GEMMA (Geo-EnvironMental Multivariate Analysis) Toolbox

---The User Guide---

0. GEMMA Toolbox requirements

To use the GEMMA toolbox, R and R-Studio desktop are required.

1. R is freely available at the link https://www.r-project.org/. In "download/CRAN" section, chose the mirror link preferred for the download suitable for your operative system.
2. Install the R CRAN
3. R-Studio desktop can be downloaded at the link https://posit.co/download/rstudio-desktop/ , chose your operative system and start the download.
4. Install the R-Studio desktop software

Note: to develop GEMMA toolbox we use both Windows 11 and macOS 12 systems, R-Studio 2023.12.1, and R 3.3.0

1. Folder Structure

The GEMMA Toolbox software includes folders (a) for its functionality. The "plot" and "tab" sub-folders store charts and tables generated from the analysis, while the "scripts" contain the software's scripts.

Note:

- Do not delete or rename the folders!
- The "the script" folder contain all the software scripts, programming-skills are required to manage the files without compromise the GEMMA Toolbox functioning.

Steps:

1. Copy your dataset to the software folder directory.
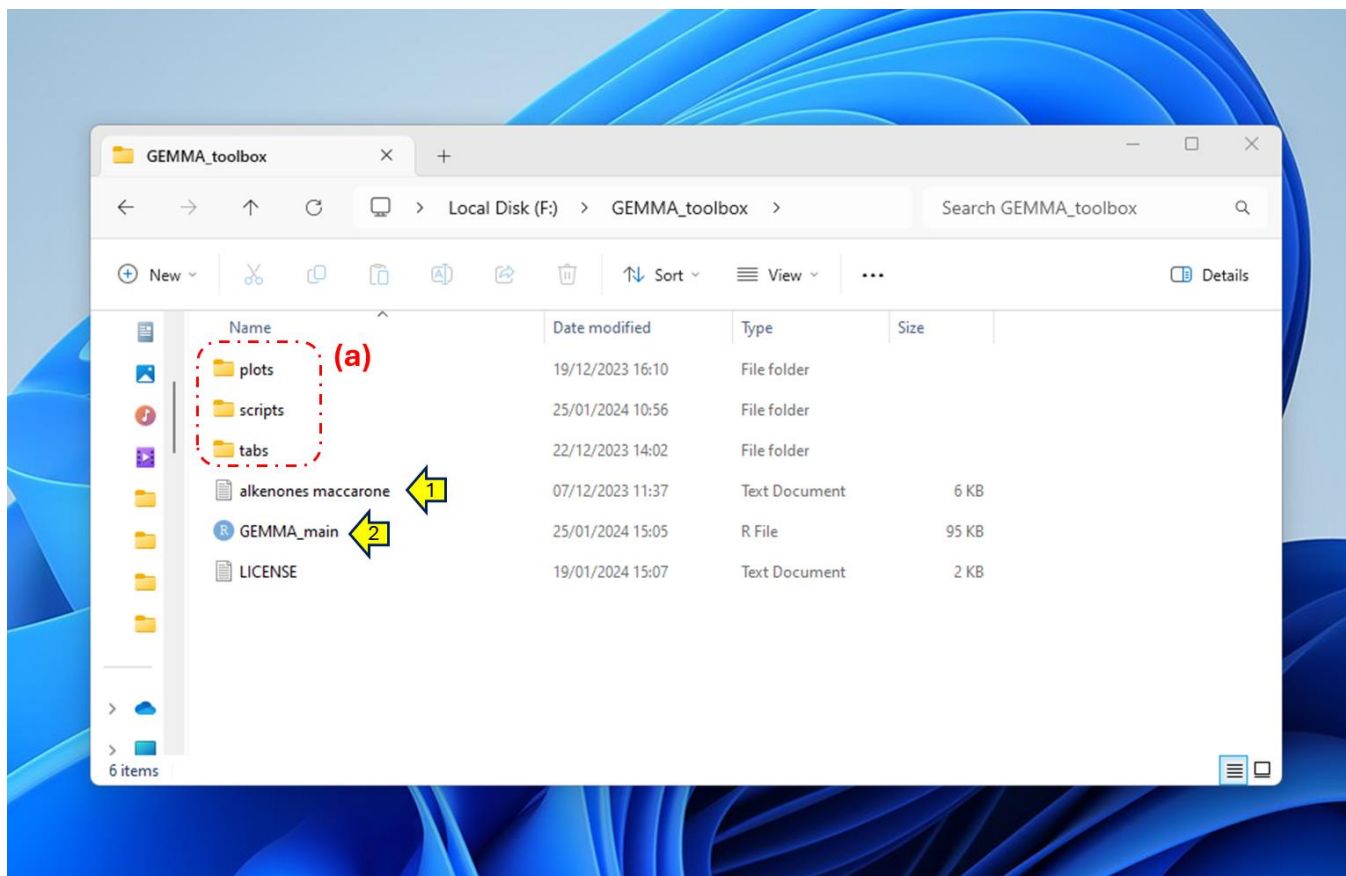2. Run the GEMMA_main.R file using R-Studio.

