer thickness >= 25 cm (v0.3.1 added per WRB 2022)

v0.3.1: thickness gate added. Limitations remaining: WRB also accepts deep cracks (>= 1 cm wide extending from the surface to >= 50 cm depth, when soil is dry) and wedge-shaped peds as alternative evidence; this implementation requires clay + slickensides. The "after mixing of upper 18 cm" clause from WRB is still deferred.

Value

A `DiagnosticResult`.

References

IUSS Working Group WRB (2022), Chapter 3.2 – Vertic properties.

`vertisol` *Vertisol RSG gate (WRB 2022 Ch 4, p 101)*

Description

WRB 2022 Ch 4 (p. 101), Vertisols: other soils having

1. a vertic horizon starting \<= 100 cm from the mineral soil surface;
2. \>= 30% clay between the mineral soil surface and the vertic horizon throughout;
3. shrink-swell cracks that start (a) at the mineral soil surface, (b) at the base of a plough layer, (c) directly below a self-mulching surface (a layer with strong granular, or strong angular or subangular blocky, structure of aggregates \<= 1 cm) or (d) directly below a surface crust, and extend to the vertic horizon.

Since v0.9.311 each item is TRUE, FALSE or NA (not recorded), and the gate is NA when an item is undecided and none fails. Only a vertic horizon that can start within 100 cm decides item 1, so unrecorded layers below it do not leave the gate undecided. Item 3 reads the cracks of each layer as `vertic_horizon` does (`cracks_width_cm`, or slickensides or an `ss` designation as evidence of them): every layer between the start and the vertic horizon must have them, and a layer whose cracks are not recorded leaves the item undecided. Until v0.9.310 item 3 asked for a layer with cracks starting within 20 cm and read missing cracks as absent.

Usage

```
vertisol(pedon, strict = NULL)
```

Arguments

pedon	A PedonRecord.
strict	Logical or NULL. When NULL (default) it resolves via getOption("soilKey.rsg_strict", FALSE). TRUE applies the Tier-3 strengthened threshold.

Value

A DiagnosticResult.

Tier-3 strict mode (v0.9.98)

With strict = TRUE the overlying-clay threshold is raised from 30% to 35%, tightening the gate against marginally clayey profiles that satisfy the vertic horizon but sit close to the Vertisol cut-off.

---

vitric_properties *Vitric properties (WRB 2022 Ch 3.2.16)*

Description

Volcanic glass \>= 5% in 0.02-2 mm fraction, Al_ox + 1/2 Fe_ox \>= 0.4%, phosphate retention \>= 25%.

Usage

```
vitric_properties(
  pedon,
  min_glass_pct = 5,
  min_alfe = 0.4,
  min_p_retention = 25
)
```

Arguments

pedon	A PedonRecord.
min_glass_pct	Numeric threshold or option (see Details).
min_alfe	Numeric threshold or option (see Details).
min_p_retention	Numeric threshold or option (see Details).

Value

A DiagnosticResult recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

vlm_pick_provider *Pick the best available VLM provider*

Description

Selects a provider based on what is reachable in the user's environment, in this preference order: local Ollama (if `ollama_is_running()`), then Anthropic, OpenAI, and Google (each requires the relevant `*_API_KEY` environment variable). Errors with an actionable installation / API-key hint when no provider is reachable.

Usage

```
vlm_pick_provider(verbose = TRUE)
```

Arguments

`verbose`	If `TRUE` (default), emits a one-line `cli` message explaining the chosen provider.

Value

Character scalar: one of `"ollama"`, `"anthropic"`, `"openai"`, `"google"`.

vlm_provider *Construct a VLM provider chat object*

Description

Returns an `ellmer` chat object configured for the given provider, ready to be passed to the extraction functions (`extract_horizons_from_pdf`, etc.). The chat object wraps API credentials and model selection; it does not itself send any request.

Usage

```
vlm_provider(
  name = c("auto", "anthropic", "openai", "google", "ollama"),
  model = NULL,
  ...
)
```

Arguments

`name`	Provider name. One of `"anthropic"` (Claude), `"openai"` (GPT-4o family), `"google"` (Gemini), `"ollama"` (local).
`model`	Optional model identifier; defaults to `default_model(name)`.
`...`	Additional arguments forwarded to the corresponding `ellmer::chat_*` constructor (e.g. `system_prompt`, `api_key`, `base_url`, `params`).

