R codes to be used within QGIS for Precision Agriculture applications

Corentin LEROUX – Aspexit – www.aspexit.com

The functionalities presented in this document allow to **manipulate and process Precision Agriculture data with R directly from QGIS**.

The R codes are available on the GitHub account of Aspexit. **Once the codes have been downloaded, they should be put in the right folder** :

« C:\Users\ASUS\AppData\Roaming\QGIS\QGIS3\profiles\default\processing\rscripts »

You can find this folder directly when clicking on the R icon in the Processing Toolbox (Figure 1), and then clicking on saving the script in the new window (Figure 2). If the R icon is not present, be sure to have installed the QGIS plugin « Processing R Provider » and to have R well interfaced with QGIS (see figure 4).

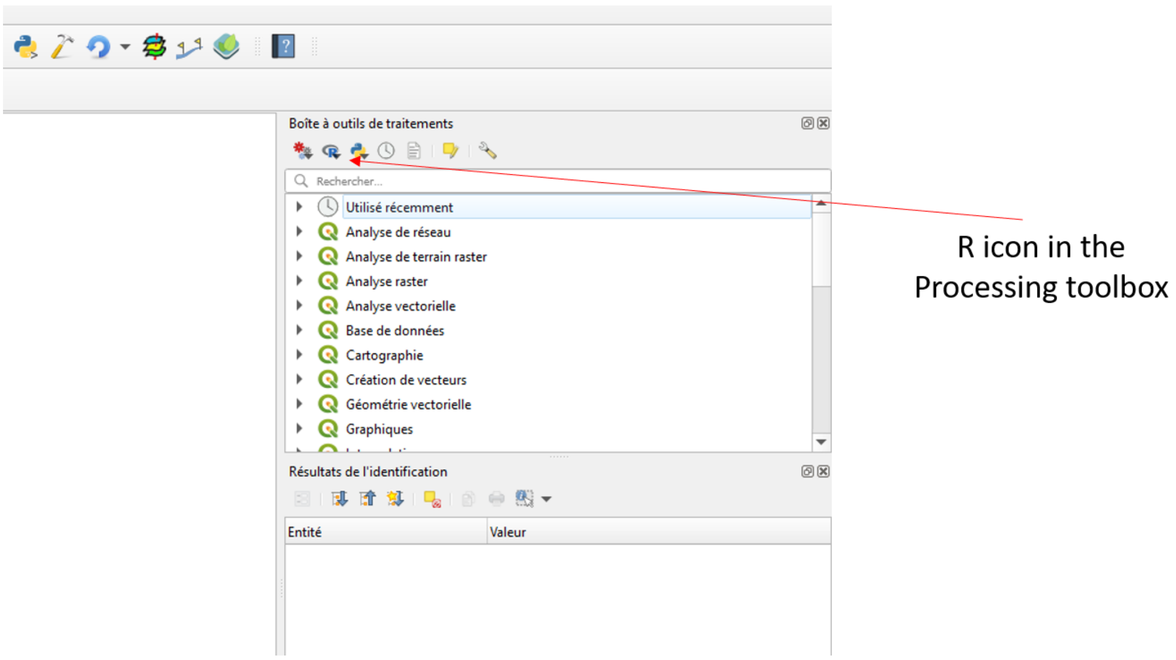

Figure 1. R icon in the Processing Toolbox.

