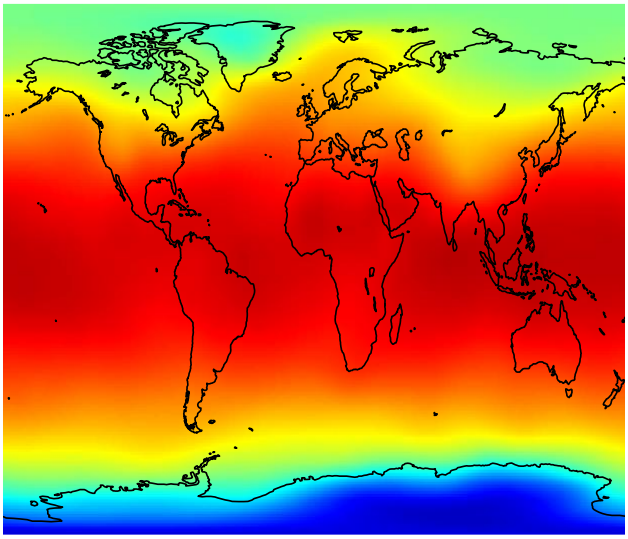
Global-Krigger

A global kriging interpolation toolbox

v1.0.0

1 May 2023



Contents

Introduction

Many applications in Earth sciences require spatial interpolation from a set of irregularly distributed observations on curved surfaces such as the surface of the Earth. These include model-data comparisons and derivation of continuous scalar fields as input for Earth system models. Kriging is a powerful interpolation technique, mainly because it provides explicit quantification of the uncertainty associated to predictions, due both to data density and measurement error. However, the theory of kriging for curved domains where the distance metric is non-Euclidean, such as the surface of the Earth, is not very well developed. Thus, the validity of the predictions of existing non-Euclidean kriging applications cannot be guaranteed.

Global-Krigger is a new kriging interpolation algorithm that guarantees the validity of predicted fields, and derives a combined uncertainty field due both to spatial variations in data density and measurement error (when defined). It includes a graphical user interface that is user-friendly, guiding users through a range of choices during data pre-interpolation analysis, kriging, and post-processing.

We encourage users to contact us when finding bugs, or for suggestions, to njc58@cornell.edu. If you use Global-Krigger for your research, please cite the study that introduced the software (this is a reference for the accepted version, please check online for the definitive reference):

Cosentino N. J., Opazo N. E., Lambert F., Osses A., van 't Wout E. (in press) Global-Krigger: A global kriging interpolation toolbox with paleoclimatology examples. Accepted for publication in *Geochemistry, Geophysics, Geosystems.*

Version history

v1.0.0 - 1 May 2023
The first complete version of Global-Krigger, after peer-review.

System requirements

Global-Krigger was written on MATLAB R2022a, and has been thoroughly tested in that version of MATLAB. To be run from the Command Window without the GUI, it requires the Curve

Fitting Toolbox, Mapping Toolbox, Signal Processing Toolbox and Statistics and Machine Learning Toolbox. To run the GUI, the installed version of MATLAB needs to be R2016b or more recent.

A standalone version of Global-Krigger's GUI exists that does not require MATLAB, for the Windows, macOS and Linux platforms. The system requirements for this installation are the same as for MATLAB R2022a, and can be checked here.

Using Global-Krigger

From MATLAB's Command Window

To run Global-Krigger from MATLAB without the GUI, download all M-files (*.m) included in this distribution to your computer. Open MATLAB and set the directory path to the folder where you downloaded the files.

You may run Global-Krigger by executing the *main.m* file in MATLAB, either by entering *main* in the Command Window and then hitting the enter key, or by pressing the *Run* button (green play symbol) in MATLAB's Editor tab. After starting the *main.m* routine, you will be prompted to enter required input for Global-Krigger.