Figure 1

2. Running GEMMA Toolbox

To run the GEMMA Toolbox software, simply click the "Run App" button, as shown in the figure (yellow arrow).

Note: If running GEMMA Toolbox for the first time, allow some time for automated library installations. The duration may vary based on network and computer performance.
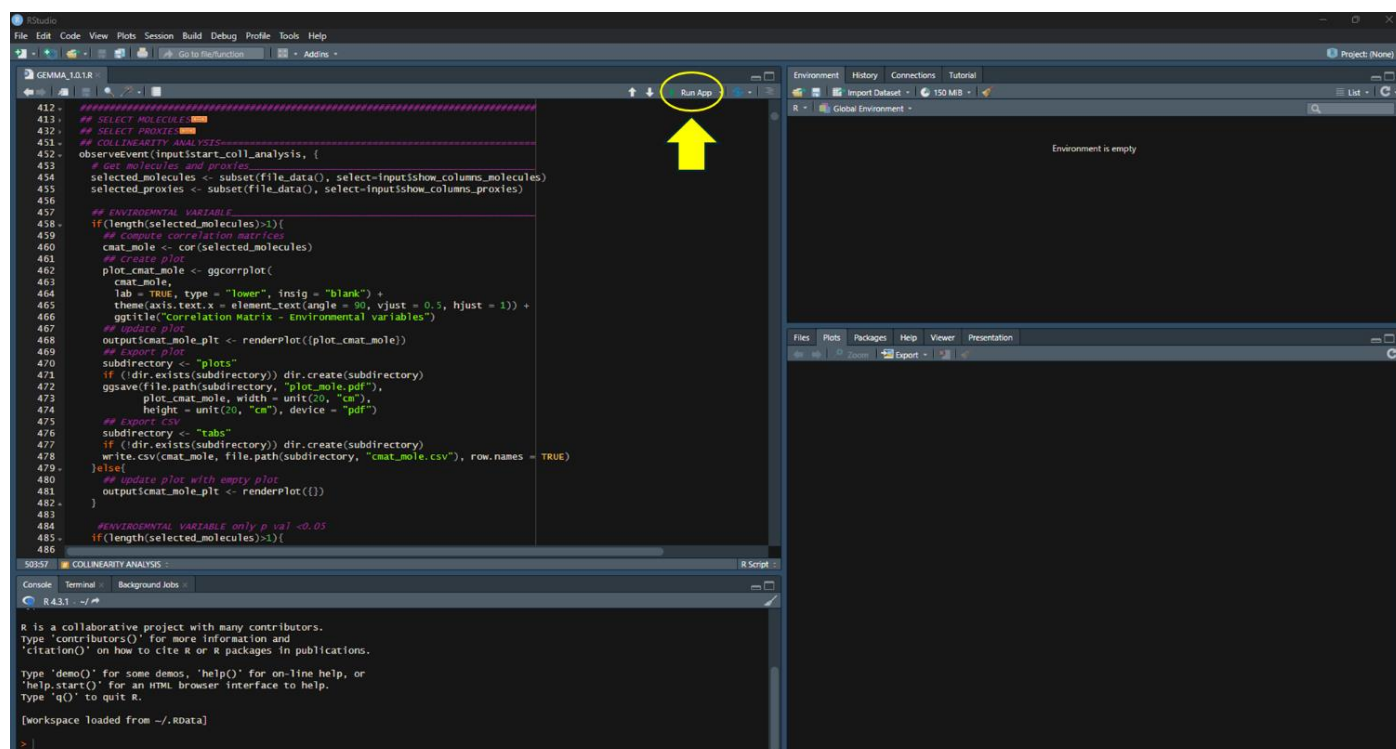