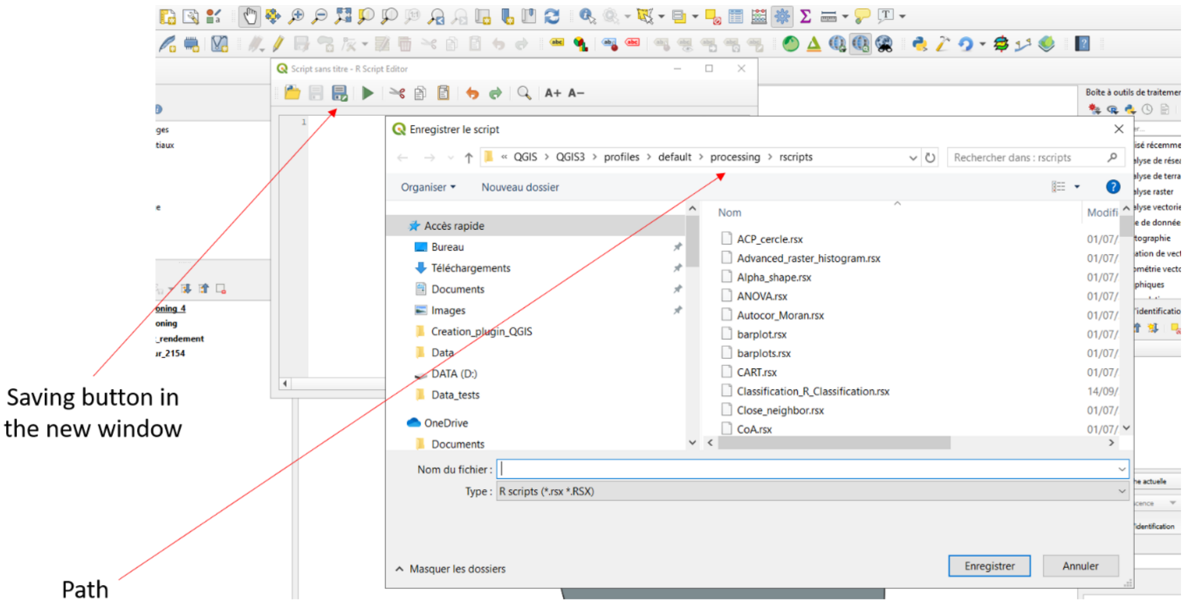

Figure 2. Path for R scripts

R codes in this folder get the extension « .rsx ». The files with extension « .rsx.help » contain the help associated to each function (this help will be displayed when you run the functions). After downloading the functions and putting them in the right folder, those should appear in the processing toolbox in the "R" folder and the "Precision Agriculture" sub-folder (Figure 3).

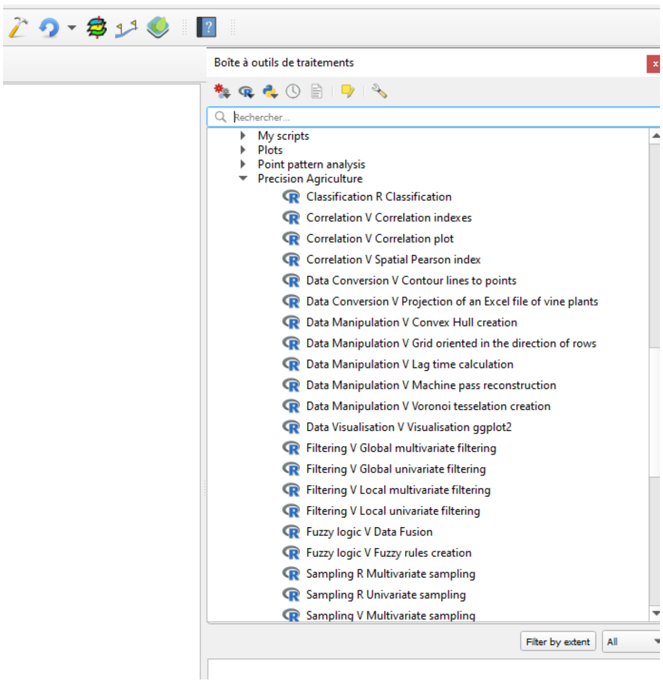

Figure 3. R scripts in the Processing Toolbox

The use of these functions requires (1) having R installed on your machine, and (2) QGIS well interfaced with R (Figure 4).