Details

This is purely a convenience wrapper: it picks a default model per provider and forwards remaining arguments (e.g. `system_prompt`, `api_key`) to the underlying ellmer constructor. `ellmer` must be installed.

Value

An `ellmer` `Chat` object exposing a `$chat()` method for sending prompts.

Local-first option

Passing `name = "ollama"` runs every extraction locally via an Ollama server (default `gemma4:e4b`, Gemma 4 edge with multimodal text+image+audio support). No data leaves the machine, which is the recommended setting for sensitive field descriptions (e.g. governmental surveys, indigenous land studies) where institutional independence and data sovereignty matter. Pull the model first:

```
ollama pull gemma4:e4b      # ~3 GB edge variant (default)
ollama pull gemma4:31b      # frontier dense variant
ollama pull gemma3:27b      # earlier generation, still solid
```

Then start an Ollama server (`ollama serve`) and the chat object returned here will dispatch over HTTP locally.

Examples

```
## Not run:
# Cloud provider (needs ANTHROPIC_API_KEY)
provider <- vlm_provider("anthropic")

# Local Gemma 4 edge model -- default, ~3 GB, runs anywhere
provider <- vlm_provider("ollama")

# Local Gemma 4 frontier dense model -- best quality
provider <- vlm_provider("ollama", model = "gemma4:31b")

# Any other multimodal model the user has pulled
provider <- vlm_provider("ollama", model = "qwen2.5vl:32b")

## End(Not run)
```

---

within_tolerance *Compare a value with a threshold that has a tolerance*

Description

Some classification systems give a tolerance to some of their printed thresholds: a value beyond the threshold but within the tolerance does not fail the class, it attaches the soil to it imperfectly (the French *Referentiel pedologique* 2008 calls this an imperfect attachment). `within_tolerance()` makes one such comparison.

Usage

```
within_tolerance(value, threshold, direction = c("min", "max"), tol = 0)
```

Arguments

value	Numeric vector: the values compared (a factor or character vector is read by its labels).
threshold	Numeric: the printed threshold, recycled.
direction	"min" (the value must reach the threshold) or "max" (the value must not exceed it).
tol	Non-negative number: the tolerance, in the unit of the value, recycled. NA or NULL reads as 0 (strict).

Details

For a minimum (direction = "min"), a value at or above the threshold passes, a value below it by no more than tol is imperfect, and a value further below fails. For a maximum ("max"), a value at or below the threshold passes, a value above it by no more than tol is imperfect, and a value further above fails. The edge of the tolerance band is inclusive and read with a relative margin of 1e-9, so that 5.3 is within 0.2 of 5.5 despite the binary representation of decimals; the threshold itself is read exactly. A tolerance of 0 makes the comparison strict (only "pass" or "fail"). A missing value gives NA.

Value

A character vector as long as value: "pass", "imperfect", "fail" or NA.

Examples

```
# organic carbon 6.5 g/100 g against a minimum of 8 with a tolerance of 2
within_tolerance(6.5, 8, "min", 2)
# horizons starting at 100, 130 and 150 cm against "within 120 cm",
# tolerance 20
within_tolerance(c(100, 130, 150), 120, "max", 20)
# a threshold with no tolerance is strict
within_tolerance(39, 40, "min", 0)
```

---

wosis_citation *Citation for data read from WoSIS-latest*

Description

Returns the citation ISRIC asks users of WoSIS-latest to give, with the date the data were read. The ISRIC Data and Software Policy requires products, web services, papers and reports that use the data to reference the provider and acknowledge that the data were acquired through ISRIC.

Usage

```
wosis_citation(accessed = Sys.Date())
```

Arguments

accessed	Date the data were read (default today).

Value

A character string.

Examples

```
wosis_citation(as.Date("2026-10-06"))
```

---

wosis_profile_to_pedon

Turn one WoSIS profile into a PedonRecord, with its horizons

Description

Fetches the profile's layers live from ISRIC (read_wosis_layers_graphql) and builds a PedonRecord. Every value carries a provenance note naming its WoSIS code, unit and method. The site metadata carries the licence, the source dataset, the date the data were read and the citation ISRIC asks for (wosis_citation), so attribution survives into anything the user later exports. CC BY requires it, and an extract that drops it turns a licensed use into an unattributed one.

