User manual – ***trident***

An R package for dental microwear texture measurement and variable analysis

Ghislain Thiery[1, 2], Arthur Francisco[3], Margot Louail[1], Gildas Merceron[1]

1. Laboratoire PALEVOPRIM, UMR 7262, CNRS & Université de Poitiers, France
2. Center for Evolutionary Origins of Human Behavior EHUB, Kyoto University Museum, Japan
3. Institut Pprime, UPR CNRS 3346, Université de Poitiers, France

---

Table of contents

Introduction....3
What is trident?....3
Installation of trident....4
License and copyright....5
Why making an app?....5
General presentation....5
Getting help....6
1. Batch analysis....7
1.1. Selecting .SUR files....7
1.2. Peak removal....8
1.3. Select parameters....10
1.3.a Complexity....10
1.3.b Height....10
1.3.c Spatial parameters....11
1.3.d Topology parameters....12
1.4. Heterogeneity....12
1.5. Adding factors from file names....15
1.6. Launch, save and export....15
1.6.a. Launch....15
1.5.b. Save....16
1.5.c. Export to R....17
2. Dataset....18
2.1. Selecting a dataset....18
2.2. Transforming the data....19
2.2.a Log transformation....19
2.2.b Box Cox transformation....19
2.3. Save and export....21
2.3.a. Save....21
2.3.b. Export to R....21
3. Variables....22
3.1. Selecting a factor....22
3.2. Correlation test....22
3.3. Multicheck....24
3.4. Top3....25
3.5. Rank by....27
3.6. Save and export....29
3.6.a. Save....29
3.6.b. Export to R....29
4. Graphics....30
4.1. Univariate....30

4.2. Multivariate....31
4.2.a. PCA....31
4.2.b. Use supplementary variables....33
4.2.c. Add supplementary individuals....33
4.3. Saving graphs....36
Conclusions....37
Concluding remarks....37
Acknowledgments....38
References....39

Introduction

What is trident?

In its current form, ***trident*** is an R package specialized in dental microwear texture analysis (DMTA). In creative hands however, it can be used in many other ways, from data visualization of multivariate analysis to classification of variables according to their discriminative power (relative to a user-defined factor, such as species, diet, locality...). It was designed with and for the R environment (R Core Team, 2021). In addition to its core functions, ***trident*** provides a complete, user-friendly app. If you have trouble using some of its functionalities, we hope that the present guide will help you master it to its full extent.

Before it was a program, TRIDENT was an interdisciplinary project started and led by Gildas Merceron in 2013, which came out in 2017 as a second project named DIET-Scratches, both funded by the French Agency of Research (https://anr.fr). Combining resources in agronomy (Mourier Farm Station, INRAe GenESI), tribology (Pprime Institute) and paleontology (PALEVOPRIM), the aim was to ultimately investigate the ecology of extinct mammals, including ungulates and primates, through dietary behavior analysis. The projects were based on (i) a controlled feeding experiment with domestic sheep and pigs, (ii) an application on large samples of wild and extant communities of mammals with known dietary ecology to serve as models. More information can be found on the projects' website: http://anr-trident.prd.fr/en/

During the early stages of the TRIDENT project, a computational routine combining R, Python and Fortran was developed by Arthur Francisco for calculating DMTA parameters and comparing them (Francisco, Brunetière and Merceron, 2018; Francisco *et al.*, 2018). The ***trident*** software is an attempt to (1) make this routine friendlier for researchers who are not fluent in R or Python and (2)

increase the accessibility to the code by making it openly available on an online repository (Github). Additional functions, such as an agile framework for Principal Components Analysis, were also added during the 3-year long development of this package. We hope you find them useful!

Installation of trident

The .tar file containing the source package can be downloaded from a Github repository at the following address:

https://github.com/nialsiG/trident/blob/main/InstallTrident_1.3.8.zip

Before starting the installation of ***trident***, you should:

1. Check that your version of R is >= 4.0.0
2. Check that Rtools40 has been installed (you can get it at the following link: https://cran.r-project.org/bin/windows/Rtools/)

Note: ***trident*** was developed in the Windows environment, and most of its functionality were designed for Windows. They might work in other systems, but we cannot guarantee that all functionalities will work as expected. While it may be limiting for some users, at the time Windows is the most widespread OS. Besides, functions from the R package should work in most existing environments.

Before installing the source package, please also check that the following packages have been installed:

```
install.packages(c("car", "DescTools", "doSNOW", "dplyr", "DT",
"factoextra", "FactoMineR", "foreach", "ggpubr", "ggplot2", "MASS",
"nortest", "parallel", "plyr", "shiny", "shinyjs", "shinyFiles", "snow",
"stats", "stringr", "utils"))
```

Once it is done, you can install ***trident*** by running the following code, then choosing the 'trident_1.3.8.tar.gz' file in the selection window:

```
install.packages(choose.files(), repos = NULL, type = "source")
```

Alternatively, you can follow the instructions in the readme.txt file.

License and copyright

Why making an app?

The previous versions of this software have been used in several publications (Francisco, Brunetière and Merceron, 2018; Francisco *et al.*, 2018; Louail *et al.*, 2021; Merceron *et al.*, 2021). However, two limitations were identified at that time. The first one concerned the repeatability of the measures: the code was not accessible to a wide range of users, and it was not open. The second limitation was the ease of use. Previous versions used a combination of Fortran, Python and R to collect the data. Furthermore, fine-tuning the content of data collection and analysis was only possible for people familiar with those languages.

Using an application solves both issues and makes measuring and analyzing DMTA data easier, faster, and a lot more intuitive for beginners.

General presentation

Launching the ***trident*** app is as simple as typing the following line directly in the R console:

```
trident::trident.app()
```

Once the interface opens, you should see something similar to Figure 1.1. The interface is composed of four tabs. Each tab covers a single major functionality of the software: Tab 1 is for

Dental Microwear Texture computations (1. Batch analysis); Tab 2 is for handling datasets (2. Dataset); Tab 3 is for variables description and classification (3. Variables); and finally, Tab 4 is for graphical visualization (4. Graphics).

Getting help

You can ask questions or push requests directly on ***trident***'s github page:

https://github.com/nialsiG/trident

You may also find there an up-to-date list of the known bugs. Please do not hesitate to amend it if you find an unnoticed bug!

1. Batch analysis

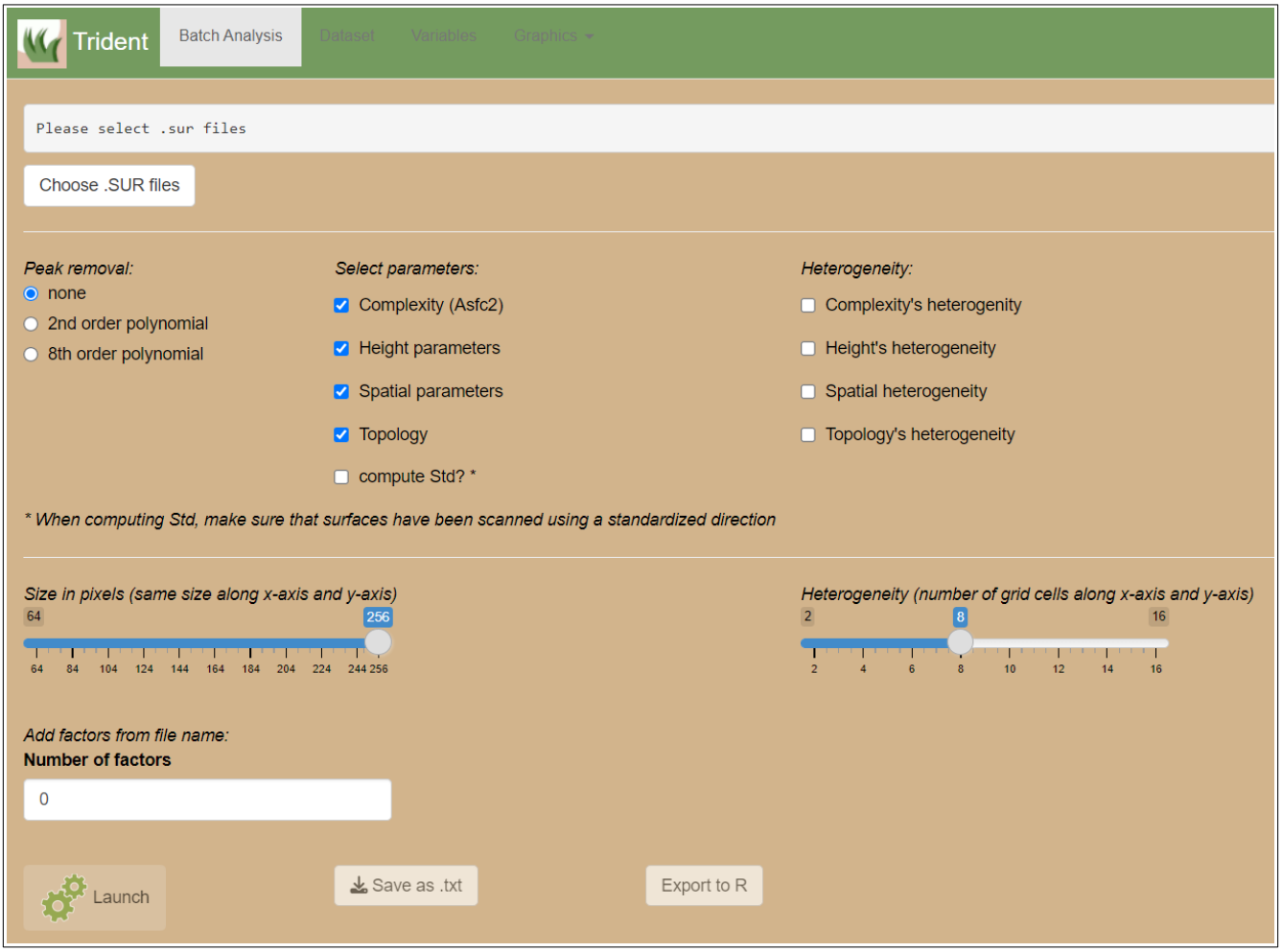


Figure 1.1. The interface of ***trident*** under the Batch analysis tab.

