Variography Sequence

- The purpose of the variography sequence is to produce an empirical variogram and variogram model from data produced by the vertex sequence.
- To learn about variography, please consult a reference such as Geostatistics for Environmental Scientists, Webster and Oliver (ISBN: 978-0470028582), chapters 4-5.

Required Packages

- spdep
- ggplot2

Pre-variography formatting

To make the process of generating various variograms in batches regular, rename and merge final vertex tables for each variable of interest. Column names for these variables should be in the form:

[FTP | PIB].[PVC | NC].[SUVR | DVR].[TP*N*]

TP*N* refers to timepoint – TP0, TP1, TP2, etc.
Examples: FTP.NC.SUVR.TP1, PIB.PVC.DVR.TP0, etc.

This process is not performed by these scripts and should be done by the user during or after the vertex sequence and before the variography sequence.

Variography_1_Detrend.R – detrends **VARIABLE** if a trend surface can be fit to the x, y patch coordinates, and z (as value of **VARIABLE**, not location) data. The script will also output a summary containing the details of trend surface regression for all of the input vertex tables.

Arguments

1. **AGGREGATES** – path to folder containing final aggregated vertex tables (as produced in vertex sequence)
2. **DETREND** – path to folder to export regression summary table
3. **VARIABLE** – variable on which to perform detrending. This will include any variable that contains **VARIABLE** in its column name e.g. an argument of “FTP” will detrend the column “FTP.PVC” as well as “FTP.NC”. To avoid this behavior, express column name more explicitly.

Objectives

a. Read in final (remapped) aggregated vertex tables from **AGGREGATES**
b. Detrend data for each column with name that includes **VARIABLE**
c. Add detrended data columns to vertex table
d. Overwrite vertex tables in **AGGREGATES** with detrended data
e. Save trend surface regression results for all input data to table (.rds) in **DETREND**

Notes

tobias.estime@gmail.com
September 2019

- Uses an excessively large data frame to store summary results for each related variable (see script for details). In some cases, this data frame may not be large enough to store all results (adjust as needed).
- This script overwrites files in **INPUT**
- _d is appended to the column names of all variables *regardless of whether or not variable was detrended*. Refer to summary table for detrend information. This notation is followed throughout the variography sequence.

Usage

```
Rscript Variography_1_Detrend.R vertex_directory/final_tables
../variography/variography_directory FTP.PVC.SUVR
```

Variography_2_Empirical.R – computes empirical (semi)variogram (either auto-variogram or cross-variogram) for the given vertex table and variable(s).

Arguments

1. **AGGREGATES** – path to folder containing final aggregated vertex tables
2. **DISTANCES** – path to folder containing distance lists (produced by Vertex_5)
3. **BINS** – path to folder containing bins, which are the pairs of vertices that contribute to each point on the variogram.
4. **EMPIRICALS** – path to folder to output empirical variograms
5. **VARIABLE** – variable to calculate variogram
6. **BIVARIABLE** – variable to calculate cross-variogram (repeat **VARIABLE** argument if producing an auto-variogram *or* cross-time variograms)
7. **STANDARD** – 0 (false) or 1 (true) to specify whether using standard space data structures
8. **AUTO** – 0 (false) or 1 (true) to specify whether to produce auto-variograms only. If **VARIABLE** and **BIVARIABLE** are the same and **AUTO** is set to 0, script will produce cross-time variograms in addition to auto-variograms.
9. **POINTS** – integer; approximate number of points to produce for each variogram; highly dependent on size of vertex tables; recommended value: 25 or greater.

Objectives

a. Read in vertex tables, distance lists
b. Create bins for specified number of **POINTS** if a bin file (.rds) is not available
c. Iterate over matching **VARIABLE** and **BIVARIABLE**
d. Calculate empirical variogram for both detrended (_d) and original data
e. Plot empirical variogram for both detrended and original data

Notes

- Operations are performed on both hemispheres concurrently.
- Binning is a time-consuming process. If bins already exist, they will not be reproduced. To prevent repeated binning, keep **POINTS** argument consistent.
- Variography sequence does not include cross-subject “hybrid” variograms.