Usage

```
wosis_profile_to_pedon(row, fetch_layers = TRUE, timeout = 30, progress = NULL)
```

Arguments

row	One row of read_wosis_profiles_graphql.
fetch_layers	If TRUE (default) the layers are read from ISRIC; FALSE builds the site metadata only, with no network call.
timeout	Seconds to wait for ISRIC (default 30).
progress	Optional function(done, total) passed to read_wosis_layers_graphql.

Details

When the layers carry a more restrictive licence than the profile listing reported, the more restrictive one is kept.

Value

A PedonRecord. If the layers could not be read, it has no horizons and site$wosis_error says why.

wrb06_code_to_rsg *WRB 2006 RSG code -> 2022 RSG name*

Description

AfSP ships WRB 2006 RSG codes (2-letter, e.g.\ LV, AC, AR). The 2-letter codes are stable across WRB editions (2006 -> 2022); only a handful of qualifier names changed. This helper maps the codes to the WRB 2022 RSG names that `classify_wrb2022` emits.

Usage

```
wrb06_code_to_rsg(code)
```

Arguments

`code`	Character vector of WRB 2006 codes.

Value

Character vector of singular WRB 2022 RSG names; `NA` for unrecognised codes.

wrb2022_canonical *WRB 2022 canonical reference (parsed IUSS Working Group WRB 2022)*

Description

Convenience wrapper for `canonical_reference("WRB_4th_2022")`. Returns a 3-element list:

- `$rsg` (118 obs): Reference Soil Group + criteria text
- `$pq` (661 obs): principal qualifiers per RSG
- `$sq` (1167 obs): supplementary qualifiers per RSG

Usage

```
wrb2022_canonical(prefer_pkg = TRUE)
```

Arguments

`prefer_pkg`	Retained for compatibility and no longer used. Until v0.9.203 `FALSE` selected a copy bundled with soilKey; there is no bundled copy any more, so the data always come from SoilTaxonomy.

Details

Source: NCSS-tech `SoilTaxonomy` R package. Original: IUSS Working Group WRB (2022). *World Reference Base for Soil Resources*, 4th edition.

Value

The canonical WRB 2022 reference data (a list / data.frame of RSG and qualifier criteria), read from the **SoilTaxonomy** package.

`yermic_properties` *Yermic properties (WRB 2022 Ch 3.2.17) – per-pedon test wrapping* `test_yermic_surface`.

Description

Yermic properties (WRB 2022 Ch 3.2.17) – per-pedon test wrapping `test_yermic_surface`.

Usage

```
yermic_properties(pedon)
```

Arguments

pedon	A `PedonRecord`.

Value

A `DiagnosticResult` recording whether the diagnostic is present, the qualifying layers, and the supporting evidence.

Index

∗ **datasets**
GSM_DEPTHS, 167
ossl_demo_sa, 258

abrupt_textural_difference, 10, *246*
acrisol, 11, *210, 211*, *229*
aeolic_material, 12
albeluvic_glossae, 12
albic, 13
alisol, 14, *210, 211*, *229*
andic_properties, 14, *211*
andosol, *14*, 16
annotate_wrb_from_usda, 17, *49*
anthraquic, 18
anthric_horizons, 18, *212*
apply_soilgrids_depth_prior, 19, *91, 92*, *161–163*
apply_spectral_preprocessing, 20
arenic_texture, 22, *212*
argic, 23, *25, 26*, *56*, *59*, *126*, *194*, *210, 211*, *214*, *216*, *222*, *224*, *229, 230*, *235, 236*, *239*, *242*, *262*, *272*, *309*
argic_aqp, *23*, 24
argic_with_strong_clay_films, 26, *30*
argillic_clay_films_test, 26
artefacts, 27
as_aqp, 28, *159, 160*
attach_lucas_spectra, 29, *37*
audit_argic_strong_films, 30
auto_set_proj_env, 31
available_esdb_attributes, 32, *206, 207*