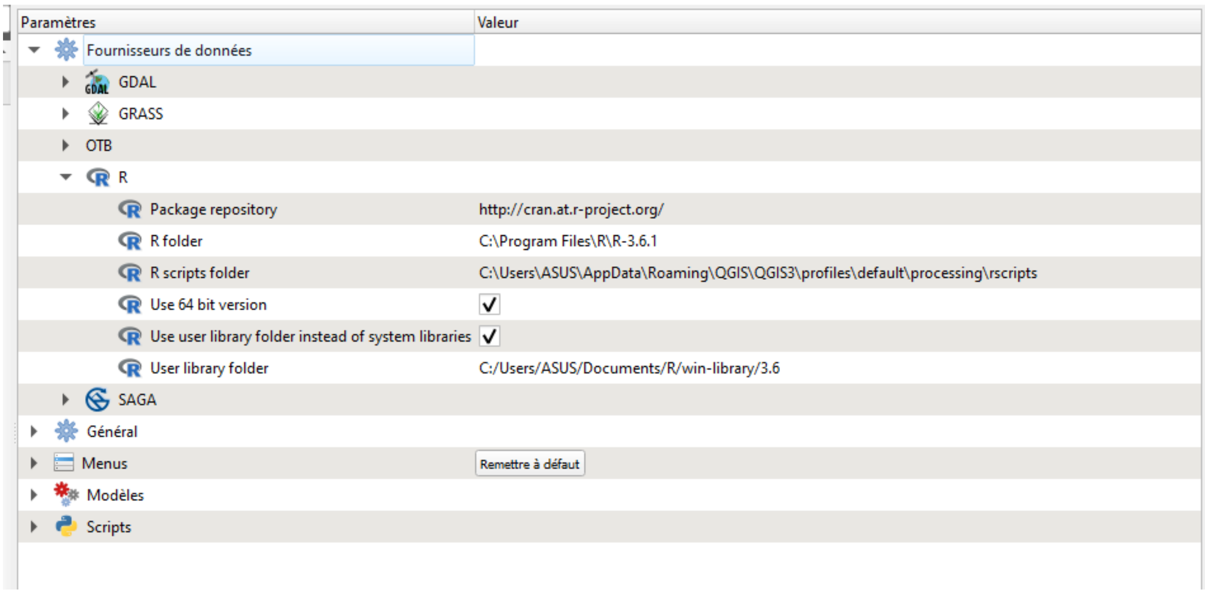

Figure 4. Interfacing R and QGIS. This can be accessed by clicking on the option button at the extreme right of the R icon that you just clicked in the processing toolbox.

The functions presented in this document are those available in R. They are grouped by themes (Correlations, Sampling, Spatial Analysis…). Functions applied on rasters have a prefix 'R'. Functions applied on vectors have a prefix 'V'.

Some of the few problems that may arise :

- If possible, try to save the outputs of the functions. Temporary layers can sometimes cause problems when use directly in another algorithms
- Always make sure that the projection of the data is okay. Otherwise, you might have some troubles
- Outputs of the functions, whatever tables or shapefiles, very often have a column named « fid ». This column can sometimes cause problems when the outputs are used as inputs in other functions because QGIS will try to generate a new fid column while there already exists one.

TABLE DES MATIERES

1. Classification

1.1. R – Classification

Allows to reclassify a raster into a user-defined number of classes using several classification methods

- **Inputs**
  - Raster layer
  - Classification method
    - Equal : Equal-intervals
    - Quantiles
    - K-means
    - Jenks
  - Number of classes

- **Outputs**
  - Classified raster

2. Correlation

2.1. V – Correlation indexes

Allows to calculate a correlation index between two fields (column) of a vector layer. Users have to select the two numerical fields for which they need the correlation index

- **Inputs**
  - Vector layer
  - Correlation method (Kendall, Spearman, Pearson)
  - The two numerical fields for which correlation is calculated

- **Outputs**
  - Correlation index

2.2. V – Correlation plot

Allows to compute a correlation plot between all the numerical variables available. The layer should solely contain numerical variables. A categorical variable can be used to color the correlation plots in different groups (ex : a column with values of 0 for group A and values of 1 for group B).

- **Inputs**
  - Vector layer
  - Color factor : Yes if a field will be used to categorize the data. No otherwise
  - Variable color : Field that is used to categorize the data

- **Outputs**
  - Correlation plot

2.3. V – Spatial Pearson index

Performs a modified version of the *t* test to assess the correlation between two spatial processes. Users have to select the two numerical fields for which they need the correlation index

- **Inputs**
    - Vector layer
    - The two numerical fields for which correlation is calculated

- **Outputs**
    - Correlation index

3. Data Conversion

3.1. V – Contour lines to points

Allows to convert elevation contour lines into elevation points

- **Inputs :**
  - Line layer
- **Outputs :**
  - Point layer

3.2. V – Projection of an Excel file of vine plants

Allows to transform an Excel spreadsheet identifying vine stocks (each cell of the spreadsheet represents a vine stock) into a geolocalized vector layer. It seems that loading tables does not work when manipulating R within QGIS. The code should then be used in R outside of QGIS.