Figure 2

3. Dataset Import

After launching the app, import the dataset:

1. Click "BROWSE..." (1).
2. Locate your dataset file in the GEMMA Toolbox directory (2) and open it (3).

Note: To load a new dataset, refresh the app first (4).
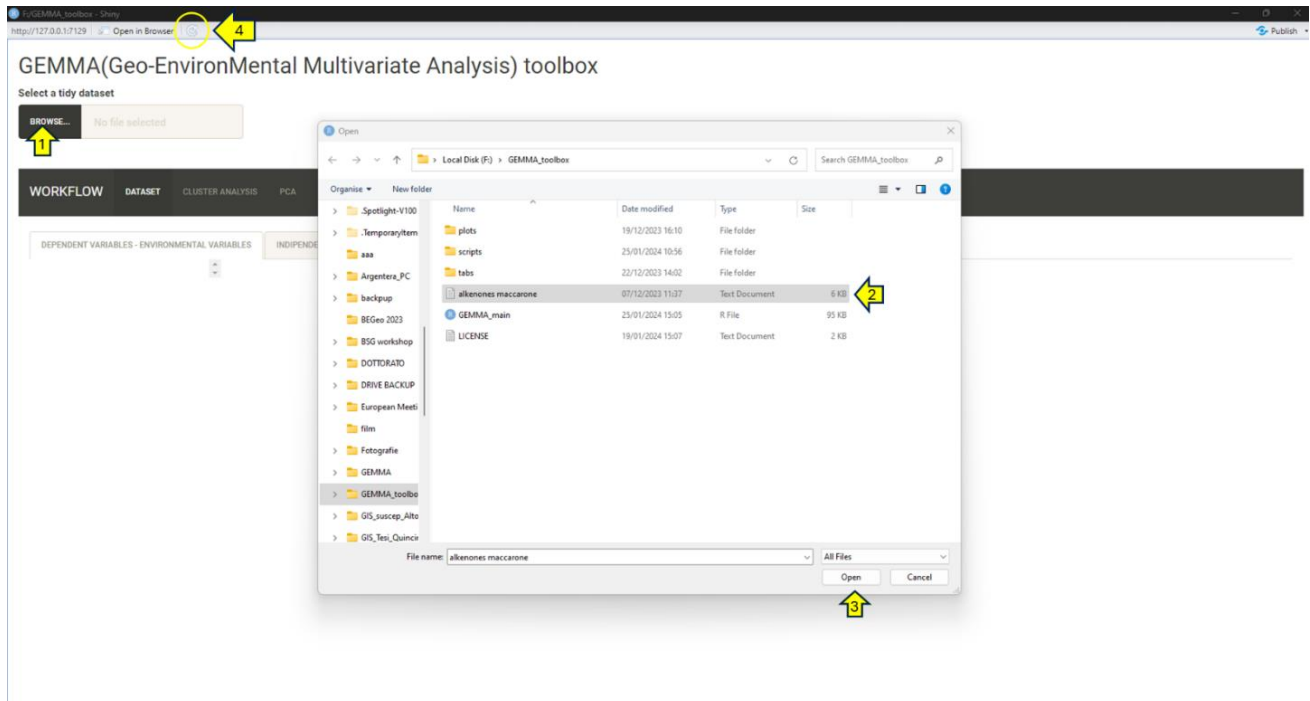
Figure 3

4. Variable Selection & Raw Data Analysis

In the "DATASET" panel, select variables for analysis in both "DEPENDENT VARIABLES" (1) and "PREDICTORS" (3) sub-panels. Use the checkbox (2) to select variables, and the central section (a) to view real-time data.