B_espodico, 50
B_incipiente, 51, *60*
B_latossolico, *26*, *30*, 52
B_nitico, *52*, 53
B_planico, 54, *337*
B_textural, *26*, *52*, 55
batch_robustness, 32
benchmark_afsp, 33, *188*
benchmark_bdsolos, 34, *49*, *190*
benchmark_isis, 36, *292*
benchmark_lucas_2018, *35*, 36, *49*, *198*
benchmark_natsoil, 39, *45*, *151*, *200*, *298*
benchmark_npdb, 40, *202*, *254*
benchmark_nsd, 41, *295*
benchmark_performance, 41
benchmark_redape, 43, *111*, *203*, *313*
benchmark_run_classification, 44, *49*
benchmark_sali, 45, *153*, *298*
benchmark_spectral_fill, 45, *266*, *299*
benchmark_unified, 47, *312, 313*
benchmark_wrb_vs_usda, 49
build_taxon_profiles, 50, *160, 161*

calcaric_material, 57
calcic, 57, *214*, *224*, *235, 236*, *239*, *242*
cambic, 58, *60*, *214*, *224*, *236*, *242*, *272*
cambic_aqp, *59*, 60
canonical_reference, 62
canonicalise_kst13ed_gg, 61
cerosidade, 62
chernic, 63
chernozem, 64, *227*, *233*
chernozem_strict, 64, *180*
claric_material, 65
classification_robustness, *32*, 69, *112, 113*
ClassificationResult, 66, *71*, *73*, *78*, *80, 81*, *83–85*, *87–89*, *92, 93*, *95*, *98*, *100*, *103*, *105, 106*, *109–112*, *114–116*, *123*, *180, 181*, *262*, *278*, *288*, *303*, *307*
classify_all, *35*, *48*, 70, *84*, *100, 101*, *155*, *162, 163*, *292*, *296*, *327*
classify_asc3, *39*, *45*, *71*, 72, *74, 75*, *199*, *296*
classify_asc3_family, *73, 74*, 74
classify_by_spectral_neighbours, 75, *134*, *298*

classify_cpcs1967, 77, *310*
classify_cscj2011, 79
classify_cssc3, *40*, 80
classify_cst2001, 82, *219*
classify_csv, 83, *101*, *296*
classify_fao, *67*, 84, *86–88*, *144, 145*, *222*
classify_fao1974, *85, 86*, 86
classify_fao1988, *86*, 87
classify_from_documents, 88, *307, 308*
classify_from_photos, *20*, 90, *123*
classify_hscs2024, 92
classify_ka5, 94
classify_kidpr2008, 96
classify_nzsc4, *41*, 98, *294, 295*
classify_profiles, 100
classify_rp2008, *67*, 102, *173, 174*, *310, 311*
classify_scsj2017, *79, 80*, 104
classify_sibcs, *30*, *34*, *36*, *43*, *70–72*, *76*, *89*, 105, *111, 112*, *120*, *156*, *162*, *315*, *328*
classify_sibcs_familia, *106*, 106, *109*
classify_usda, *34*, *70–73*, *76, 77*, *79*, *81, 82*, *85*, *87*, *89*, *93, 94*, *96*, *99*, *103, 104*, 108, *110*, *112*, *115*, *162*, *304*, *328*
classify_usda_family, *108*, 109
classify_via_smartsolos_api, 110, *120*
classify_with_engine_heuristic, 112
classify_with_uncertainty, *69*, 112, *163*
classify_wrb2022, *34*, *36*, *49*, *66*, *70–72*, *76*, *89, 90*, *112*, 114, *162*, *278*, *304*, *328*, *331*
classify_za1991, 115
clear_kst13_cache, 117, *182*
clear_ossl_cache, 117
cohesic, 118
combine_priors, 119
compare_engines, 119
compare_smartsolos, *111*, 120
compute_ki, 121
compute_kr, 121
compute_per_attribute_evidence_grade, *92*, *113*, 122
continuous_rock, 123
coverage_report, 123
cryic_conditions, 124, *125, 126*, *217*
cryic_horizon, 125

DiagnosticResult, *11–15*, *17–19*, *22*, *24–28*, *51–55*, *57–60*, *63–65*, *118*, *123*, *125, 126*, 126, *128*, *136–138*, *147–149*, *154*, *156, 157*, *159*, *164*, *166*, *168*, *170, 171*, *173–177*, *179, 180*, *184–186*, *210*, *244*, *246–250*, *257, 258*, *261, 262*, *269–271*, *273*, *275–277*, *287*, *289, 290*, *303*, *310*, *322–325*, *329–331*, *335–343*, *345*, *347, 348*, *353–355*, *361*
distrofico, 128
dolomitic_material, 128
download_bdsolos, 129, *190*
download_extdata_cache, 130
download_ossl_subset, *117*, 131, *133, 134*
download_ossl_subset_with_labels, 133
download_redape_dataset, 135, *203*
duric_horizon, *136*, 136, *220*
duripa, 136