- **Inputs :**
  - CSV file
  - Coordinates
    - The precise X/Y position of the first vine stock of the first row of vines
    - The precise X/Y position of the last vine of the first row of vines
    - The precise X/Y position of the first vine of the last row of vines
  - The inter row
  - The intra row
- **Outputs :**
  - Spatial dataset

4. Data Manipulation

4.1. V – Convex hull creation

Allows to build a convex envelope (contour) around a point vector layer

- **Inputs :**
  - Point layer
- **Outputs :**
  - Contour of the points

4.2. V- Grid oriented in the direction of rows

For a given point layer oriented in a particular direction (following rows or a machine passage for example), the function builds a grid oriented in the majority direction of the points. The size of the grid is defined by the user. Prerequisite: The order of the points in the layer must follow the data acquisition.

- **Inputs :**
  - Point layer
  - Contour of the field
  - Mesh length and width
- **Output**
  - Grid

4.3. V – Lag time calculation

Allows to compute the machinery lag time in high resolution spatial datasets arising from mounted sensors. Spatial observations are shifted until the correlation in the data is maximal.

- **Inputs :**
  - Point layer
  - Numerical field that will be used to calculate the lag time
- **Output**
  - Lag time value
  - Plot Lag time vs. Temporal Shift

4.4. V – Machine pass reconstruction

For a point vector layer acquired following a machine passage by a sensor onboard the machine (e.g. a yield sensor onboard a combine harvester, a dose level sprayed by a sprayer), the function reconstructs the spatial footprint of the machine at the level of each collected observation. This function makes it possible to find out whether or not there are overlaps between the different machine passes. The user must enter the working width of the machine

- **Inputs :**
  - Point layer
- **Output**
  - Spatial footprint polygons

4.5. V – Voronoi tesselation creation

Allows to build tesselations from point vector data

- **Inputs :**
  - Point layer
- **Output**
  - Voronoi polygons

5. Data Visualisation

5.1. V – Visualisation ggplot2

Allows to generate R graphics from a vector layer using the ggplot 2 package.

- **Inputs :**
  - Vector layer
  - Graph type : Points, Lines, Boxplots
  - X-axis
  - Y-axis
  - Color type : no colour ; coloring ; filling. Users have first to select the color type before selecting the field to be used as coloring factor.
  - The use of facets to categorize the displayed data (users have to select whether they want facets or not, and then choose the field to be used as facet).
  - Font sizes (x and y axes, x and y titles, facets 1 and 2)

- **Outputs :**
  - ggplot2 graph

6. Filtering

6.1. V – Global multivariate filtering

Allows to perform a multivariate filtering to remove outliers from a vector layer. The filtering methods are NOT spatialized, they only look at the global distribution in the data.

- **Inputs :**
  - Vector layer
  - Filtering method :
    - DBSCAN (for two variables) : the main group of observations is considered normal. Remaining observations are considered outliers (inspired from Leroux et al. 2018 - A general method to filter out defective spatial observations from yield mapping datasets)
    - Mahalanobis distance (for two variables and more). Robust covariance matrix is constructed. The method assumes that data is normally distributed.
  - Percentile : Find the Xth percentile of the Chi-Squared distribution (for Mahalanobis method only)
  - Epsilon neighbourhood and minimum number of points in epsilon region : for DBSCAN method only
  - Outlying borders (for DBSCAN method only) : whether points at the border of the main group are considered outliers or not.
  - Outliers handling : whether outliers are marked or removed.

- **Outputs :**
  - Outliers

6.2. V – Global univariate filtering

Allows to perform a univariate filtering to remove outliers from a vector layer. The filtering methods are NOT spatialized, they only look at the global distribution in the data.

- **Inputs :**
  - Vector layer
  - Filtering method :
    - Normal : outliers removal assuming normal distribution in the data. Outliers are removed if outstide the range mean+/- 3 standard deviations
    - Tukey : Tukey interquartile method
    - Skewed : outliers removal when data is skewed (inspired from Hubert,

M., & Van der Veeken, S. (2008). Outlier detection for skewed data)
    - Outliers handling : whether outliers are marked or removed.

- **Outputs :**
    - Outliers

6.3. V – Local multivariate filtering

Allows to perform a multivariate spatial filtering to remove outliers from a vector layer.