1.1. Selecting .SUR files

At the program start, the Launch button is disabled. In order to launch the batch analysis, the user must select surface files first. Please make sure that **surface files are in the .SUR file format**. Select surfaces by clicking the Choose .SUR files button, then navigate to the file where the surfaces are located (Fig. 1.1). You can select multiple surfaces using the Shift key.

Once surfaces are selected, the Launch button is re-enabled: you can start the batch analysis right away, although it is a good idea to check the configuration of the analysis before (see below).

Note: The routine may have trouble managing both peak removal (surface pre-treatment) and DMTA computation (see below). We advise you to first pre-treat the surfaces to generate a set of new pre-treated .SUR files, on which you can run the batch analysis (with the peak-removal checkbox unchecked!).

1.2. Peak removal

One of the risks when investigating surface roughness is that the "large-scale" tooth surface geometry might conceal the "small-scale" variation of height attributable to microwear. The ***trident*** program has an algorithm to solve this issue. Following a surface treatment procedure presented on previous studies (Francisco, Brunetière and Merceron, 2018; Francisco *et al.*, 2018), the primary surface S1 is first numerically cleaned of any abnormal peaks. Then, considering the large-scale tooth surface geometry as a second order polynomial (PS2), the latter is subtracted via a least square approximation. Removing PS2 could be deemed sufficient due to the small area measured (0.2 mm), as dental facets are concave or convex with a single preferred axis. However, in order to enhance the finest scale roughness, the subtraction can instead be applied with an eighth order polynomial (PS8).

This can be done by selecting the radiobutton corresponding to the desired level of peak removal, PS2 or PS8 (Fig. 1.2). Once the computation is done, the treated surface will be saved at the same location as the original surface file, its name showing the suffix "_X1_002_Y1_002" or "_X1_008_Y1_008", respectively. Note that the order of the polynomial applies to each direction.

Note: Sometimes the interface might close during the process, but the surfaces have been treated and saved.

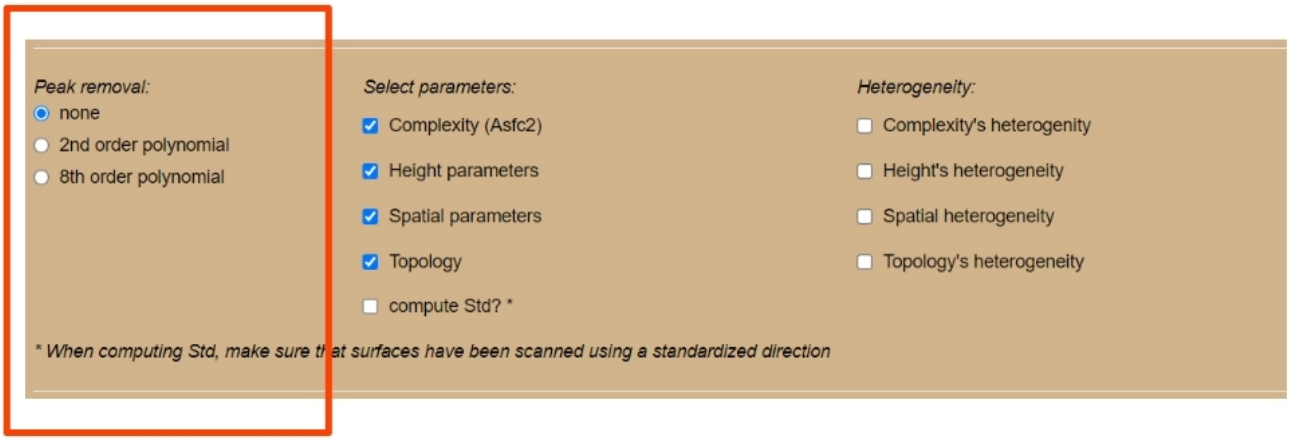


Figure 1.2. Location of the “Peak removal” menu. Note the 3 radiobuttons, with default selection on “None”.

1.3. Select parameters

Next, you must select which parameters you want to measure by checking their respective boxes (Fig. 1.3). As of version 1.3.0, the following DMTA parameters can be measured (see also Francisco, Brunetière and Merceron, 2018):

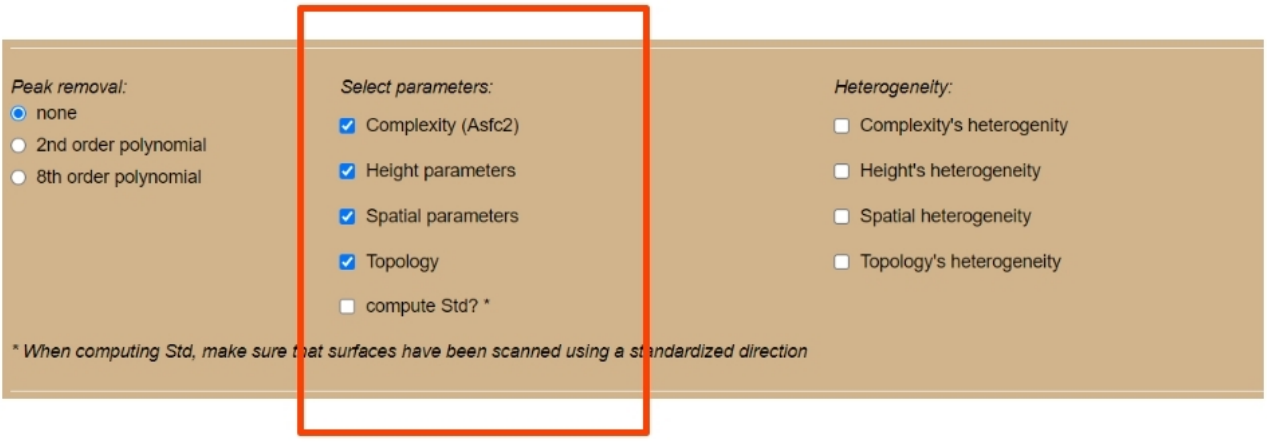


Figure 1.3. Location of the “Select parameters” menu. All parameter categories except Std are checked by default.

1.3.a Complexity

The only available complexity parameter (Asfc2; Table 1) depicts the scale-dependent density of structures over the microwear surface. It is also an estimate of surface roughness.

Table 1. Complexity parameters

Parameter	Significance
Asfc2	Area scale fractal complexity (modified after Asfc in Scott *et al.*, 2006; see also supplementary materials in Francisco, Brunetière and Merceron, 2018)

1.3.b Height

Height parameters (Table 2) describe the central variables of absolute heights (Sa, Sm, Smd), as well as their statistical dispersion (Sp, Sq, Sv), their variation (Ssk, Sku). Height parameters also includes direct estimates of relief such as the relative area (Sdar, which is the ratio between the developed surface area of a 3d object over the projected 2d area of the same object).

Table 2. Height parameters (* ISO 25178)

Parameter	Significance
Sa	Arithmetic mean of the absolute of the heights (*)
Sp	Absolute of the largest height (*)
Sq	Height standard deviation (*)
Sv	Absolute of the smallest height (*)
Ssk	Height skewness (*)
Sku	Height kurtosis (*)
Sdar	Relative area (developed area/projected area)
Sm	Mean height (0 for the whole surface, but non-zero for its samples)
Smd	Median height

1.3.c Spatial parameters

Spatial parameters (Table 3) describe the distribution of structures in space. Some of them are akin to the heterogeneity of Asfc (HAsfc; Scott, Teaford and Ungar, 2012), for instance Rmax, Sal or Stri (= Str^{-1}). Other parameters such as b.sl, s.sl and r.sl are based on the slope of the autocorrelation function fACF(tx,ty), which quantifies the self similarity of a surface shifted along a (tx, ty) vector (Francisco *et al.*, 2018). These are again estimates of spatial heterogeneity.

Lastly, texture direction (Std) is an estimate of the surface dominant inclination; in practice however, it is strongly impacted by the positioning of specimens whence scanning them. It is not recommended to use texture direction without a controlled protocol for surface orientation.

Table 3. Spatial parameters. (* ISO 25178)

Parameter	Significance
Rmax	Semi-major axis of the fACF ellipsis
Sal	Semi-minor axis of the fACF ellipsis (*)
Std	Texture direction (*)
Stri = Str $^{-1}$	Rmax/Sal ratio (derived from Str(*))
b.sl	Highest slope of fACF at the distance rs from the origin
r.sl	b.sl/s.sl ratio
s.sl	Smallest slope of fACF at the distance rs from the origin

1.3.d Topology parameters

Lastly, topology parameters (Table 4) measure the proportion of the surface above standard height thresholds. This proportion is either a relative surface area (Sk1, Sk2), the median of the relative areas of isolated cells above a threshold (Smc1, Smc2) or the number of such cells (Snb1, Snb2). The final parameter Sh is also a percentage, but instead of height thresholds it uses texture direction thresholds.

Table 4. Topological parameters.

Parameter	Significance
Sk1, Sk2	Relative area of the surface above h1 and h2 respectively
Smc1, Smc2	Median relative area of the cells with heights exceeding h1 and h2 respectively
Snb1, Snb2	Number of cells with heights exceeding h1 and h2 respectively
Sh	Percentage of quasi-horizontal faces (normal within a 4 cone)

1.4. Heterogeneity

The heterogeneity of a (dental) surface is related to the spatial distribution of its features: for instance, a single pit in the enamel implies more heterogeneity than several pits uniformly distributed through the enamel surface (Scott *et al.*, 2006). Following Francisco et al. (2018),

trident uses a fast and intuitive approach for estimating heterogeneity: the surface is divided in nxn grid cells, and DMTA variables are computed for each grid cell. Then, descriptive and distribution statistics are computed from the nxn values obtained for each variable (Table 5).