ensure_horizon_schema, 137, *172*, *348*
eutrofico, *128*, 138
evaluate_rsg_tests, 138, *317*
extract_horizons_from_pdf, *89, 90*, 139, *356*
extract_munsell_from_photo, *89*, *92*, 140
extract_site_from_fieldsheet, *89*, 141

familia_mineralogia_argila_geral, 142
FamilyAttribute, *75*, *106, 107*, *109*, *143*, 143
fao_crosswalk, 144
febr_index_munsell, 145, *291, 292*
ferralic, *59*, *126*, 146, *214*, *216*, *222*, *224*, *230*, *235, 236*, *239*, *242*, *261, 262*
ferralsol, *114*, 148
ferric, 149
fetch_aafc_normals, 149, *255*
fetch_natsoil, 150, *199, 200*
fetch_sali, 152, *296–298*, *322*
fibrico, 153
fill_from_spectra, *45–47*, *132*, 154, *241*, *258*, *280*, *298, 299*, *345*
fill_munsell_from_spectra, 155, *282*
fluvic_material, 156, *223*
folic, 157
format_wrb_name, 158
fragic, 158, *159*
fragipa, 159
from_aqp, *28*, 159

gapfill_by_predicted_taxon, *50*, 160

gapfill_derive_horizon, 161
gapfill_within_pedon, *48*, *71*, *106*, *108*, *115*, *161*, 162
get_perturbation_scale, *112, 113*, 163
gleyic_properties, 164, *165*, *224*, *240*, *336*
gleysol, 165
grapes-or-or-grapes, 166
GSM_DEPTHS, 167, *168–170*
gypsic, 167, *214*, *224*, *235, 236*, *239*, *242*
gypsiric_material, 168

harmonize_to_gsm, *48, 49*, 168
hemico, 170
histic_horizon, *157*, 170, *225*
horizon_attribute_spec, *100*, 171, *262*, *326*, *348*
horizon_column_spec, *171, 172*, 172, *262*, *267*, *348*
hortic, 172
humus_form_rp, *67*, 173
hydragric, 174
hypersulfidic_material, *175*, 175
hyposulfidic_material, 175

identify_horizons, *82, 83*, *95, 96*, *98*, *116, 117*
inspect_bdsolos_csv, *130*, 176, *189, 190*
irragric, 177
isis_citation, 178, *292*
isis_profile_to_pedon, *81*, 178, *292*

kastanozem, 179, *227*, *233*
kastanozem_strict, 180
key_trace_table, 180
kidpr2008_types, *97*
kst13_canonical, *62*, 181, *182, 183*
kst13_codes, *123, 124*, 182, *183*
kst13_criteria, *182*, 183

leptic_features, 183, *229*
limnic_material, 184
limonic, 185
lithic_discontinuity, 186
lixisol, 186, *210, 211*, *229*
load_afsp_pedons, *33*, 187, *188*
load_afsp_sample, *33*, 188
load_bdsolos_csv, *34, 35*, *130*, *176*, 189, *292*
load_embrapa_pedons, 190
load_febr_pedons, 191
load_kssl_nasis_sample, *188*, 192
load_kssl_pedons, 193
load_kssl_pedons_gpkg, *17*, *49*, 194, *195*
load_kssl_pedons_with_nasis, *192*, 194
load_kssl_sample, *192*, 195
load_lucas_pedons, 196
load_lucas_soil_2018, *29, 30*, *37, 38*, 197
load_natsoil_pedons, *39*, *150, 151*, 199, *297, 298*
load_npdb_pedons, *40*, 201, *254*
load_ossl_models, *324*
load_ossl_models *(*save_ossl_models*)*, 324
load_redape_pedons, *43*, 202
load_rules, 203, *315–318*, *327*
load_wosis_sample, 204
load_wosis_stratified_sample, *188*, 205
lookup_esdb, *32*, *36*, *38*, *198*, 206, *207–209*
lookup_mapbiomas_solos, 207, *209*
lookup_soilgrids, *38*, *198*, *208*, 208
luvisol, *210*, 210, *211*, *229*