The dataset you will work with also needs to be in the same directory. Only files with extensions .txt, .csv, .mat, .xlsx and .xls are permitted. Three columns must be included: the first for longitude, the second for latitude, and the third for a scalar variable. A fourth column for the variable 1-sigma absolute uncertainty is optional. For .mat files, the file must include only one array and should not include headers. For .txt, .csv, .xls and .xlsx files, as many non-numeric headers as desired are permitted. Only the first sheet in multiple-sheet spreadsheet files will be read.

Launching the GUI from MATLAB

The main window

To launch Global-Krigger's GUI from MATLAB, first open MATLAB, then install the GUI by double-clicking the *GlobalKrigger_v1.0.0_MATLAB_GUI.mlappinstall* executable included in this distribution. Once installed, this executable may be discarded. Then, click on the Apps tab in MATLAB and select the Global-Krigger icon.

Figure 1 shows the first window that pops-up when Global-Krigger's GUI is launched. At first, the plots on the right half of the screen will not show. Only when a dataset is imported using the

Import Data button (1) will the histogram and data point map be shown. Use the *Reset* button (2) at any point during your session to undo applied changes to the dataset.

The *Data Visualization* tab (3, Figure 1) displays the distribution of the dataset's variable on a histogram, as well as a map with the location of the dataset's points. The *Dataset* tab (4, Figure 1) shows a table with the data, which is updated after any change is produced to the original dataset. Finally, the *Kriging Visualization* tab (5, Figure 1) will display the results of the kriging interpolation once the process of kriging is complete (this tab will remain blank until the process is complete).

The *Data Analysis* tab (6, Figure 1) includes two options: treatment of outliers and data transformation. The former allows eliminating the points with the lowest/highest 2% or 5% of the values. The latter allows transforming the data. Either step (or both) is recommended when the dataset is too skewed to perform kriging, when doing so reduces the skewness ratio of the dataset. This value, together with the histogram and map plots are modified when these modifications to the datasets are applied.

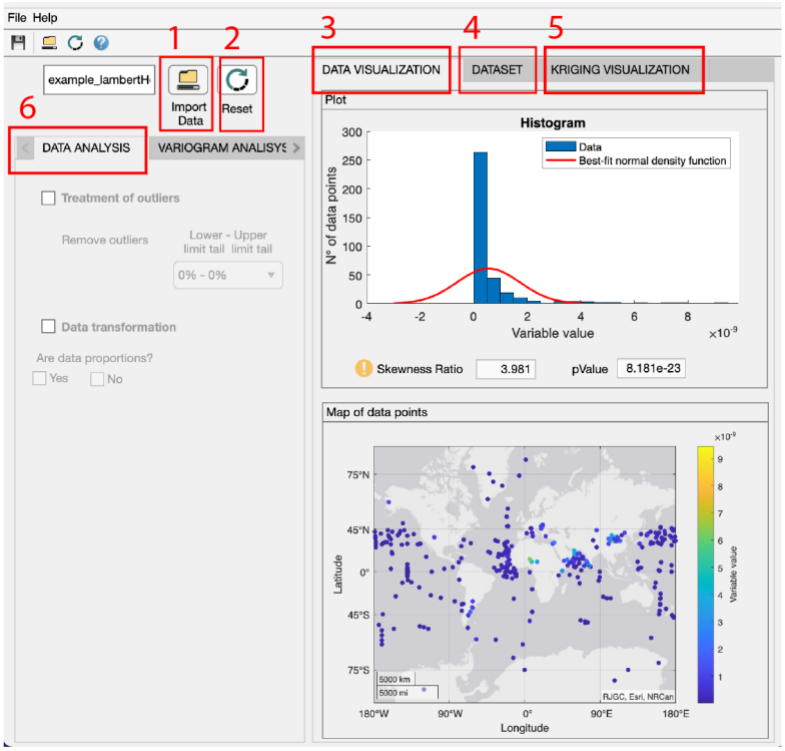


Figure 1. The main window of Global-Krigger's GUI.