Table 5. Heterogeneity statistics computed for every DMTA variable. (*) *Warning: we advise you not to consider these values in forthcoming analyses, as it could target an aberrant topographic relief*

Parameter	Significance
min	Minimal value (*)
max	Maximal value (*)
std	Standard deviation
mean	Arithmetic mean
med	Median
fst.05	5th percentile
lst.05	95th percentile
min.05	Mean of values under the 5th percentile
max.05	Mean of values above the 95th percentile
fst.25	1st quartile
lst.25	3rd quartile
min.25	Mean of values under the 1st quartile
max.25	Mean of values above the 3rd quartile
skw	Skewness of the histogram of distribution
kurt	Kurtosis of the histogram of distribution

To compute heterogeneity variables, you have to check the box corresponding to a parameter family in the “Heterogeneity” menu (Fig. 1.4.1). You can change the size of the surface (in pixels) and the number of grid cells that will be used to compute heterogeneity using sliders (Fig. 1.4.2). Please note that the number of pixels in the X and Y directions are the same. This is also the case for the number of grid cells: this ensures that the total number of grid cells will always be a square number (8*8 = 64, 10*10 = 100, 12*12 = 144, 16*16 = 256 etc.).

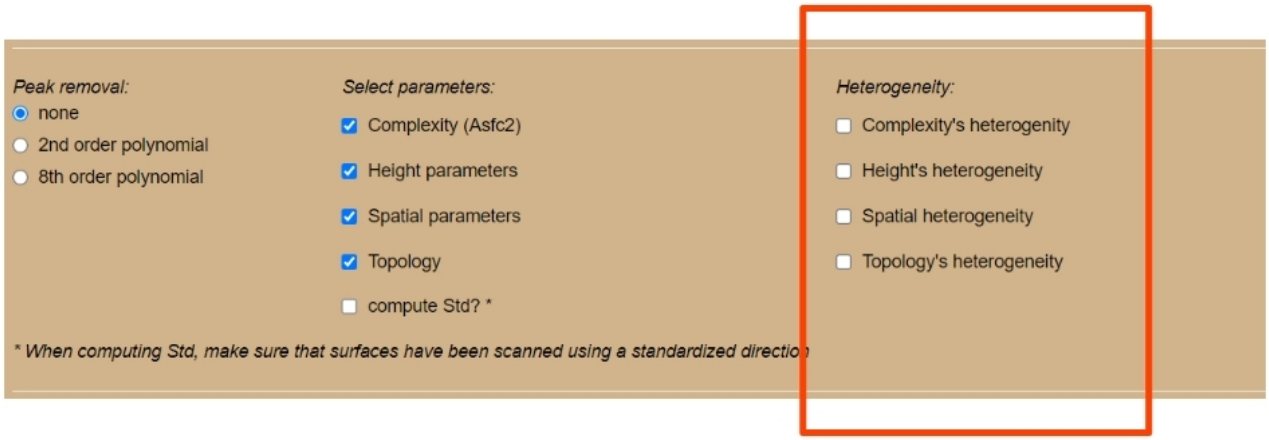


Figure 1.4.1. Location of the “Heterogeneity” menu.

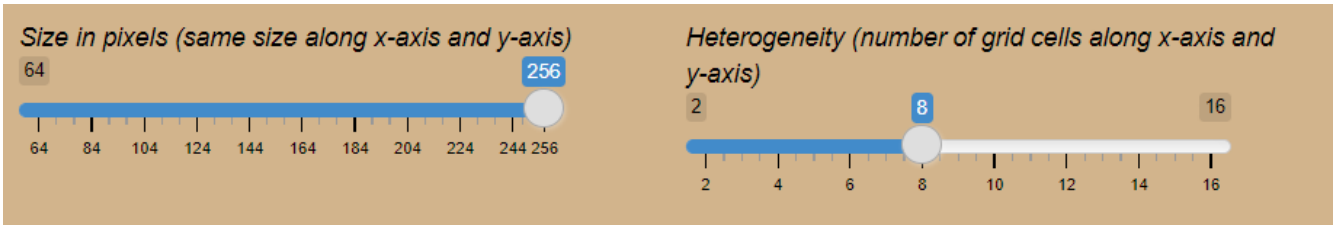


Figure 1.4.2. Sliding cursors for selecting the size of the surface to be considered (left cursor, unit = pixels) and the number of cells composing the grid for heterogeneity (right cursor, unit = number of cells)

Note: The calculation of heterogeneity of Asfc2 is time-consuming and requires a significant amount of memory.

1.5. Adding factors from file names

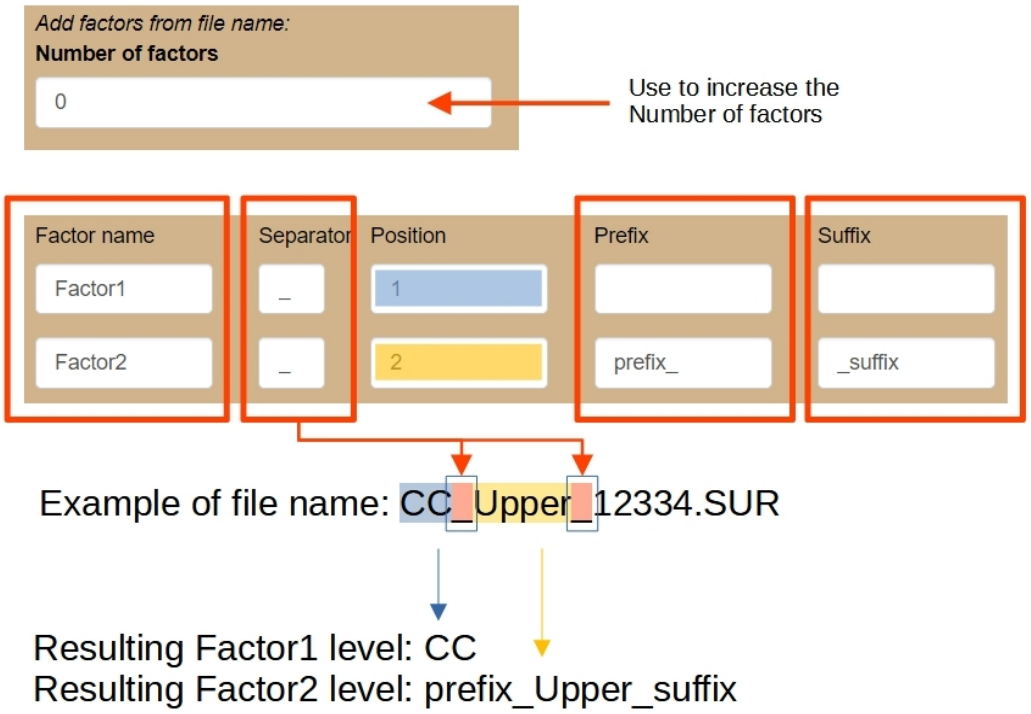


Figure 1.5. Input for adding factors from filename.

Note: The number of characters used as separators should remain constant for all the filenames of the surfaces being processed in the same batch analysis. If you find the following error:

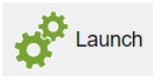
Warning in matrix(unlist(Strings), nrow = length(Strings[[1]])):

data length [x] is not a sub-multiple or multiple of the number of rows [y]

...you should check if the number of separators is the same for all filenames. Be especially wary of very similar characters e.g. "_" and "-".

1.6. Launch, save and export

1.6.a. Launch

As soon as you have selected a .SUR file, the 

button will be enabled. Once you are satisfied with the configuration of your analysis, you can click it. You should note:

- That computation time is dependent on several factors, including the number of specimens, the size of the surfaces, the number of grid cells for computing heterogeneity, whether you perform height polynomial removal or not... Also, the computation of Asfc2 is significantly longer than that of all height variables, for instance. This likely comes from the fractal nature of area scale fractal complexity.
- That Std central and heterogeneity variables are disabled by default. This comes from the observation that Std is a non-sense if the user did not use a standardized orientation protocol. If your protocol takes that matter into account, then do not hesitate to able it.
- **That batch analysis uses parallel computing and will requisite all of your machine's cores, except one. So make sure you do not need your computer for something else!**

Once the computation is over, a table akin to Fig. 1.6.1 should be displayed (if you cannot see it, try scrolling down)