make_acrisol_canonical, 210
make_alisol_canonical, 211
make_andosol_canonical, 211
make_anthrosol_canonical, 212
make_arenosol_canonical, 212
make_argissolo_canonical, 213
make_asc3_example, *74*, 213
make_calcisol_canonical, 214
make_cambisol_canonical, 214
make_cambissolo_canonical, 215
make_chernossolo_canonical, 215
make_chernozem_canonical, 216
make_cpcs1967_example, *78*, 216
make_cryosol_canonical, 217
make_cscj2011_example, *80*, 217
make_cssc3_example, *81*, 218
make_cst2001_example, *83*, 219
make_durisol_canonical, 220
make_empty_horizons, 220
make_espodossolo_canonical, 221
make_fao_example, *86–88*, 221
make_ferralsol_canonical, 222
make_fluvisol_canonical, 223
make_gleissolo_canonical, 223
make_gleysol_canonical, 224
make_gypsisol_canonical, 224
make_histosol_canonical, 225
make_hscs2024_example, *93*, 225

make_ka5_example, *95*, 226
make_kastanozem_canonical, 227
make_kidpr2008_example, *98*, 228
make_latossolo_canonical, 228
make_leptosol_canonical, 229
make_lixisol_canonical, 229
make_luvisol_canonical, 230
make_luvissolo_canonical, 230
make_neossolo_canonical, 231
make_nitisol_canonical, 231
make_nitossolo_canonical, 232
make_nzsc4_example, *100*, 232
make_organossolo_canonical, 233
make_phaeozem_canonical, 233
make_planosol_canonical, 234
make_planossolo_canonical, 234
make_plinthosol_canonical, 235
make_plintossolo_canonical, 235
make_podzol_canonical, 236
make_retisol_canonical, 236
make_rp2008_example, *104*, 237
make_scsj2017_example, *105*, *217*, 239
make_solonchak_canonical, 239
make_solonetz_canonical, 240
make_stagnosol_canonical, 240
make_synthetic_pedon_with_spectra, 241
make_technosol_canonical, 241
make_umbrisol_canonical, 242
make_vertisol_canonical, 242
make_vertissolo_canonical, 243
make_za1991_example, *117*, 243
match_horizon_sequence, *94*
mineral_material, 244
mollic, *126*, *166*, *214*, *216*, *222*, *224*, *227*, *230*, *233*, *235, 236*, *239*, *242*, 245, *262*
mudanca_textural_abrupta, *55*, 246
mulmic_material, 247

natric_horizon, *240*, 247, *309*
nitic_horizon, *231*, 248
nitisol, *249*, 250
normalise_febr_sibcs, *34, 35*, 250
normalise_febr_usda, *35*, *251*, 251
normalise_febr_wrb, *35*, *251*, 252
normalise_kssl_subgroup, 252
npdb_dry_colour_model, 253, *254*
npdb_scenario, *40*, *150*, *253*, 254
npdb_soil_climate, *150*, *254*, 255
nsd_citation, 256, *294*

ollama_is_running, 256
organic_material, 257
organotechnic_material, 257
ornithogenic_material, 258
ossl_demo_sa, 258
ossl_library_template, 260, *345*
oxic_usda, *147*, 261

panpaic, 261
pedon_json_schema, 267, *351*
pedon_to_spc, 268
PedonRecord, *11–20*, *22, 23*, *25–30*, *32–34*, *36, 37*, *39–41*, *43–45*, *49, 50*, *52–60*, *63–65*, *69–74*, *77*, *79–82*, *84, 85*, *87–89*, *91–94*, *96*, *98–100*, *102–104*, *106–110*, *112–115*, *118*, *120*, *122, 123*, *125, 126*, *128*, *136–139*, *141–143*, *147–149*, *153–157*, *159–162*, *164, 165*, *167–180*, *184–191*, *193–195*, *197–203*, *210–236*, *238–247*, *249, 250*, *253, 254*, *257, 258*, *261, 262*, 262, *265, 266*, *268–273*, *275–277*, *279, 280*, *282*, *287*, *289–296*, *298*, *302, 303*, *307*, *309*, *315–325*, *327*, *329–334*, *336–343*, *345–348*, *350–353*, *355*, *359*, *361*
pedons_from_spectral_table, 265, *299*
petrocalcic, 268
petroduric, 269
petrogypsic, 270
petroplinthic, 270
phaeozem, *227*, *233*, 271
pi_to_confidence, *155*, 274
pick_engine, 271, *272*
pick_engine_batch, 272
pisoplinthic, 273
plaggic, 274
planic_features, *234*, 275
planosol, 276
plinthic, *224*, *235, 236*, *242*, *248*, 277
posterior_classify, 278
predict.soilKey_pls_model, 279
predict_from_spectra, *30*, *37, 38*, *198*, 279, *346*
predict_lab_from_spectra, 280, *285*