Usage

```
## Native space auto-variograms, for example
```

tobias.estime@gmail.com

```
Rscript Variography_2_Empirical.R ../vertex_sequence/vertex_directory/final_tables
../vertex_sequence/vertex_directory/distance_lists variography_directory/bins
variography_directory/empiricals FTP.PVC.SUVR FTP.PVC.SUVR 0 1 25

## Or native space cross-time variograms

Rscript Variography_2_Empirical.R ../vertex_sequence/vertex_directory/final_tables
../vertex_sequence/vertex_directory/distance_lists variography_directory/bins
variography_directory/empiricals FTP.PVC.SUVR FTP.PVC.SUVR 0 0 25

## Or native space cross-variograms

Rscript Variography_2_Empirical.R ../vertex_sequence/vertex_directory/final_tables
../vertex_sequence/vertex_directory/distance_lists variography_directory/bins
variography_directory/empiricals FTP.PVC.SUVR PIB.PVC.DVR 0 0 25
```

Variography_3_Model.R – models empirical variograms. The “stable” variogram model is the only valid model used for fitting. A separate “pure nugget” (invalid noise) model is also fit for comparison to valid models. This script does not select a “best” model, but generates possible models, which can then be evaluated by some criteria in **Variography_4_Model_Summary.R**.

Arguments

1. **EMPIRICALS** – path to folder containing empirical variograms
2. **MODELS** – path to folder to output model data structures (.rds)
3. **HEMISPHERES** – list of hemispheres (as additional arguments) to model

Objectives

a. Read in empirical variogram
b. Model various lengths of the empirical variogram with either nugget or stable variogram models.
c. Export model (.rds) files, which contain multiple possible models for each input variogram

Notes

- The process of batch modeling is not always successful. It may be possible to improve outcomes by adjusting input parameters.
- The sill parameter is a partial sill.
- The entire variogram is never used to fit the model.
- If a model cannot be fit and results in an error, the result will be stored as a default nugget model.
- Multiple models are contained within each exported model file. These are fit to different fractions of the variogram.

Usage

```
Rscript Variography_3_Model.R variography_directory/empiricals
variography_directory/models lh_d rh_d
```

---

tobias.estime@gmail.com

Variography_4_Model_Summary.R – selects one model from the series of models output in previous script for each empirical variogram by iterating through different models and selecting based on some criteria. Stores parameters for each variogram in file.

Arguments

1. **EMPIRICALS** – path to folder containing empirical variograms
2. **MODELS** – path to folder containing model (.rds) files
3. **HEMISPHERES** – list of hemispheres (as additional arguments)

Objectives

a. Read in empirical variogram and models file
b. Compare models for each empirical variogram
c. Select model based on some criteria
d. Store parameters for each model in summary file

Notes

- The output of this script is used in the mapping sequence.

Usage

```
Rscript Variography_4_Model_Summary.R variography_directory/empiricals
variography_directory/models lh_d rh_d
```

Variography_5_Model_Plot.R – plot variogram models along with empirical data

Arguments

1. **SUMMARY** – path to file containing model summary produced by previous script
2. **EMPIRICALS** – path to folder containing empirical variograms
3. **PLOTS** – path to folder to export plots (.png)
4. **HEMISPHERES** – list of hemispheres (as additional arguments)

Objectives

a. Read in models summary and empirical variograms
b. Export plots of models (solid line) against empirical data (points)

Notes

- Dotted lines are plotted at range and sill

Usage

```
Rscript Variography_5_Model_Plot.R variography_directory/models/summary_models.rds
variography_directory/empiricals variography_directory/models/plots lh_d rh_d
```

---

tobias.estime@gmail.com
September 2019