Figure 4

Analysis Steps:

1. Click "START COLLINEARITY ANALYSIS" (5) in the "COLLINEARITY ANALYSIS" sub-panel (4).
2. For statistical analyses, check the "ENVIRONMENTAL VARIABLES" (1) checkbox. Redundancy analysis requires "PREDICTOR" (3) selection.
3. Note: Detrended Correspondence Analysis (DCA) cannot process negative values; consider data transformation for positivity.
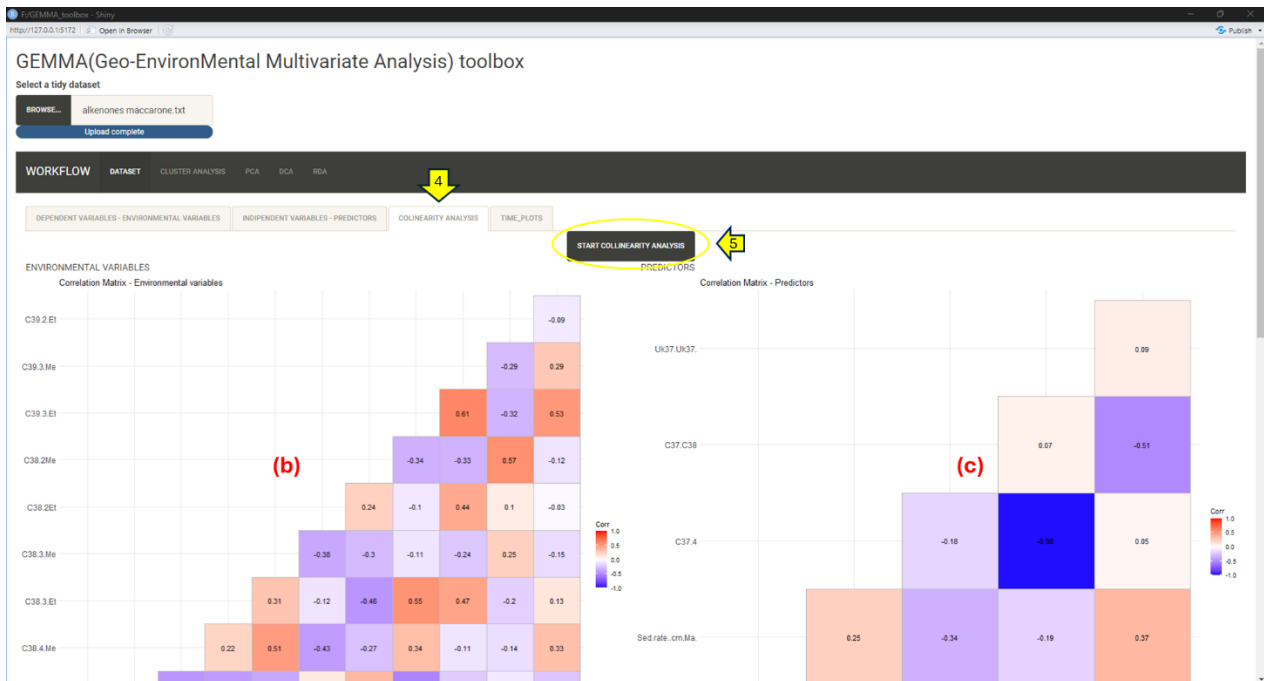

Figure 5

The "TIME_PLOT" sub-section (6) follows a standard structure for raw data, Principal Component Analysis (PCA), Detrended Correspondence Analysis (DCA), and Redundancy Analysis (RDA). Customize the analysis using the dropdown menu (7) and checkbox lists (8, 9). Click "PLOT" buttons (9, 11) to show the relative plots.

Note: the two plots can be displayed independently.

Figure 6

5. Cluster Analysis (CLA)

Perform Cluster Analysis (CLA) in the "CLUSTER ANALYSIS" panel (1):

1. Choose CLA method from the dropdown menu (2).
2. Optionally tune GAP statistic parameters (3).
3. Click "CLUSTER NUMBER" to evaluate cluster number (4) through statistical methods (a).
4. Select the method chosen (5) or set the cluster number manually with the "Manual" option.
5. Click "START CLA" (6) to clustered samples. Results appear in section (b).