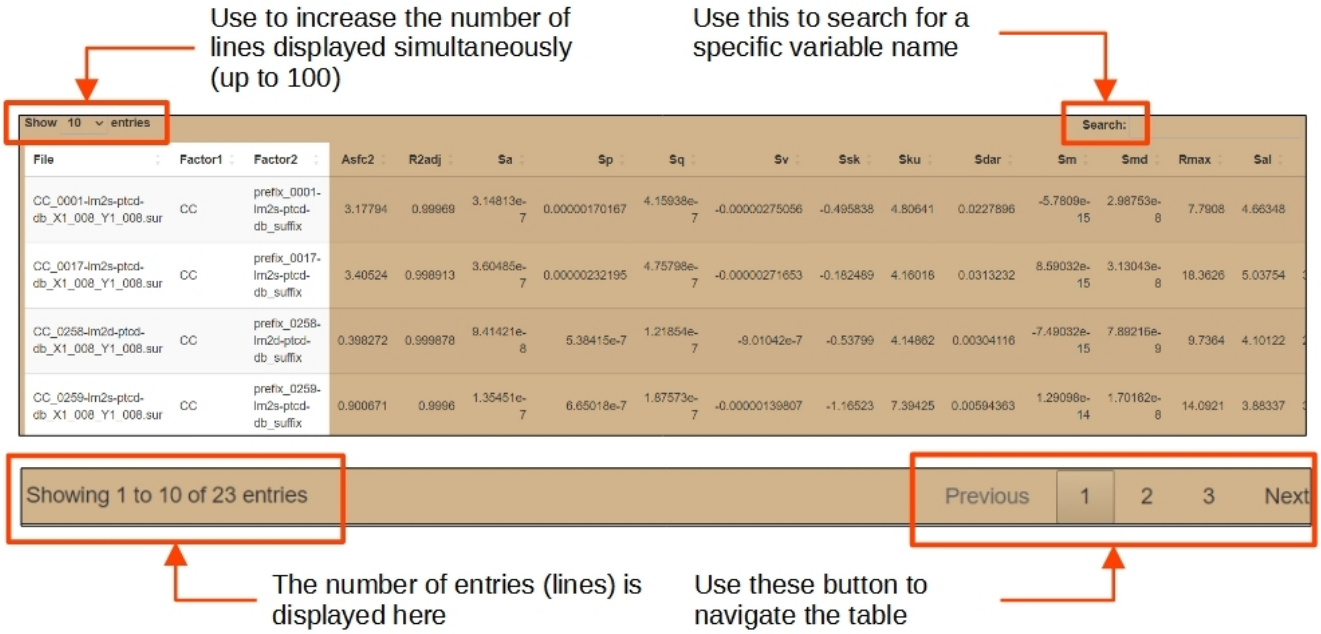
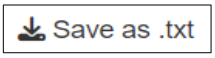


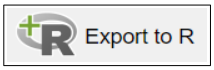
Figure 1.6.1. Result of a batch analysis

1.5.b. Save

The table can (and should!) be saved in your computer using the 

button. Please note that you need to save a table if you want to use it for other analyses involving ***trident***. At the moment, the only available file format is .txt, with tab-separated columns.

1.5.c. Export to R

While you may want to continue using the ***trident*** interface, you may also want to export the table to your R session by clicking the 

button. You will be prompted to give a name to your R object: **make sure that this name has not been taken yet, or the object will be replaced!** If you are using Rstudio, be also wary that your object might not appear immediately into your "Environment" panel – you may need to refresh it (Fig. 1.6.2). Finally, since the interface uses the current R session, you need to close the interface before using functions on your R object.

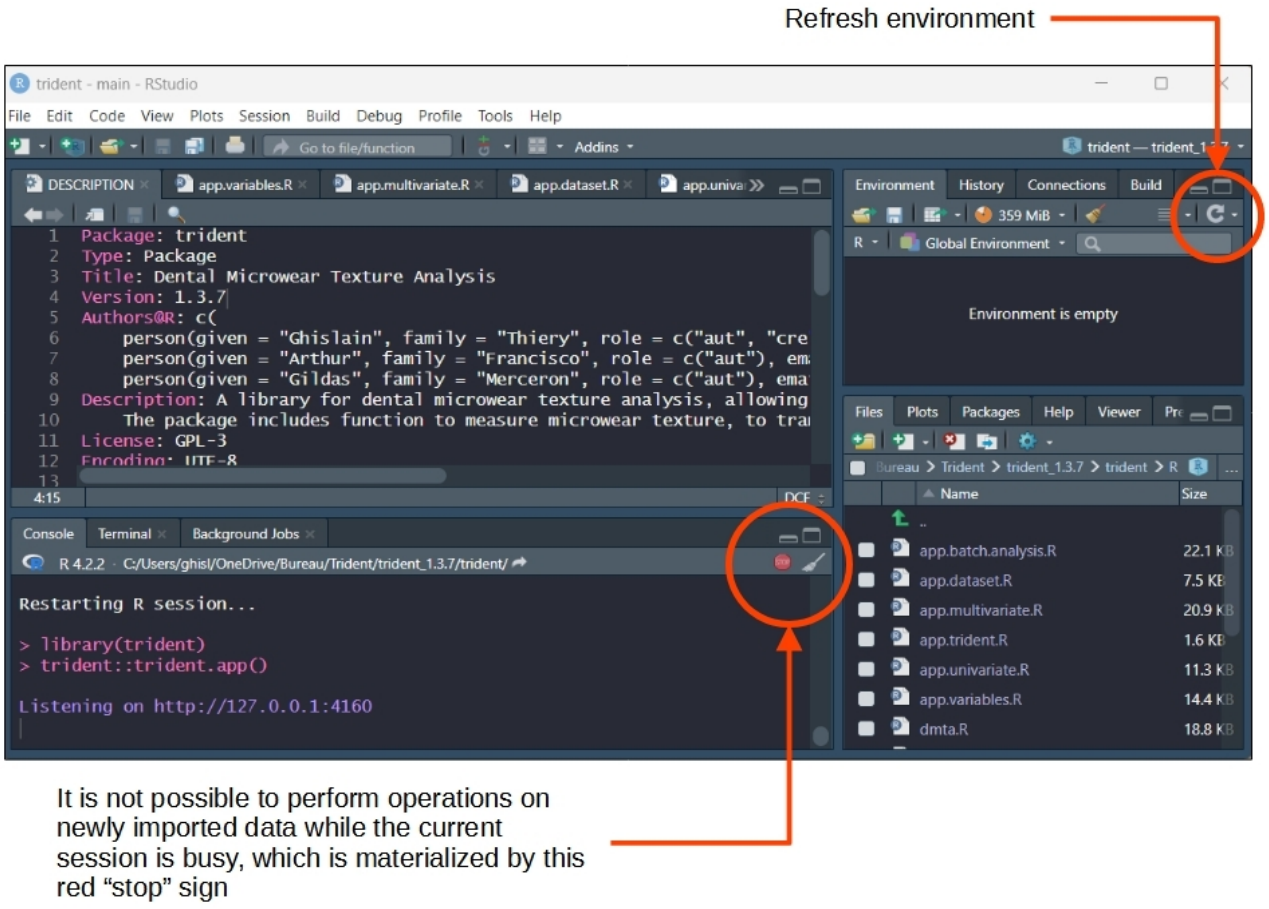


Figure 1.6.2. Exporting data to RStudio might require you to refresh the "Environment" panel

2. Dataset

2.1. Selecting a dataset

By default, the buttons for transforming, exporting and saving data, but also the functionalities of the other tabs, are disabled until the user loads a dataset. To do so, first click the button, and select a tab-separated .txt file.

Note: The default is set to tab-separated txt files, but you can change it to space-separated. Be wary that trying to import the wrong format file might crash ***trident***. If that happens, stop the activity of the current R session

Before importing the data, you can decide if you want to remove all the columns which contain any number of NA values; some latter functionalities (rank by, top3, PCA…) are indeed impacted by the presence of NA values. To do so, simply check the “Remove NA values” box (Fig. 2.1).

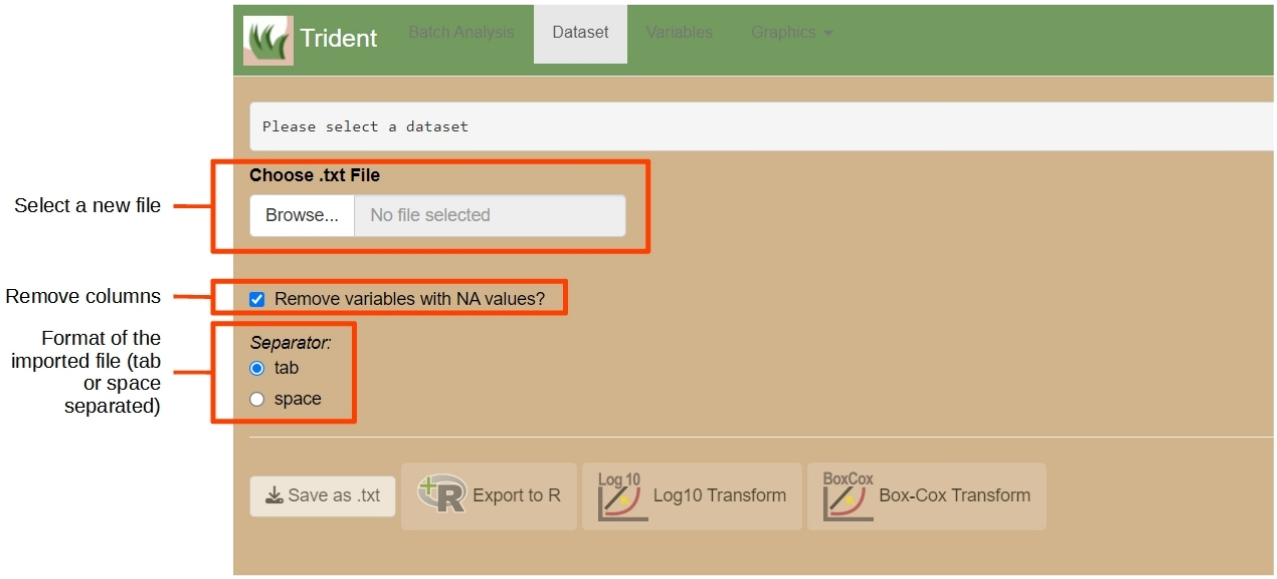


Figure 2.1. Selecting a dataset.

2.2. Transforming the data

2.2.a Log transformation

Trident allows you to perform base 10 log transformations. To do so, click on the 

button. After a short wait, the data should be transformed.

Note: After transformation, the button is disabled and the name of every numeric variable will be added the suffix ".log10". This must be taken into consideration for further analysis that require column names to be exactly the same e.g., adding supplementary individuals to the PCA.

2.2.b Box Cox transformation

The Box Cox transformation is a transformation of a non-normal dependent variables into a normal shape. Normality is an important assumption for many functionalities of ***trident***, so if your data is not normally distributed, you might want to apply a Box Cox transformation.