- **Inputs :**
    - Vector layer
    - Filtering method :
        - DBSCAN (for two variables) : the main group of observations is considered normal. Remaining observations are considered outliers (inspired from Leroux et al. 2018 - A general method to filter out defective spatial observations from yield mapping datasets)
        - Mahalanobis distance (for two variables and more). Robust covariance matrix is constructed. The method assumes that data is normally distributed.
    - Percentile : Find the Xth percentile of the Chi-Squared distribution (for Mahalanobis method only)
    - Neighbourhood distance
    - Epsilon neighbourhood and minimum number of points in epsilon region : for DBSCAN method only
    - Outlying borders (for DBSCAN method only) : whether points at the border of the main group are considered outliers or not.
    - Outliers handling : whether outliers are marked or removed.

- **Outputs :**
    - Outliers

6.4. V – Local univariate filtering

Allows to perform a univariate spatial filtering to remove outliers from a vector layer.

- **Inputs :**
    - Vector layer
    - Filtering method :
        - CV : Coefficient of variation (inspired from Spekken et al. (2013). A

simple method for filtering spatial data

  - IDW : Each point value (Zi) is compared to interpolated value from neighbours (Zj). If Zi is outside the range Zj+/- n standard deviations of neighbourood values, the point is considered an outlier (inspired from Simbahan et al., 2004 : Screening yield monitor data improves grain yield maps)
  - Moran : remove outliers based on Moran spatial autocorrelation statistics (inspired from Gozdowski et al., 2010 – Evaluation of methods for the detection of spatial outliers in the yield data of winter wheat).
  - Normal : outliers removal assuming normal distribution in the data
- Maximum CV (for CV method only)
- Distance neighbourhood
- Number of standard deviations
- Min and max neighbourhood distance (for Moran method only)
- Outliers handling : whether outliers are marked or removed.

- **Outputs :**
  - Outliers

7. Fuzzy logic

7.1. V – Data Fusion

Allows to aggregate multiple numerical fields that were previously assigned a membership function. Because the function requires users to set lots of parameters (membership functions for each numerical fields), users are asked to edit the code manually. The function is available on the GeoFIS software also (https://www.geofis.org/en/documentation-en/data-fusion/)

- **Inputs :**
  - Vector layer
  - Aggregation
    - WAM (Weighted arithmetic mean)
    - OWA (Ordered Weighted average)

- **Outputs**
  - A new column in the vector layer specifying the membership of each observation

7.2. V – Fuzzy rules creation

Allows to generate fuzzy rules for a numerical field

- **Inputs :**
  - Vector layer
  - Shape (triangular, trapezoidal, trapezoidal inferior, trapezoidal superior)
  - Rule_name : prefix of the new column name that will be based on the specified fuzzy rule
  - Parameters of kernel and support (lower kernel, lower support, upper kernel, upper support) depending on the membership function selected

- **Outputs**
  - A new column in the vector layer specifying the membership of each observation

8. Sampling

8.1. V – Multivariate Sampling

Allows to perform a multivariate sampling via a conditioned Latin Hypercube method. The layer should only contain the numeric fields that will be used for sampling.

- **Inputs :**
  - Vector layer
  - Number of samples

- **Outputs :**
  - Sampling points
  - Boxplots to show the representativity of sampling points.

8.2. V – Sampling within polygon

Allows to carry out a sampling in a polygon. The number of points is defined by the user and several sampling schemes are available

- **Inputs :**
  - Polygon Contour
  - Sampling scheme (random, regular, stratified, Non-aligned, Clustered)

- **Outputs :**
  - Sampling points

8.3. V – Univariate Sampling

Allows to sample a user-defined number of points in a vector layer. Several sampling methods are available. The layer should only contain the numeric fields that will be used for sampling.

- **Inputs :**
  - Vector layer
  - Sampling method (Random, Quantiles, K-means)
  - Number of groups if the objective is to perform a stratified sampling (only for quantiles and k-means)
  - Number of sampling points in each group (groups are not considered for the random method).

- **Outputs :**
  - Sampling points

9. Spatial analysis

9.1. R – Landscape metrics

Allows to calculate a set of landscape ecological metrics on a classified raster. Several indexes are available : Shannon, Richness, Cohesion, Division, Contagion, Interspersion and juxtaposition

- **Inputs :**
  - Classified raster
  - Choice of Neighbourhood : Queen/Rook

- **Outputs :**
  - Landscape metrics

9.2. V – Gaussian random fields generation

Allows to generate a raster layer according to a user-defined spatial structure. For the spatial footprint considered by the user, the user must fill in the mean of the attribute, its variance, and the components of the variogram (range, nugget effect, sill).