Note: the minimum number of clusters is "2", if "1" is selected the CLA will not perform.
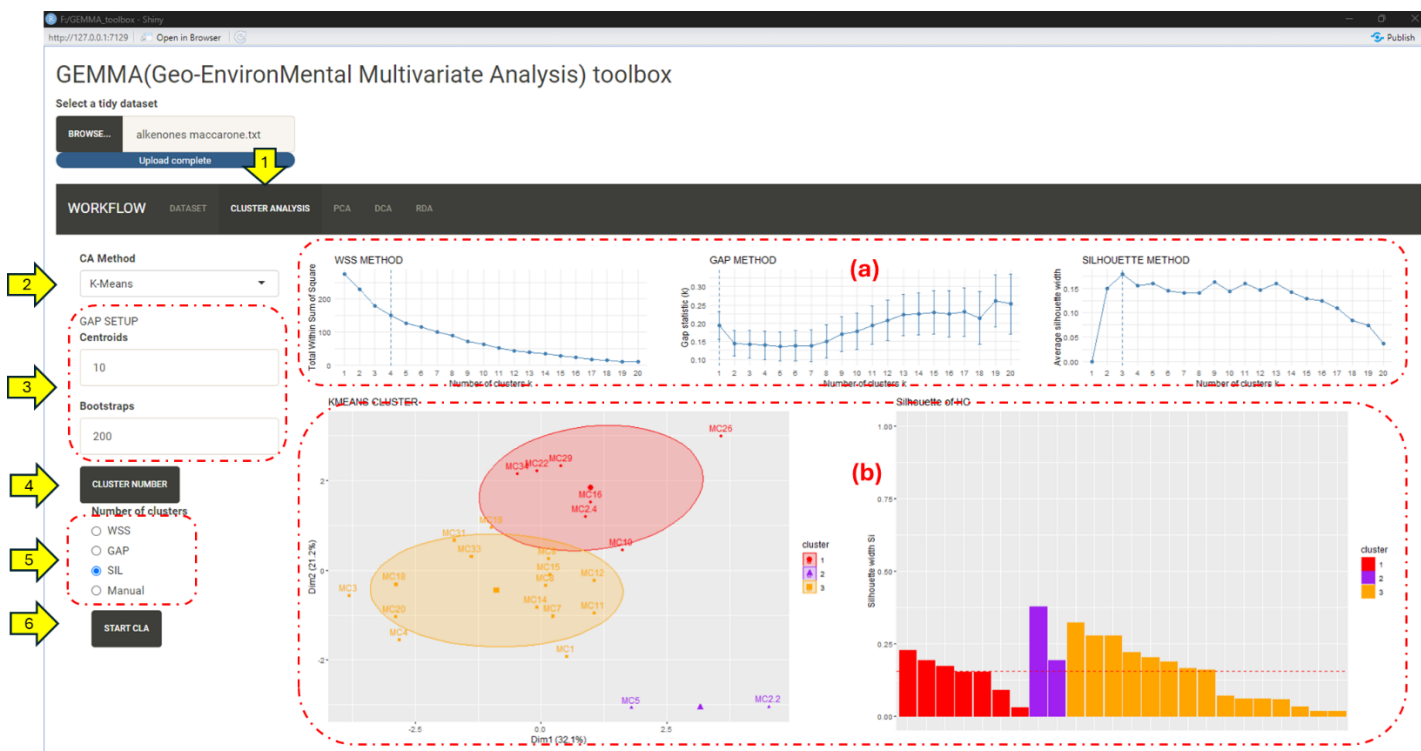
Figure 7

6. Principal Component Analysis (PCA) & Detrended Correspondence Analysis (DCA)

Access PCA and DCA in respective panels (1, 2):

