

Package ‘GeoVersa’

July 6, 2026

Title Design-Based Residual-Correction Forests for Digital Soil Mapping

Version 0.3.0

Description Implements DB-TARF (Design-Based Targeted Adaptive Residual Forest) for large-scale digital soil and ecological mapping evaluated under the design-based paradigm of Wadoux et al. (2021) [<doi:10.1016/j.ecolmodel.2021.109692>](https://doi.org/10.1016/j.ecolmodel.2021.109692). A random forest is augmented by a cross-fitted, out-of-fold-selected residual correction (residual forests, ordinary kriging, recalibration), together with design-based conformal prediction intervals.

License MIT + file LICENSE

Encoding UTF-8

Imports ranger, caret, stats, withr

Suggests Cubist, gstat, sp, nnet, testthat (>= 3.0.0)

RoxygenNote 7.3.3

Config/testthat/edition 3

NeedsCompilation no

Author Hugo Rodrigues [aut, cre]

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

Repository CRAN

Date/Publication 2026-07-06 12:30:14 UTC

Contents

dbtarf_train_predict	2
Index	4

dbtarf_train_predict *Fit DB-TARF and predict a test set*

Description

DB-TARF (Design-Based Targeted Adaptive Residual Forest) fits a random-forest base learner on `train_df` and adds a cross-fitted, out-of-fold-selected residual correction, then predicts `test_df`. A correction is adopted only when it lowers the out-of-fold RMSE of the calibration sample.

Usage

```
dbtarf_train_predict(
  train_df,
  test_df,
  response_name,
  predictor_names = NULL,
  coord_names = "auto",
  ...
)
```

Arguments

<code>train_df</code>	A data frame with the response, predictors and (optionally) coordinate columns.
<code>test_df</code>	A data frame with the same predictor (and coordinate) columns as <code>train_df</code> .
<code>response_name</code>	Character; name of the response column in <code>train_df</code> .
<code>predictor_names</code>	Character vector of predictor column names, or NULL (default) to infer them from <code>train_df</code> .
<code>coord_names</code>	Length-2 character vector of coordinate column names, or "auto" (default) to detect them.
<code>...</code>	Further arguments passed to the internal training routine (e.g. <code>n_folds</code> , <code>lambda_grid</code> , <code>rf_tune</code> , <code>train_seed</code>), and the top-k profile-ensemble controls <code>ensemble_top_k</code> (integer ≥ 1 , default 1: the single best profile, i.e. no blending), <code>ensemble_weighting</code> ("softmax", "inverse" or "uniform") and <code>ensemble_temperature</code> (positive, default 0.75); see <code>dbtarf_default_params</code> . The ensemble blends the top-k RF profiles (ranked by out-of-fold RMSE) and only activates when <code>fair_profile_search</code> is TRUE, (<code>rf_tune</code> or <code>resid_rf_tune</code>) is TRUE and <code>ensemble_top_k</code> > 1.

Value

A list with the test predictions (`pred_test`), the base-RF predictions (`pred_test_base`), conformal prediction-interval half-widths (`pi_q90`, `pi_q95`), per-run diagnostics and the `candidate_table`. When `ensemble_top_k` > 1 the diagnostics additionally record `ensemble_applied`, `ensemble_size`, `ensemble_weighting`, `ensemble_temperature`, `ensemble_profiles`, `ensemble_profile_oof_rmse` and `ensemble_weights`, and `candidate_table` gains `ensemble_member`, `ensemble_rank` and `ensemble_weight` columns. Note that when the ensemble is applied the conformal half-widths

(pi_q90, pi_q95 and the _w/_sp variants) are inherited from the single best (top-ranked OOF) profile and are *not* recalibrated against the blended pred_test; the conformal coverage guarantee therefore pertains to the best single profile, not to the blended point estimate (diagnostics\$ensemble_pi_from_best flags this).

Examples

```
set.seed(1)
n <- 120
tr <- data.frame(y = rnorm(n), a = rnorm(n), b = rnorm(n))
te <- tr[1:15, ]
out <- dbtarf_train_predict(tr, te, "y", c("a", "b"),
                           coord_names = NULL, rf_tune = FALSE,
                           fair_profile_search = FALSE)
head(out$pred_test)
```

Index

dbtarf_train_predict, [2](#)