predict_munsell_from_spectra, *30*, *38*, *155*, *198*, 281, *285*
predict_ossl_mbl, *76*, *131*, *132*, *155*, *258*, *260*, 282, *283*, *284*
predict_ossl_plsr_local, *131*, *155*, *260*, 283
predict_ossl_pretrained, *155*, 284, *345*
predict_xyz_from_spectra, *280*, *281*, 285
preprocess_spectra, *21*, *76*, *155*, 285, *346*
pretic, 287
print.soilKey_pls_model, 287
prior_consistency_check, *114*, 288
protocalcic_properties, *64*, 288
protogypsic_properties, 289
protovertic, 290

read_febr_pedons, *146*, 291
read_isis, *36*, *178*, 292
read_nsd_pedons, *41*, *256*, 293
read_pedon_csv, *83*, *84*, 295, *348*
read_sali_pedons, *45*, *152*, *153*, 296, *322*
read_spectral_library, *47*, *266*, 298
read_wosis_layers_graphql, 300, *359*
read_wosis_profiles_graphql, *300*, 301, *359*
reducing_conditions, *165*, 303
report, *68*, *89*, *90*, 303, *305*, *306*, *308*
report_html, *68*, 305, *306*, *307*
report_pdf, *68*, 306
report_to_qgis, 307
resolve_wrb_qualifiers, *158*, 308
retic_properties, *236*, 309
rp2008_crosswalk, 310, *312*
rp2008_from_label, *311*, 311
run_all_benchmarks, 312
run_classify_app, 313
run_demo, 314
run_sibcs_grande_grupo, *105*, 315
run_sibcs_key, 316
run_sibcs_subgrupo, 316
run_sibcs_subordem, *105*, 317
run_taxa_list, *315–317*, 317, *318*
run_taxonomic_key, *317*, *318*, 318
run_usda_great_group, 319
run_usda_key, 319
run_usda_subgroup, 320
run_usda_suborder, 320
run_wrb_key, 321
runApp, *314*

sali_citation, *153*, *298*, 322
salic, *214*, *224*, *235*, *236*, *239*, *242*, 321
saprico, 323
save_ossl_models, 324
shrink_swell_cracks, 324
sideralic_properties, 325
site_attribute_spec, *100*, *262*, 325, *348*
soil_classes_at_location, 327
soil_organic_carbon, 329
soilgrids_usda_lut, 326
soilgrids_wrb_lut, 326, *333*, *334*
soilkey_systems, *66*, *71*, *74*, *78*, *80*, *81*, *83*, *86*, *93*, *95*, *98*, *100*, *104*, *105*, *117*, *203*, *304*, 327
solimovic_material, 330
sombric, 330
spatial_prior, *114*, *278*, *288*, 331, *334*
spatial_prior_embrapa, *331*, 332
spatial_prior_soilgrids, *20*, *326*, *329*, *331*, 333
spodic, *50*, *224*, *235*, *236*, *242*, 334
st_features_canonical, *62*, 336
stagnic_properties, *240*, *276*, 335
subgrupo_planossolo_espessos, 337
subgrupo_planossolo_mesicos, 338
subgrupo_plintossolo_endico_concrecionario, 338
subgrupo_plintossolo_endico_litoplintico, 339
subgrupo_plintossolo_espessos, 340

takyric_properties, 340
technic_features, *241*, 341
technic_hard_material, 342
tephric_material, 342
terric, 343
texture_class_from_pct, *268*, 343
thionic, 344
train_pls_from_ossl, *37*, *279*, *280*, *324*, 345
tsitelic, 346

umbric_horizon, *166*, *242*, 347
unrecognised_attributes, *101*, *262*, *296*, *326*, 348
usda_to_wrb_rsg, *17*, *192*, *195*, 349

validate_horizon_geometry, 350
validate_pedon_json, 351

vertic_horizon, *194*, *290*, 351, *354*
vertic_properties, *242*, 353
vertisol, 354
vitric_properties, *14*, 355
vlm_pick_provider, *256*, 356
vlm_provider, *88–90*, *139*, *141, 142*, 356

within_tolerance, *104*, 357
wosis_citation, *301, 302*, 358, *359*
wosis_profile_to_pedon, *301, 302*, 359
wrb06_code_to_rsg, 360
wrb2022_canonical, *62*, *123, 124*, 360

yermic_properties, 361