1. In the "MAIN" sub-panel (3), select variables and click "START PCA/DCA" (5).
2. Choose PCs/DCs axis (6), set Montecarlo permutations (7), and run the Montecarlo test by clicking "START MONTECARLO" (8).
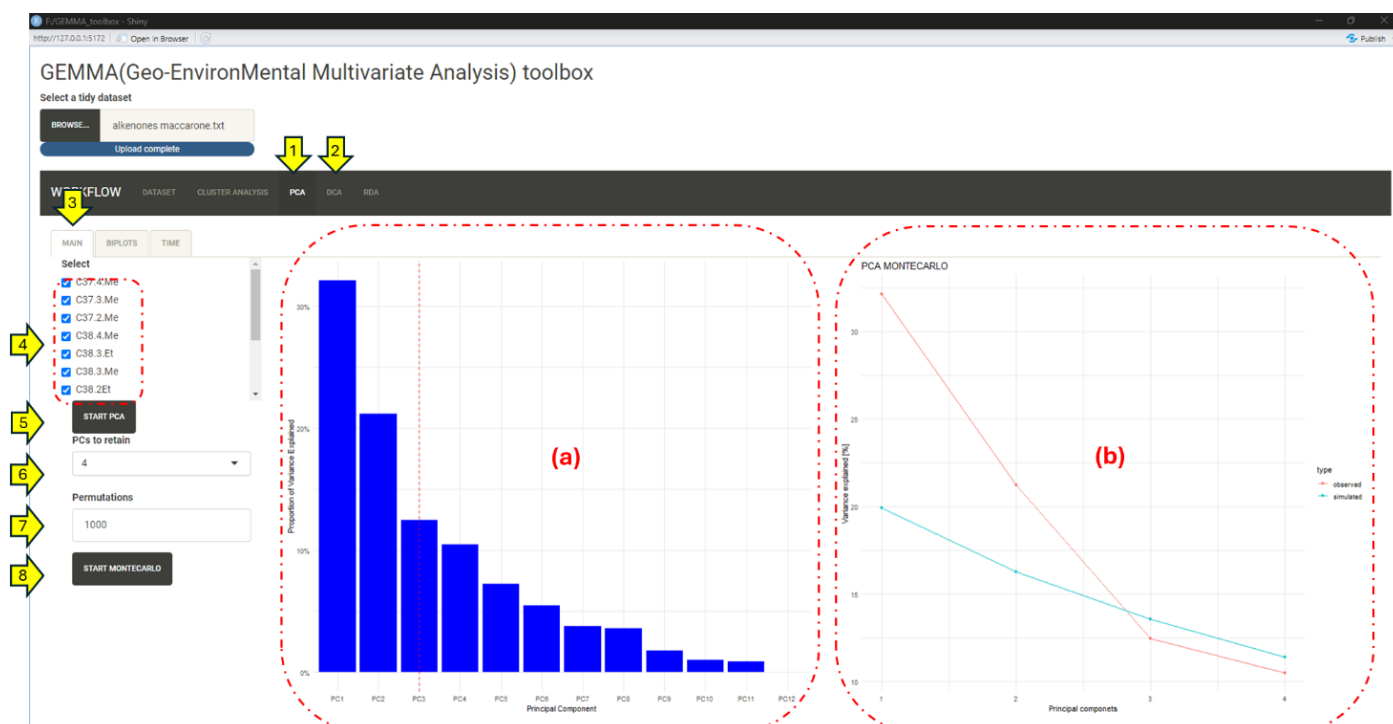

Figure 8

Explore results in the "BIPLOTS" sub-panel (9), selecting components (10) and using the "BIPLOT" button (12).

