

Spatial model diagnostics - overview of code and data

Alexander Brenning

13 November 2021

1 Files

All code is written in R version 4.1.0. CRAN or Github versions of required packages can be installed using the install script.

The following files are included in this data and code publication, using the following file naming conventions:

- `casestudy` is the name of the case study, i.e., `meuse` or `maipo`;
- `model` represents the model, for example `rf` for random forest, or `splda` for NN-LDA.
- `est` represents the estimation technique, such as `loo` for (non-spatial) LOO-CV, or `spcv` for spatial CV using k -means clustering of sample coordinates.
- `.rda` and `.Rdata` files were saved using `save()`, while `.rds` files used `saveRDS()`.

Code files:

- `spdiag_install.R`: Install required R packages from CRAN or Github, as appropriate.
- `spdiag_casestudy.R`: Prepare data and other objects defining, for example, model settings. Result is saved in file `run_casestudy.rda`, in the Maipo case study also in `wrp_maipo.rda`.

Data and results files:

- The raw data of the Meuse and Maipo case studies is included in the `sp` package as `data(meuse)` and in the `sperrorest` package as `data(maipo)`, respectively.
- `run_casestudy.rda`: Preprocessed data and other object defining, for example, model settings.
- `wrp_maipo.rda`: Preprocessed data and transformation objects for model interpretation from a transformed perspective in the Maipo case study.
- `casestudy_res_model.rda`: Results of spatial LOO estimation, to be used to compute SPEPs and SVIPs.
- `casestudy_res_est.rda`: Results of various performance estimation procedures such as spatial CV (`est: spcv`) or resubstitution (`train`).
- `casestudy_smry.rds`: Summary object that contains data and (processed) results from the case studies for use in the RMarkdown manuscript file.

2 Preparing the data and computing environment

The code has been executed with R 4.1.0 and requires several additional R packages.

1. Run `spdiag_install.R`.
2. Run `spdiag_casestudy.R`. You can skip this and use the pre-computed files instead.

3 Running the computations

The following steps perform all computations made for this paper. Skip this section to use pre-computed results.

3. Run each of the `run_casestudy_model.R` scripts.
 - This is the computationally most intensive part, which requires a high-performance computing environment since the code has not been optimized for performance. Note that memory requirements are high in parallel processing. Currently not all scripts make use of parallelization.
4. Run each of the `run_casestudy_est.R` scripts.
 - This is computationally intensive especially in the Maipo case study and in the hybrid OK-RF model. Currently not all scripts make use of parallelization.

4 Summarizing and plotting the results

5. Run `spdiag_casestudy_smry.R`.
6. Run `spdiag_casestudy_plot.R` step by step to generate the figures from the manuscript.