Variogram analysis

Figure 2 shows the screen that is displayed when the *Variogram Analysis* tab (1) is selected. Choosing the *Variogram Cloud* option (2) plots the variogram cloud, which shows the semivariance of each pair of points in the dataset versus their distance. If Global-Krigger's GUI is launched from MATLAB, the variogram cloud plot is interactive and is coupled to the map

plot. If one point is selected in the variogram cloud plot and then the *See selected data on map* button is clicked, the two points that together define the selected point in the variogram cloud are displayed in the map. More than one point can also be selected in the variogram cloud. If desired, the selected points can be removed from the dataset by clicking on the *Remove data* button. Visualizing the variogram cloud is useful to identify data points that are outliers in terms of their semivariance. This kind of outlier points may survive even when outliers in terms of the original variable were eliminated previously.

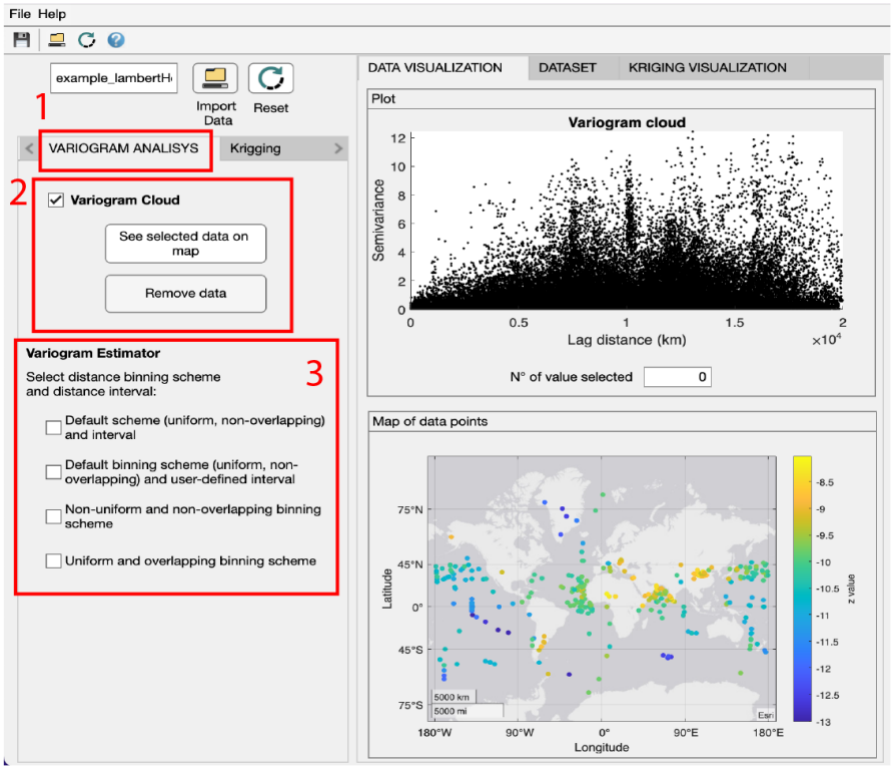


Figure 2. The *Variogram Analysis* window of Global-Krigger's GUI.

The *Variogram Estimator* option (3, Figure 2) is the first required step towards kriging after loading the dataset. This will calculate the variogram estimators (mean semivariance) for each distance bin that is defined. The default scheme is one in which all distance bins are equal in width (uniform) and do not overlap. In this case, the distance bin width is calculated as the mean of the minimum distances between pairs of points. Other options are available, as described in the options of the *Variogram Estimator* window. After a choice is made for the distance binning scheme, the variogram cloud plot is replaced by the variogram estimator plot.