Note: The "TIME" sub-plot workflow is consistent with raw data analysis.
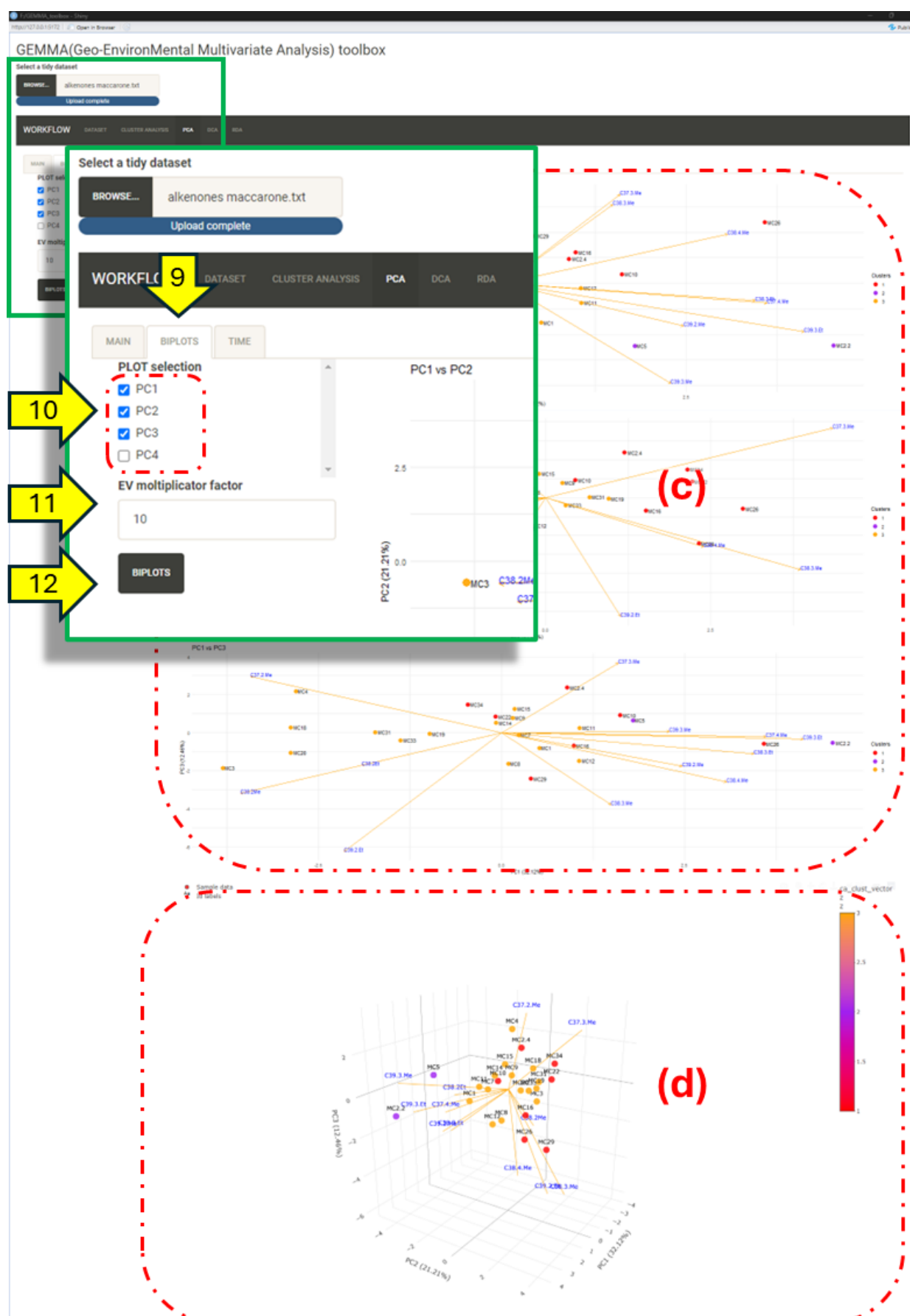

Figure 9

7. Redundancy Analysis (RDA)

Select the "RDA" panel (1) and the "MAIN" sub-panel (2) top perform the RDA analysis.

An iterative process is required to perform the best-balanced model:

    i.   Select predictors from the "PREDICTORS" subset list (3).
   ii.   (*Optional*) Set the Montecarlo iteration number (4).
  iii.   Run RDA test (5).
  iv.   Select the test from the test record list (6) and show the test (7)
   v.   Evaluate RDA model metrics (GLOBAL MODEL, RDA_AXIS, MULTICOLLINEARITY, PREDICTORS TERMS).
  vi.   Iterate (i-v) until optimal metrics align with research goals.