- **Inputs :**
  - Polygon contour
  - Method to generate sampling points within polygon
  - Number of sampling points
  - Mean and coefficient of variation of the future variable
  - Variogram parameters

- **Outputs**
  - Spatial points with specified spatial structure

9.3. V – Geary/Moran statistics

Allows to calculate Geary or Moran spatial statistics. During the calculation, matrices are row-normalized (style « W »). Other styles exist : it can be changed by editing the code.

- **Inputs :**
  - Layer : Spatial Points
  - Number of k-nearest neighbours
  - Maximum lag order
  - Choice of index : Geary / Moran
  - Computation method :
    - Spatial : spatial weights matrix in weights list form
    - Monte-Carlo : permutations

- **Outputs :**
  - Moran/Geary statistics

9.4. V – Getis Ord statistics

Allows to calculate Getis Ord spatial statistics

- **Inputs :**
  - Layer : Spatial Points
  - Neighbourhood distance

- **Outputs :**
  - Getis Ord statistics

9.5. V – Joint Count statistics

Allows to calculate joint count statistics (for categorical variables)

- **Inputs :**
  - Layer : Spatial Points
  - Neighbourhood distance
  - Computation method :
    - Spatial : spatial weights matrix in weights list form
    - Monte-Carlo : permutations

- **Outputs :**
  - Joint count statistics

9.6. V – Moran correlogram

Allows to plot a Moran correlogram. During the calculation, matrices are row-normalized (style « W »). Other styles exist : it can be changed by editing the code.

- **Inputs :**
  - Layer : Spatial Points
  - Number of k-nearest neighbours
  - Maximum lag order

- **Outputs :**
  - Moran correlogram

9.7. V – Moran scatterplot

Allows to plot a Moran scatterplot. During the calculation, matrices are row-normalized (style « W »). Other styles exist : it can be changed by editing the code.

- **Inputs :**
  - Layer : Spatial Points
  - Number of k-nearest neighbours

- **Outputs :**
  - Moran scatterplot

10. Spatial Interpolation

10.1. V – Experimental variogram

Generates an experimental variogram from a user-selected numeric field (column) in a vector layer. The user can choose the maximum distance up to which the experimental variogram is constructed, and the distance step (lag) for which the semi-variance is calculated. The vector layer must be projected !

- **Inputs :**
  - Vector layer
  - Max_dist_vario : maximum distance up to which the experimental variogram is constructed
  - Lag : distance step for which the semi-variance is calculated
  - Trend : constant or spatial

- **Outputs :**
  - A .png image presenting the experimental variogram
  - A spreadsheet with the input data of the experimental variogram (number of peers, lags, semi-variances...)

10.2. V – Optimal grid size for interpolation

Defines the optimal size of an interpolation grid for a given field (column) of a vector layer. The user must fill in the components of the theoretical variogram related to the chosen field [from Tisseyre et al. (2018) How to define the optimal grid size to map high resolution spatial data?

- **Inputs :**
  - Vector layer
  - Field contour
  - Variogram parameters (model, range, partial sill, nugget effect)

- **Outputs :**
  - Grid size

10.3. V – Optimal grid size for interpolation GSTAT

Calculate, for a given variogram model, ordinary block kriging standard errors as a function of sampling spaces and block sizes (ossfim function in the gstat package)

- **Inputs :**
    - Vector layer
    - Field contour
    - Variogram parameters (model, range, partial sill, nugget effect)
- **Outputs :**
    - Grid size

10.4. V – Spatial interpolation

Generates the interpolation of a user-defined numeric field (column) of a vector layer. A raster is generated. The user can choose the size of the interpolation grid and the interpolation method.