To perform a Box Cox transformation, click on the 

button. After a short wait, the data should be transformed. You may have noticed that this button is disabled by default. To re-enable it you need to select a factor variable: just click anywhere in the column, and it should turn blue, indicating that it has been selected (Fig. 2.2). Deselect by clicking in the column again.

Note: After transformation, the button is disabled and the name of every numeric variable will be added the suffix ".boxcox". This must be taken into consideration for further analysis that require column names to be exactly the same e.g., adding supplementary individuals to the PCA.

Figure 2.2. Activation of the Box Cox transformation button.

2.3. Save and export

2.3.a. Save

If you want to save the data after it was transformed, the table can be saved in your computer using the Save as .txt button. At the moment, the only available file format is .txt, with tab-separated columns.

2.3.b. Export to R

While you may want to continue using the ***trident*** interface, you may also want to export the table to your R session by clicking the Export to R button. You will be prompted to give a name to your R object: **make sure that this name has not been taken yet, or the object will be replaced!** If you are using Rstudio, be also wary that your object might not appear immediately into your "Environment" panel – you may need to refresh it (Fig. 1.6.2). Finally, since the interface uses the current R session, you need to close the interface before using functions on your R object.

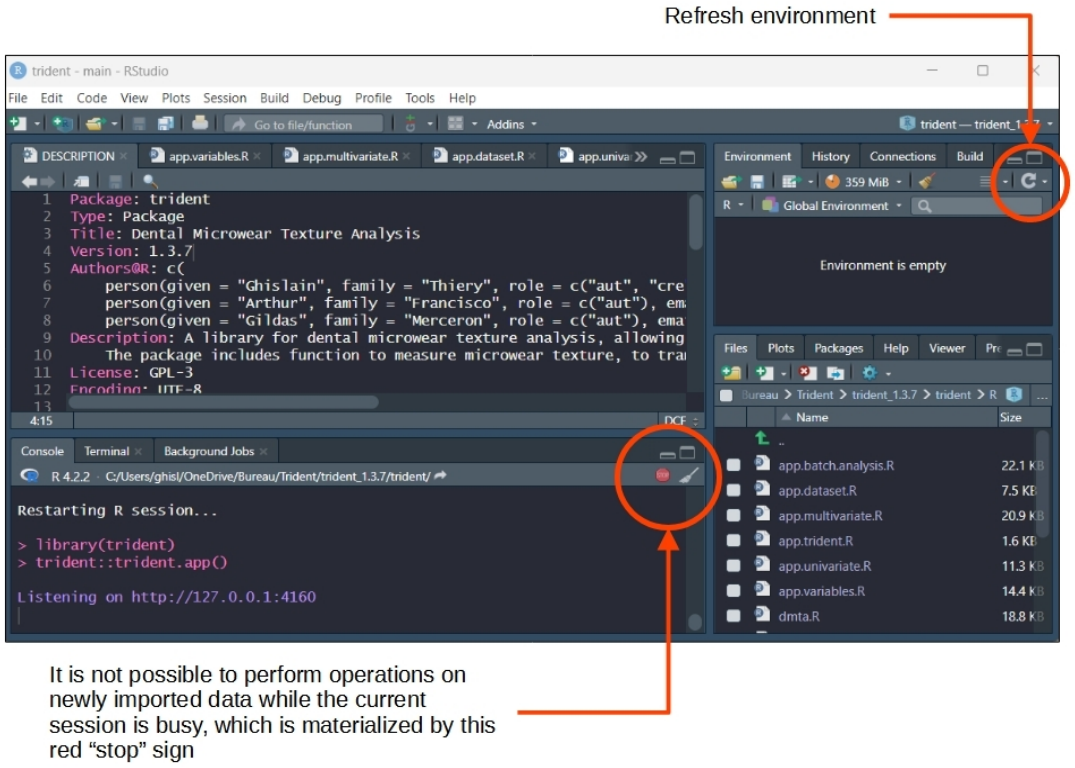


Figure 2.3. Exporting data to RStudio might require you to refresh the "Environment" panel

3. Variables

3.1. Selecting a factor

The majority of functions in the Variables and Graphics tabs will require at least one factor variable. When importing a dataset to ***trident***, any column made of character strings will be considered a factor variable by default. Make sure that your factor variable does not have too much levels for what you intend to do: for instance, every time you add one level to the n levels of a factor, the number of pairs to compare with the Top3 function increases by n, which can dramatically extend computation time, and more than 8 levels results in a very long computation time. Similarly, **graphics in *trident* are limited to 24 types of points**, so you cannot make graphs if you have 25 or more levels.

Note: this behavior is different from the general behavior in R >4.0.0, where character strings are NOT considered as factors by default. It is however much simpler for the user this way, as it removes the need for a “transforming variable to factor” step.

If your dataset comes directly from ***trident*** (1.6. Launch, save and export), the factor variables that you created from file name should be there. You should also have a factor variable entitled File and displaying the file names. **You should not use the “File” variable as factor**, even if you have a small number of files: your factor would have only one point for each level, preventing from computing dispersion statistics etc.

To select a factor, simply click on the table line which corresponds (Fig. 3.1).

3.2. Correlation test

Click on the button to perform Pearson’s linear correlation test. The results will be displayed in a table below the button’s area, along with their level of significance (NS Non significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$). For more details on the actual computation, see the *cor.table* function from the *picante* package (Kembel *et al.*, 2010). The table can be saved and exported to R.

Figure 3.1. Selecting a factor

3.3. Multicheck

This function will perform several statistical tests in order to check whether your data show a normal distribution (Shapiro-Wilk's test, skewness ratio), the homogeneity of variables i.e. homoscedasticity (Bartlett's and Levene's tests, variance ratio) and their ability to tell categories apart i.e. discriminant ability (ANOVA and Kruskal-Wallis' test). Note that this analysis require a factor variable (3.1. Selecting a factor).

Click on the 

button and after some time, a table similar to Fig. 3.3. will be displayed. **Be patient: this test can be long, especially for factors with many levels.** Like other datasets, it can be saved or exported to R.

Show 10 entries

variable	anderson_p.value	anova_p.value	bartlett_p.value	kruskall_p.valu	is.normal	is.homoscedastic	is.discriminant
Asfc2	0.000020080444947492 5	0.140718478337345	0.000010394597029813	0.42565792082	nearly	FALSE	FALSE
Sa	0.00310300023535258	0.0260095269224494	0.254124996777665	0.02413724662	nearly	TRUE	TRUE
Sp	0.000218929010834603	0.224145004827429	0.26728106256373	0.24607698241	nearly	TRUE	FALSE
Sq	0.00445877564560245	0.0620528605122208	0.399521212755476	0.03815398751	nearly	TRUE	TRUE
Sv	8.62523034917366e-11	0.686236066825681	4.57871923303001e-7	0.70279910311	nearly	nearly	FALSE
Ssk	7.21890358550148e-15	0.112823561542172	1.88288172249644e-12	0.08063079444	nearly	nearly	FALSE
Sku	3.7e-24	0.209552282137263	9.06119744914115e-26	0.13152379091	nearly	nearly	FALSE
Sdar	1.04542797207487e-7	0.306660482949642	5.14375251163321e-9	0.65650993191	nearly	FALSE	FALSE
Sm	0.0000066822991526371	0.000707664745831858	4.60613560907238e-18	0.02347077331	nearly	FALSE	FALSE
Smd	0.00255113527185852	0.547937135330039	0.136318686424926	0.55728828056	nearly	TRUE	FALSE

Showing 1 to 10 of 23 entries

Figure 3.3. A multicheck table, with the last three columns indicating whether variables are normally distributed, have homogeneous variances, and can separate the categories defined by the selected factor; "nearly" indicates cases when a question (for example is data normally distributed?) was given different answers by the tests.

3.4. Top3

For a given factor, the first method of classification in ***trident*** is used to extract the three best discriminating variables **for each pair of categories**. This is quite important, as the number of pairs will increase dramatically with the number of levels in your factor. For example, imagine you have 4 species in your sample, then you will extract 3 variables for 3+2+1 pairs, for a total of 18 variables. If you had 5 species, you would have 4+3+2+1 pairs, thus 30 variables; for 6 species, 5+4+3+2+1 pairs, thus 45 variables; and so on.

Following Francisco et al. (2018), this methods automatically transforms data using the Box Cox transformation (see 2.2. Transforming the data). Then, a multicheck routine is performed (see 3.3. Multicheck), an ANOVA is performed and non-discriminating variables are removed. Afterwards, for each pair, discriminant variables are ranked by ascending p value of Tukey's honestly significant difference (HSD) test.

The function can be launched by clicking the 

button. Note that this button requires you to select a factor in order to be re-enabled (3.1. Selecting a factor). After a few moments (computation time will be proportional to the number of variables, individuals and levels in the factor), a window with the top 3 variables for each pair, together with various information such as multicheck results, opens. You can save this table or export it to R.

Note: Each variable might be present several times in the list, as long as it is among the best 3 variables for several pairs of factor levels. For this reason, the Top 3 function is not used to classify the variables in the Graphics tab (see 4. Graphics).

Note: Only variables with a significant p value ($\alpha = 0.05$) will be kept. However, if there are less than 2 discriminant variables in total, a popup window will display a message and the top 3 procedure will be stopped.