Variogram modeling and kriging

Figure 3 shows the two panels of the *Kriging* tab (1), the *Variogram Model* panel (2), and the *Kriging Interpolation* panel (3). In the *Variogram Model* panel, a set of parameters are chosen to fit a curve (model) to the variogram estimators. An upper bound to distance may be defined, above which a constant value is modeled. Another key decision is whether to define weights for

modeling the variogram estimators or not, and if defined, how to do so. Some variogram estimators may be better defined than others, due, for example, to more pairs of points used to define some variogram estimators than others. Four weighting scheme options are included, which are explained in full in the question mark icon in panel 2. Once one weighting scheme is chosen, the variogram estimators are modeled with several functions, and the system recommends one model based on its goodness of fit and its number of parameters, both included in the Akaike Information Criterion (AIC) displayed below the upper plot. The model with the lowest AIC is recommended.

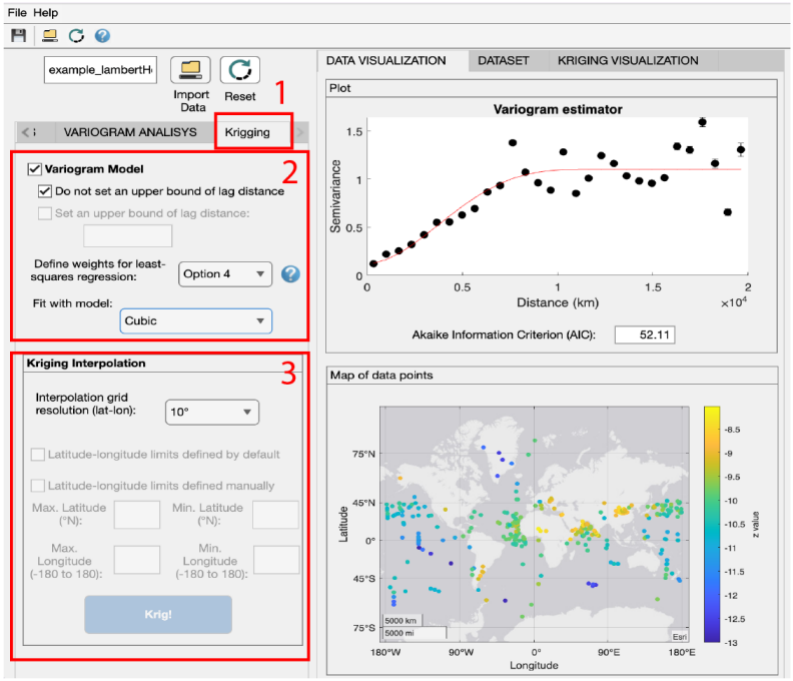


Figure 3. The *Variogram Modeling and Kriging* window of Global-Krigger's GUI.

Finally, once a model is chosen, an interpolation grid needs to be defined, as well as the latitude-longitude extent of the interpolation domain. If the default option is chosen for the latter, then the maximum and minimum latitude and longitude are defined as those of the dataset (so that no extrapolation is performed).

Click on the *Krig!* button to perform kriging. Once the process is complete, go to the *Kriging Visualization* tab to find the interpolated and uncertainty fields.

Data export

In the *File* drop-down the option to export figures (as .jpg, .png, .pdf or .svg) and data (as .dat, .txt, .csv, .mat, .xls and .xlsx) are included.

Launching the GUI without MATLAB

MATLAB is not required to run Global-Krigger: you may download included standalone programs for macOS (*GlobalKrigger_v1.0.0_noMATLAB_GUI_macOS.zip*), Linux (*GlobalKrigger_v1.0.0_noMATLAB_GUI_Linux.install*) and Windows (*GlobalKrigger_v1.0.0_noMATLAB_GUI_Windows.exe*), included in this distribution. The software capabilities are the same as those of the MATLAB-run GUI, except for the interactive plots of the variogram cloud and map (see *Variogram Analysis* section, page 5). The files that contain the dataset that will be used with the toolbox need to be placed in the same directory as the executable.