

Research Data Archive for NERC funded project

Detecting soil degradation and restoration through a novel coupled sensor and machine learning framework

Project name: Signals in the Soil: Detecting soil degradation and restoration through a novel coupled sensor and machine learning framework

Grant reference: NE/T012307/1

Project manager: Professor John Quinton

Project data manager: Dr. Mengyi Gong

1 Data description

The archived soil modelling data consist of three parts,

- (1) computer code in the form of R scripts for implementing the changepoint-based method to soil moisture time series from soil sensors and analysing the model output,
- (2) soil water and precipitation time series data from the National Ecological Observatory Network (NEON) open-sourced data portal (<https://data.neonscience.org/>) after pre-processing (i.e., down-sampling and imputing missing values) saved in “rds” format,
- (3) model output data from statistical analysis saved in “rds” format.

All R scripts are stored in the root folder. Specifically,

- (a) R script `PELT_soil_function_nlm.R` contains essential functions for implementing the method developed for this project. The details of the method can be found in this manuscript via the link <https://arxiv.org/abs/2310.17546>.
- (b) R script `PELT_soil_app_run.R` contains the code for implementing the changepoint method on the soil moisture time series from the top layer from nine selected field sites from NEON data portal, and `PELT_soil_app_parameters.R` contains the code for the follow up statistical analysis.
- (c) R script `PELT_soil_app_3dep_run.R` contains the code for implementing the changepoint method on the soil moisture time series from top three layers from three selected field sites, and `PELT_soil_app_delay.R` contains the code for analysing the delays in the soil moisture peaks down the soil profile.

- (d) R scripts `PELT_soil_penalty_learning.R` provides an example of using precipitation to assist penalty selection, which is an essential part of the changpoint-based method.

All data files are stored in folder “Application_data”. The files with names containing the word “PELT” store soil moisture time series from the top layer from nine selected sites. The files with names containing the word “Dep123” store the soil moisture time series from the top three layers from three selected sites. The file `Rain_SM_data_SRER.rds` contains soil moisture time series and precipitation time series to be used in the penalty selection example.

All output files are stored in folder “Applicaton_result”. The files `PELT_OSBS_3dep_output.rds`, `PELT_SRER_3dep_output.rds`, `PELT_TALL_3dep_output.rds` and `PELT_9sites_output_model.rds` contain the model output. The files `PELT_9sites_output_par.rds`, `PELT_9sites_output_se.rds` and `PELT_3sites_delayed_peaks.rds` contain the derived output.