Show 10 entries Search:

variable	group	f.stat	aov.p.value	k	k.p.value	hsd.CC.vs.CE	hsd.CC.vs.OG	hsd.CC.vs.RR
s.sl.boxcox	CC vs. CE	2.82591750986781	0.0502351927549268	8.14052719036911	0.0431949794548164	0.151055737715551	0.99993981853905	0.999086315360263
Sal.boxcox	CC vs. CE	3.58469972344682	0.021196617284481	9.4137945510264	0.0242663485685213	0.319994759459205	0.937225573299364	0.636762893418094
Snb1.boxcox	CC vs. CE	2.92946916878776	0.0441870229110728	6.24372006107049	0.100336387976272	0.718237060517969	0.368758376242149	0.031763898817077
s.sl.boxcox.1	CC vs. OG	2.82591750986781	0.0502351927549268	8.14052719036911	0.0431949794548164	0.151055737715551	0.99993981853905	0.999086315360263
Sal.boxcox.1	CC vs. OG	3.58469972344682	0.021196617284481	9.4137945510264	0.0242663485685213	0.319994759459205	0.937225573299364	0.636762893418094
Snb1.boxcox.1	CC vs. OG	2.92946916878776	0.0441870229110728	6.24372006107049	0.100336387976272	0.718237060517969	0.368758376242149	0.031763898817077
Rmax.boxcox	CC vs. RR	5.18396498852001	0.00379076127585755	15.2368154912969	0.00162505947601302	0.736330144821867	0.0381925168515302	0.424139901443761
Stri.boxcox	CC vs. RR	2.84409770578415	0.0488636288392692	9.55166373444359	0.0227880492831564	0.99837176092687	0.0456594530855713	0.925032752612902
r.sl.boxcox	CC vs. RR	2.98557517862969	0.0416750175697381	8.64534985046294	0.0343973875736912	0.998215067097656	0.05224074178954	0.619918789233913
Rmax.boxcox.1	CE vs. OG	5.18396498852001	0.00379076127585755	15.2368154912969	0.00162505947601302	0.736330144821867	0.0381925168515302	0.424139901443761

Showing 1 to 10 of 18 entries Previous 1 2 Next

Figure 3.4. An example of a Top3 table. Observe how the content of the ‘group’ column (e.g. CC vs. CE, CC vs. OG...) is repeated on three rows everytime.

3.5. Rank by

Sometimes, the Top 3 variables of each pair will not capture the information you were looking for. Alternatively, you can rank variables using four other filters (Fig. 4.5):

1. Rank by the p value of an ANOVA. Option to remove non-discriminant variables.
2. Rank by the p value of a Kruskal-Wallis ranked-sum analysis. Option to remove non-discriminant variables.
3. Rank by the mean of all p values of the post-hoc analysis of an ANOVA (only p values under 0.05 are used when computing this mean). Options to use Tukey's Honestly Significant Difference test (HSD) or Fisher's Least Significant Difference test (LSD); to remove non-discriminant variables; to use the geometric mean of the p values (instead of the arithmetic mean); or to arrange the ranked data by number of discriminated groups first, then by ascending p values.
4. Rank by the p value of the post-hoc analysis of an ANOVA for a selected pair of levels from the selected factor. Options to use Tukey's Honestly Significant Difference test (HSD) or Fisher's Least Significant Difference test (LSD); to remove the non-discriminant variables; to arrange the ranked data by number of discriminated groups first; and of course to select the preferred pair of levels.

Note: if you check the 'Remove non-discriminant variables' box, only variables with a significant p value ($\alpha = 0.05$) will be kept. However, if there are less than 3 discriminant variables, a popup window will display a message and the ranking procedure will be stopped.

For all four methods, the function can be launched by clicking the 

button. Be wary that just like the Multicheck and Top3 buttons, this button is disabled by default and can be re-enabled by selecting a factor (3.1. Selecting a factor). After a few instants (computation time will be proportional to the number of variables, individuals and levels in the factor), the table with the ranked variables, together with various information such as multicheck results, will display. You can save this table or export it to R.

Note: The order of variables ranked using the Rank by button will transfer to the list of variables in the graphics tab (see 4. Graphics). If you checked the 'Remove non-discriminant variables' box, those will also be removed from the graphics tab.

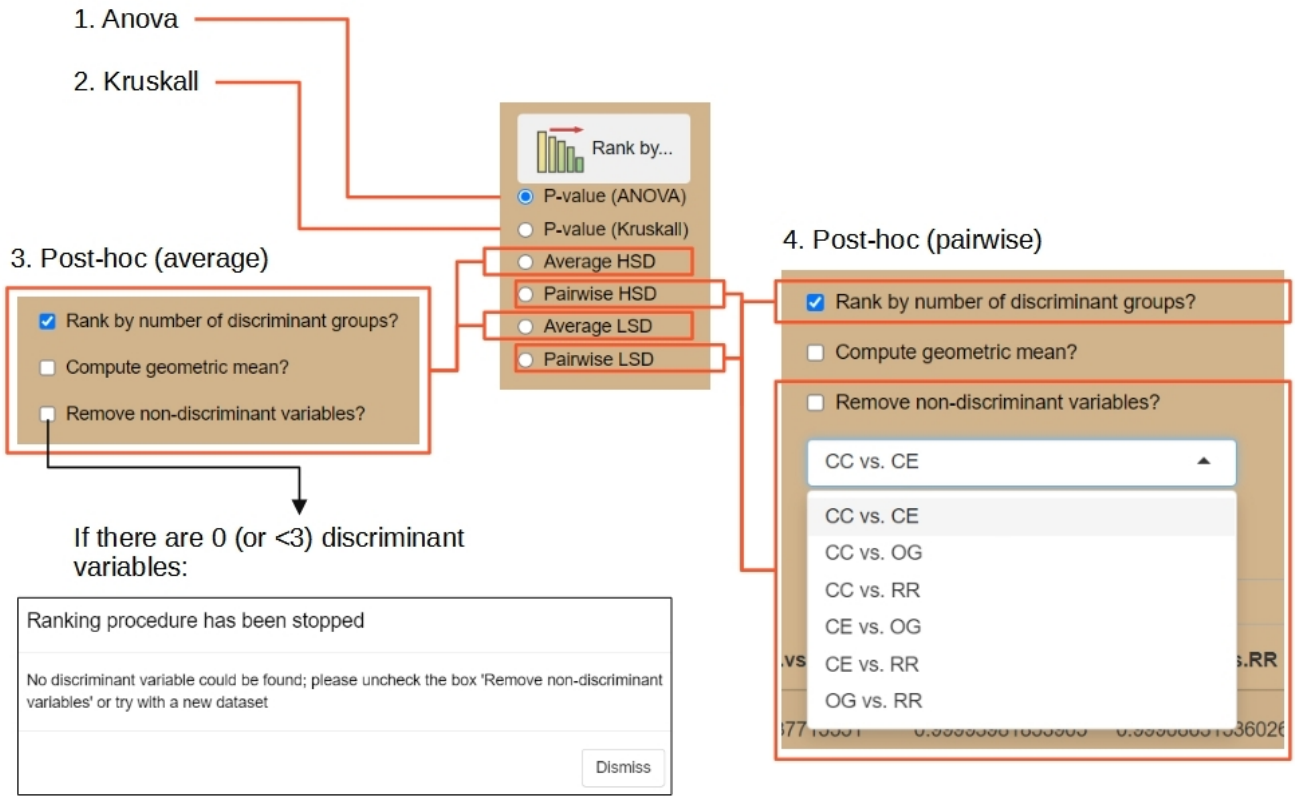


Figure 3.5. The different ranking procedures.

3.6. Save and export

3.6.a. Save

If you want to save the data after it was transformed, the table can be saved in your computer using the Save as .txt button. At the moment, the only available file format is .txt, with tab-separated columns.

3.6.b. Export to R

The table can also be exported to your R session by clicking the Export to R button. You will be prompted to give a name to your R object: **make sure that this name has not been taken yet, or the object will be replaced!** If you are using Rstudio, be also wary that your object might not appear immediately into your "Environment" panel – you may need to refresh it (Fig. 4.6). Finally, since the interface uses the current R session, you need to close the interface before using functions on your R object.

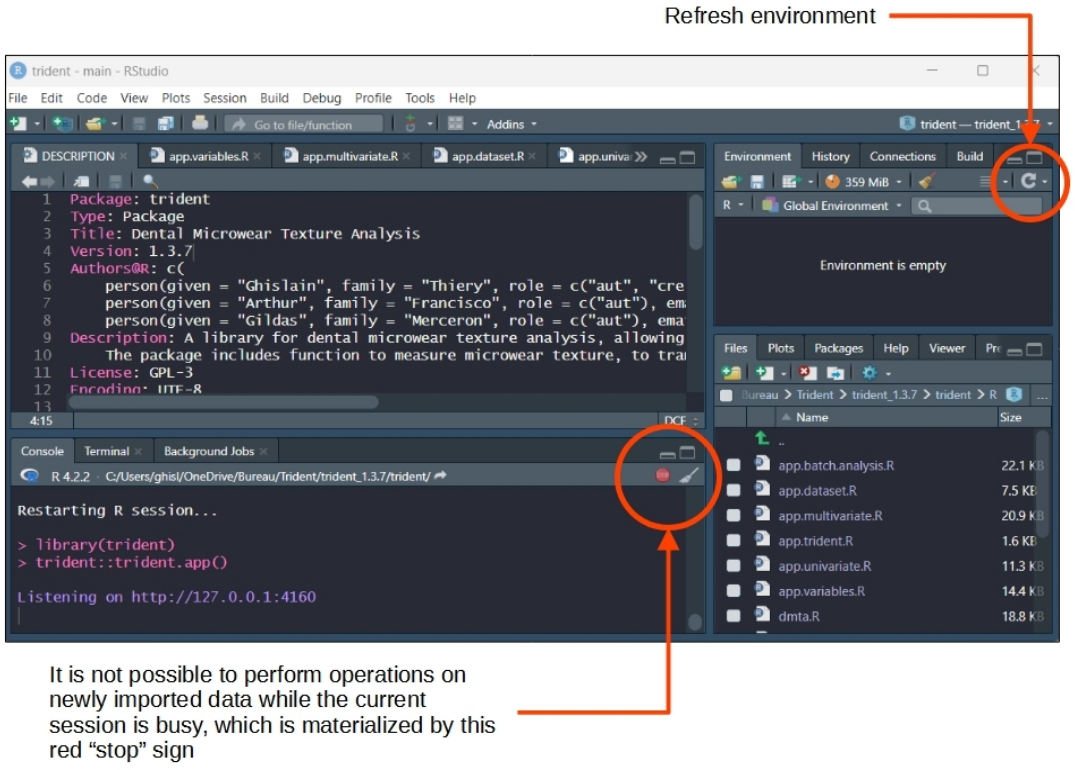


Figure 3.6. Exporting data to RStudio might require you to refresh the "Environment" panel

4. Graphics

4.1. Univariate

There are two options of univariate graphs in ***trident***: boxplot and violin plot. However, their respective buttons are disabled by default. To re-enable them, it is required to select a factor and a numeric variable: just click on the desired variables on the factor table (right) and numeric table (left), and they should be highlighted in blue, indicating that they are selected.