- **Inputs :**
    - Vector layer
    - Interpolation method
        - Inverse distance weighting: the user can choose a slightly finer weighting method (inverse, exponential, ...), and choose the interpolation power.
        - Point kriging: the user fills in a theoretical variogram model (see function X). The interpolation map is accompanied by a kriging variance map (interpolation uncertainty).
        - Block kriging : the user enters a theoretical variogram model (see function X). The interpolation map is accompanied by a kriging variance map (interpolation uncertainty).
    - Size_grid_interpolation
    - Dist_max_neighbours
    - IDW_power : power of interpolation (for IDW method only)
    - Variogram model (for point and block kriging method only)
    - Variogram display (Max_dist_vario, Lag)
    - Variogram parameters (range, nugget, sill)

- **Outputs :**
    - Interpolated data

10.5. V – Variogram modelling

Allows to adjust a theoretical variogram to an experimental variogram. The adjustment is done automatically after entering the initialization values of the theoretical variogram (range, nugget effect, sill, variogram model). Several variogram models can be parameterized: Gaussian, Spherical, Linear, Exponential. The user can choose the maximum distance up to which the experimental variogram is constructed, and the distance step (lag) for which the semi-variance is calculated. Users get variogram-based variability indexes (see Leroux and Tisseyre (2018) How to measure and report within-field variability: a review of common indicators and their sensitivity). The vector layer must be projected!

- **Inputs :**
    - Vector layer
    - Max_dist : maximum distance up to which the experimental variogram is constructed
    - Lag : distance step for which the semi-variance is calculated
    - Trend : constant or spatial
    - Variogram parameters (model, range, partial sill, nugget effect)

- **Outputs :**
    - A .png image showing the theoretical variogram fitted to the experimental variogram.
    - A spreadsheet with the parameters of the theoretical variogram"
    - A spreadsheet with spatial structure indicators calculated from the theoretical variogram model
        - Cambardella index
        - MCD index

11. Zoning

11.1. V – Fusion of zones

Allows to merge the zones of a zoning according to constraints set by users. The final number of zones is smaller than the initial number of zones

- **Inputs :**
  - Initial zoning
  - Method
    - Number zones : Zones are merged iteratively until the final number of zones is reached
    - Minimal distance : Zones are merged iteratively until the threshold difference between zones is reached
    - Both : Zones are merged iteratively until the number of zones is reached and the constraints on the threshold difference between zones is respected.
  - Final number zones : Final number of zones in the zoning
  - Threshold : Minimal difference in value between zones

- **Outputs :**
  - Final zoning

11.2. V – Geozoning

Allows to generate a zoning based on the R geozoning package (inspired from Loisel et al. 2019 – An optimisation-based approach to generate interpretable within-field zones)

- **Inputs :**
  - Point layer
  - Step : Related to the iterations of the algorithms. The higher the step parameter, the longer the computation time ; but the algorithm might find better zonings.
  - Number classes : Zones are created so that they at least are exhaustive of the number of classes that is required.

- **Outputs :**
  - Final zoning

11.3. V – Merging two zonings

Allows to merge two zonings of the same field into one single zoning. The final zoning can be considered as a microzoning that combines the zones of both initial zonings

- **Inputs :**
  - First zoning
  - Second zoning

- **Outputs :**
  - Merged zoning

11.4. V – Region Growing

Allows to generate a zoning (region growing algorithm) from punctual spatial observations. Inspired from Leroux et al. (2018)

- **Inputs :**
  - Point layer
  - Contour of the field
  - IDW smoothing : Yes/no. In case the initial data is very noisy
  - Size_grid_interpolation : (applies if IDW smoothing is performed)
  - IDW power : power of interpolation (applies if IDW smoothing is performed)
  - Dist border : distance to edges of the field. Prevent core of zones to be too close to borders to avoid border effects
  - Step :
  - Grid footprint

- **Outputs :**
  - Zoning

11.5. V – Zoning quality index

Allows to calculate a zoning quality indicator in relation to available auxiliary information. The indicator should fill in the zoning vector layer, and the vector layer to which the zoning is to be compared. Several indices are available :

- **Inputs :**
  - Zoning
  - Point layer for which the quality of zoning need to be adressed

- Index
    - Variance reduction (from Bobryk et al., 2016 - Validating a Digital Soil Map with Corn Yield Data for Precision Agriculture Decision Support)
    - Zoning Opportunity Index (inspired from Roudier et al. 2011 - A technical opportunity index adapted to zone-specific management)
- Machine width (for Zoning Opportunity index) : machine footprint
- Step (for Zoning Opportunity index) : Sensibility of the machinery. A step of 5 means that the machine is able to apply two treatments levels differing from 5 units.

- **Outputs :**
    - Zoning quality index