In the lower section of the "MAIN" panel, the screeplot (a) and heatmap (b) of the showed test are plotted.

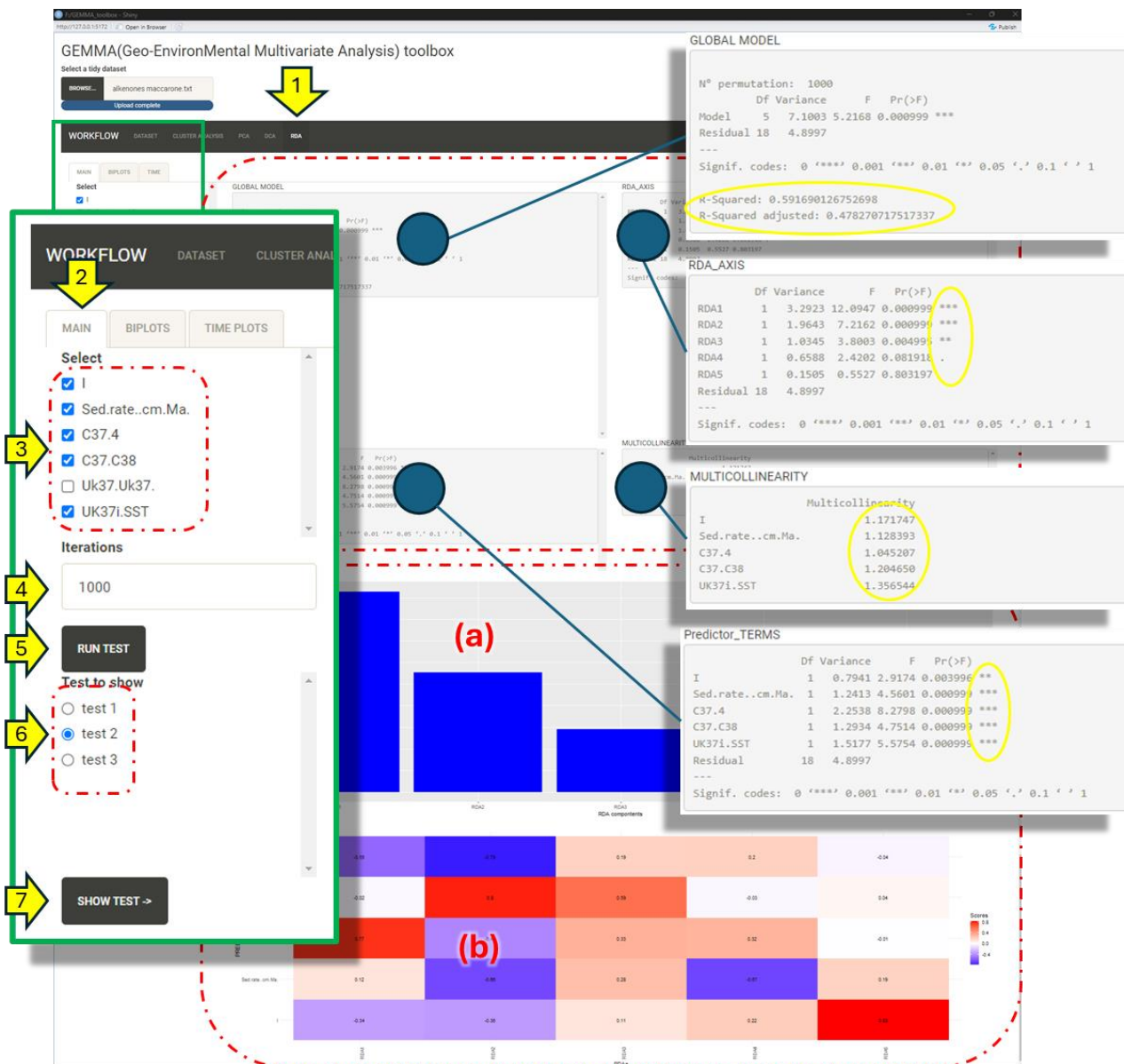**Note:** "BIPLOT" and "TIME" sub-plot workflows follow the same structure as PCA/DCA.

Figure 10