Tip: To unselect a variable, click again on that variable. The selection is maintained even if you navigate between the different table panels.

After selecting your variables, a box-and-whiskers graph will be built by default (Fig. 4.1). You can change the type of graph to display using the Box Plot and Violin Plot buttons.

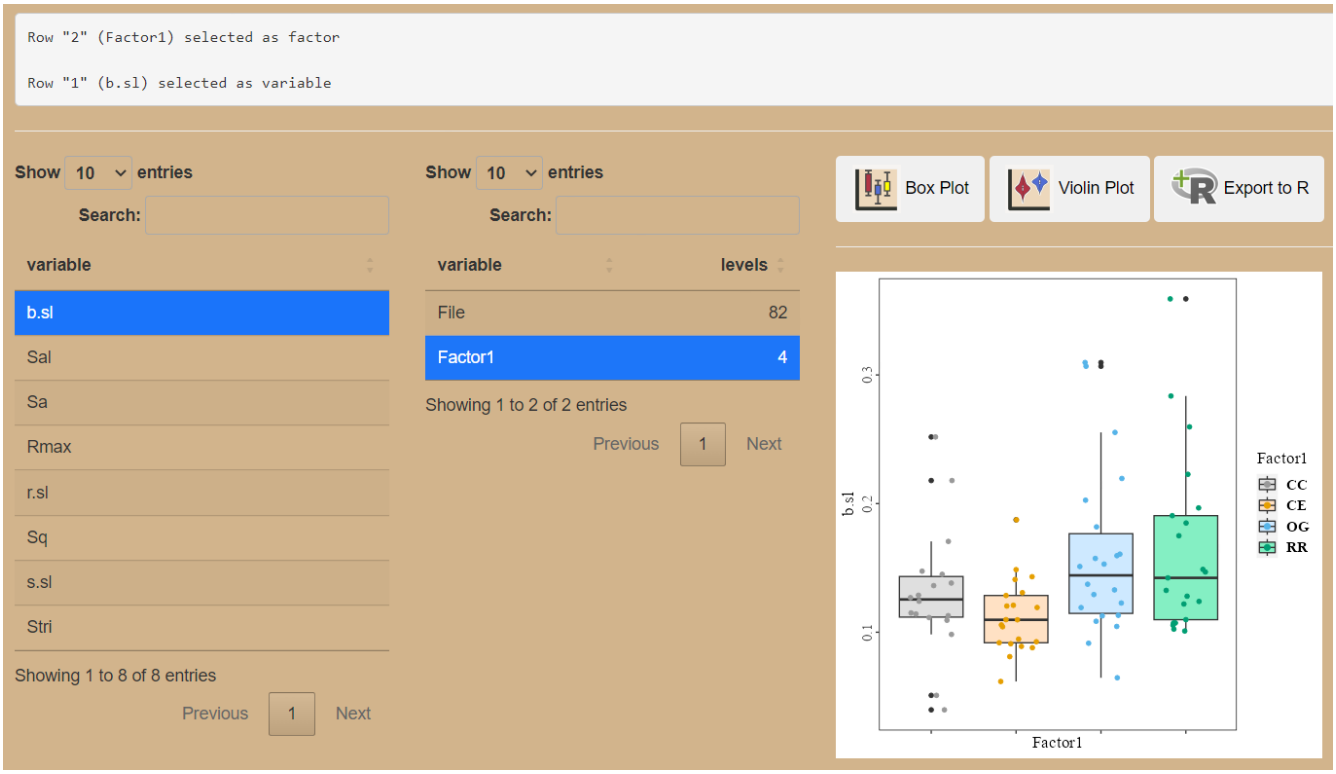


Figure 4.1. A box-and-whiskers graph displayed in ***trident***. Note the variables used in the graph are currently selected, as indicated by a blue highlight.

4.2. Multivariate

4.2.a. PCA

The multivariate tab presents an interface for the configuration of a principal component analysis (PCA). Most of the features are disabled by default: the user needs to select one factor variable and 2 or more numeric variables to re-enable them (Fig. 4.2.1).

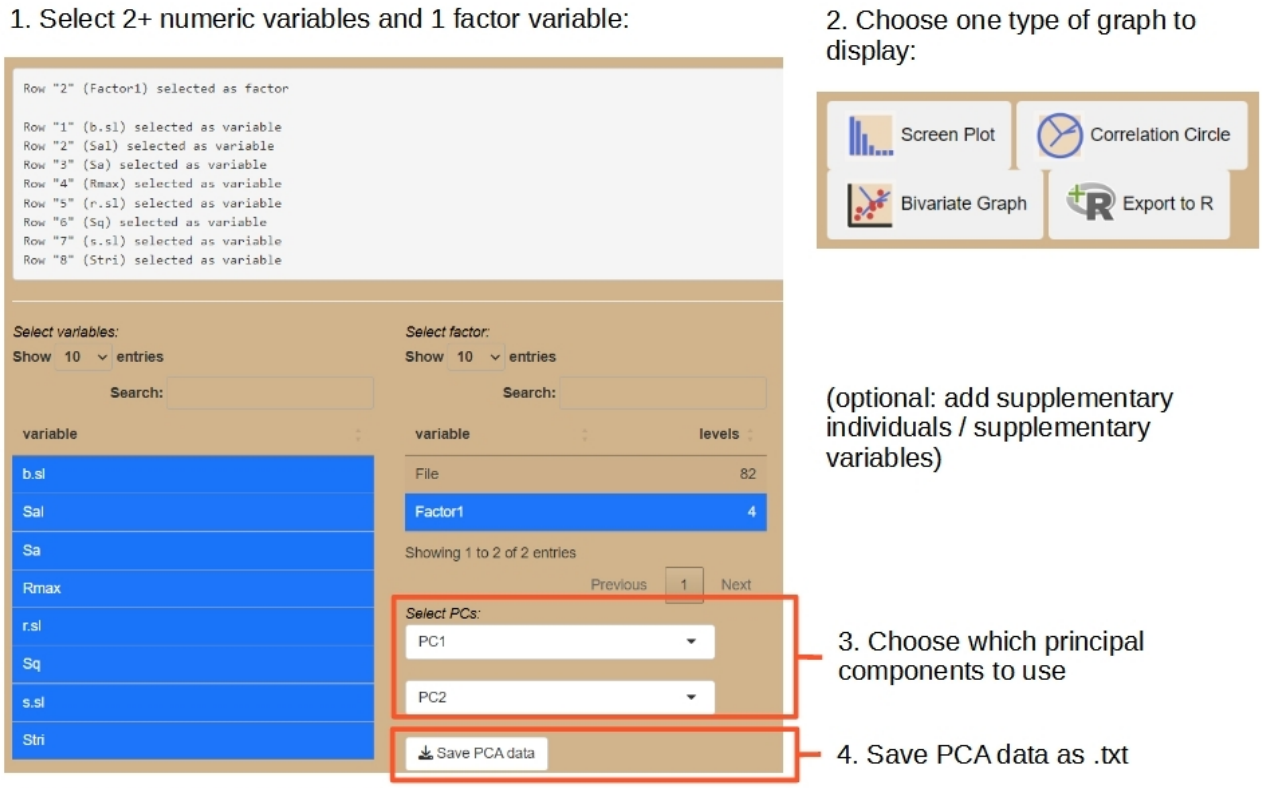
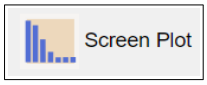


Figure 4.2.1. Configuration of PCA in ***trident***.

Afterwards, the user has to select one type of graph. Currently ***trident*** can produce 3 types of graphical outputs (produced using the factoextra R package; Kassambara and Mundt, 2020):

1. Screen plot: Click on 

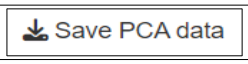
to display the screen plot with the percentage of explained variance for each principal component (Fig. 4.2.2).

2. Correlation circle: Click on 

to display the correlation circle, with any supplementary variables displayed in pink (Fig. 4.2.3). You can choose which principal components (PCs) are displayed on the graph (Fig. 4.2.1).

3. Bivariate graph: Click on 

to display the bivariate graph of the PCA, with any supplementary individuals displayed in pink (Fig. 4.2.4). You can choose which principal components (PCs) are displayed on the graph (Fig. 4.2.1).

Tip: The results of the PCA can also be saved as .txt using the 

button.

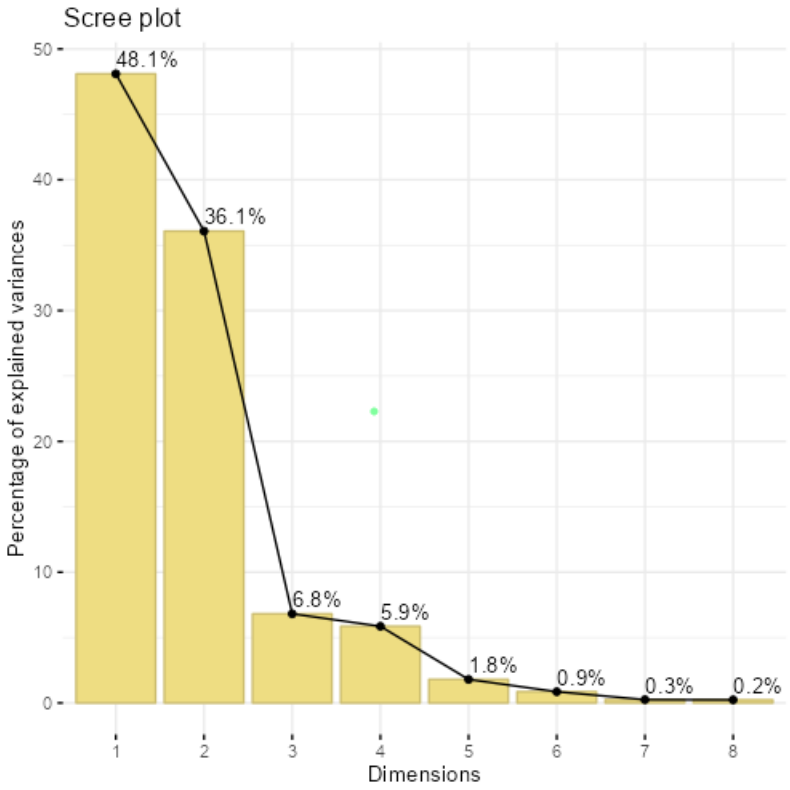


Figure 4.2.2. A screen plot produced using ***trident***.

4.2.b. Use supplementary variables

If you scroll down, you may have noticed a third table entitled “Select supplementary variables” (Fig. 4.2.3). Click on any variable in this table to add it to the PCA. These variables will not interfere

with the result of the PCA. Instead, their predicted correlation with PCs will be calculated and they will be added as supplementary arrows on the correlation plot.

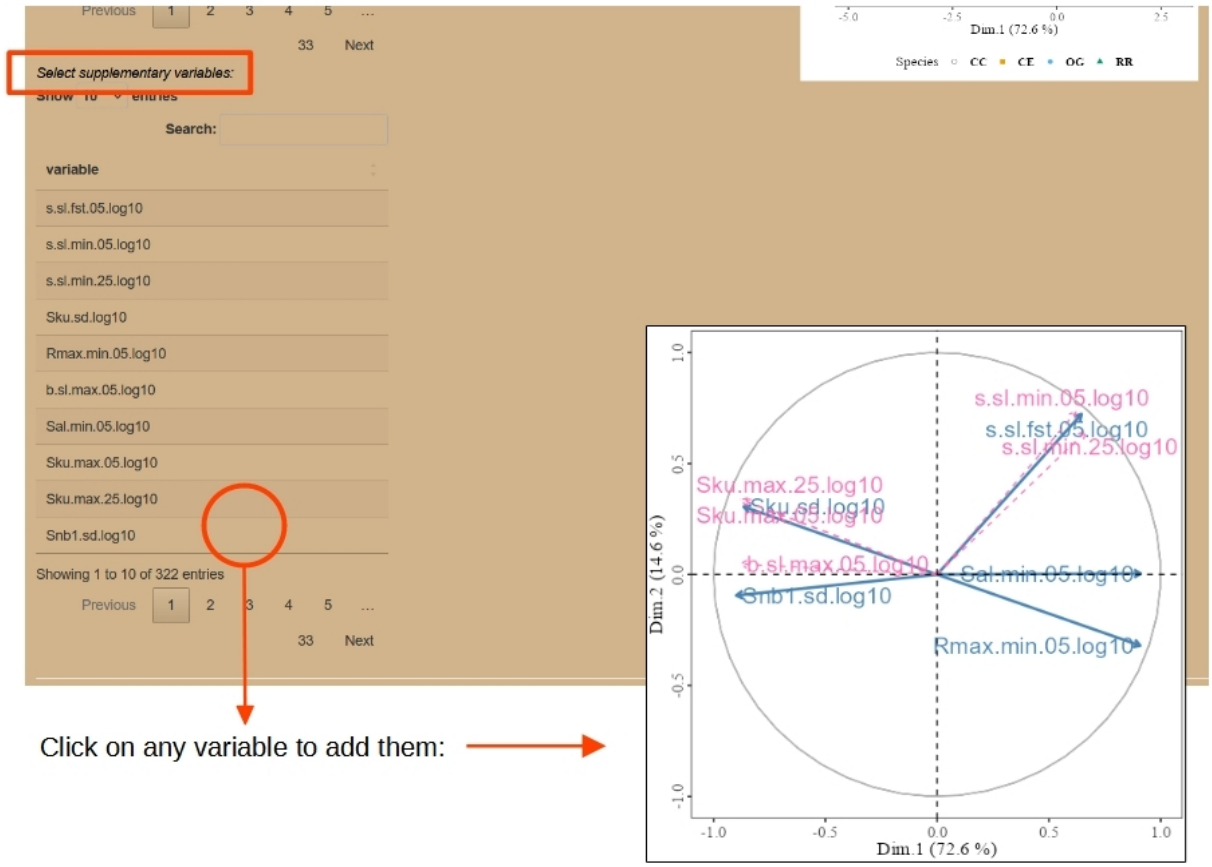


Figure 4.2.3. Using supplementary variables, with the resulting correlation circle.

Tip: Both supplementary variables and individuals are represented in pink

4.2.c. Add supplementary individuals

Click on the Browse... button to add supplementary individuals. Upon clicking the button, you will be prompted to select a .txt file.

Note: The default is set to tab-separated txt files, but you can change it to space-separated. Be wary that trying to import the wrong format file might crash ***trident***. If that happens, stop the activity of the current R session

Supplementary individuals will not interfere with the result of the PCA. Instead, their predicted coordinates will be calculated and they will be added as supplementary dots on the plot. Currently, supplementary individuals are only displayed with numbered pink dots, regardless of any factor. However, you can always export the graph in R to modify it.

Note: It is essential that the column names of the table provided for supplementary individuals matches the entirety of the variables used for the PCA. If you transformed your data, for instance with the log 10 transformation (see 2.2. Transforming the data), then you should also have transformed your supplementary data, and made sure that the variables' names were changed to "name.log10", as it will be automatically generated in ***trident*** for the original dataset used to generate the PCA model.

Note 2: Beware that Box Cox transformation is sample-dependent. One cannot make a Box Cox transformation on PCA individuals, then on supplementary individuals.

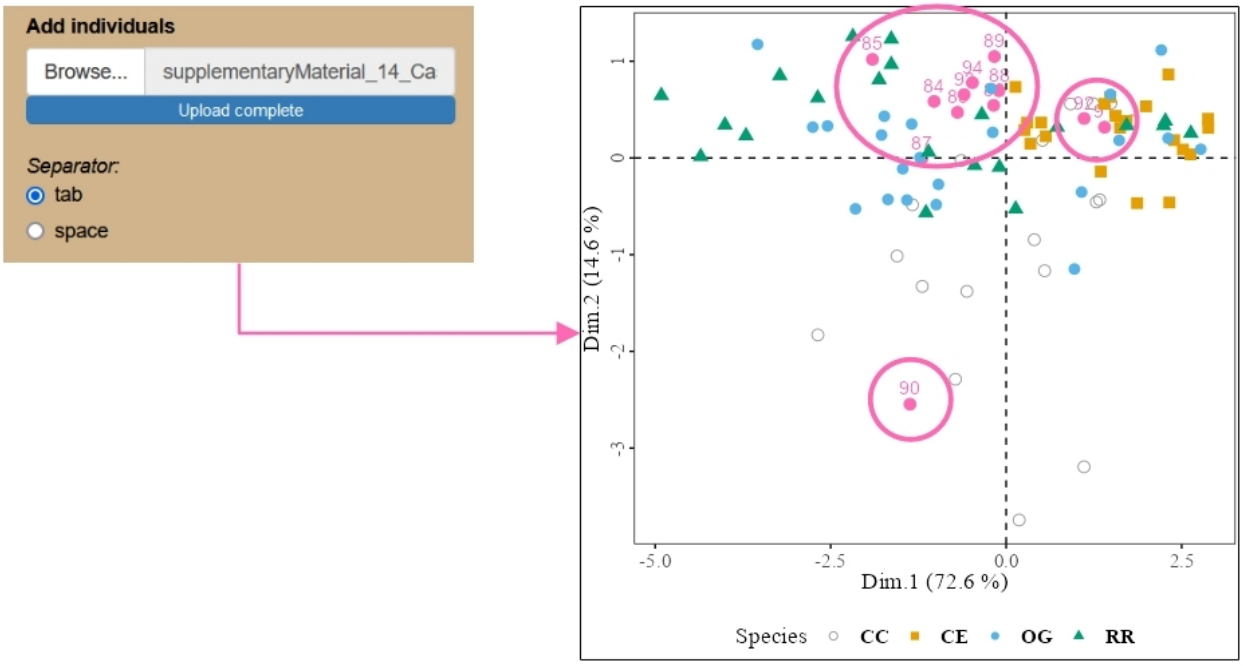


Figure 4.2.4. Adding supplementary individuals, with the resulting bivariate graph.

Tip: Both supplementary variables and individuals are represented in pink

4.3. Saving graphs

The graphs can be saved at a low resolution by right-clicking on the plot panel, then choosing 'Save'. Graphs can also be exported as ggplot2 objects (Wickham *et al.*, 2023) to your R session by clicking the Export to R button. You will be prompted to give a name to your R object: **make sure that this name has not been taken yet, or the object will be replaced!** If you are using Rstudio, be also wary that your object might not appear immediately into your "Environment" panel – you may need to refresh it (Fig. 1.6.2). Finally, since the interface uses the current R session, you need to close the interface before using functions on your R object.

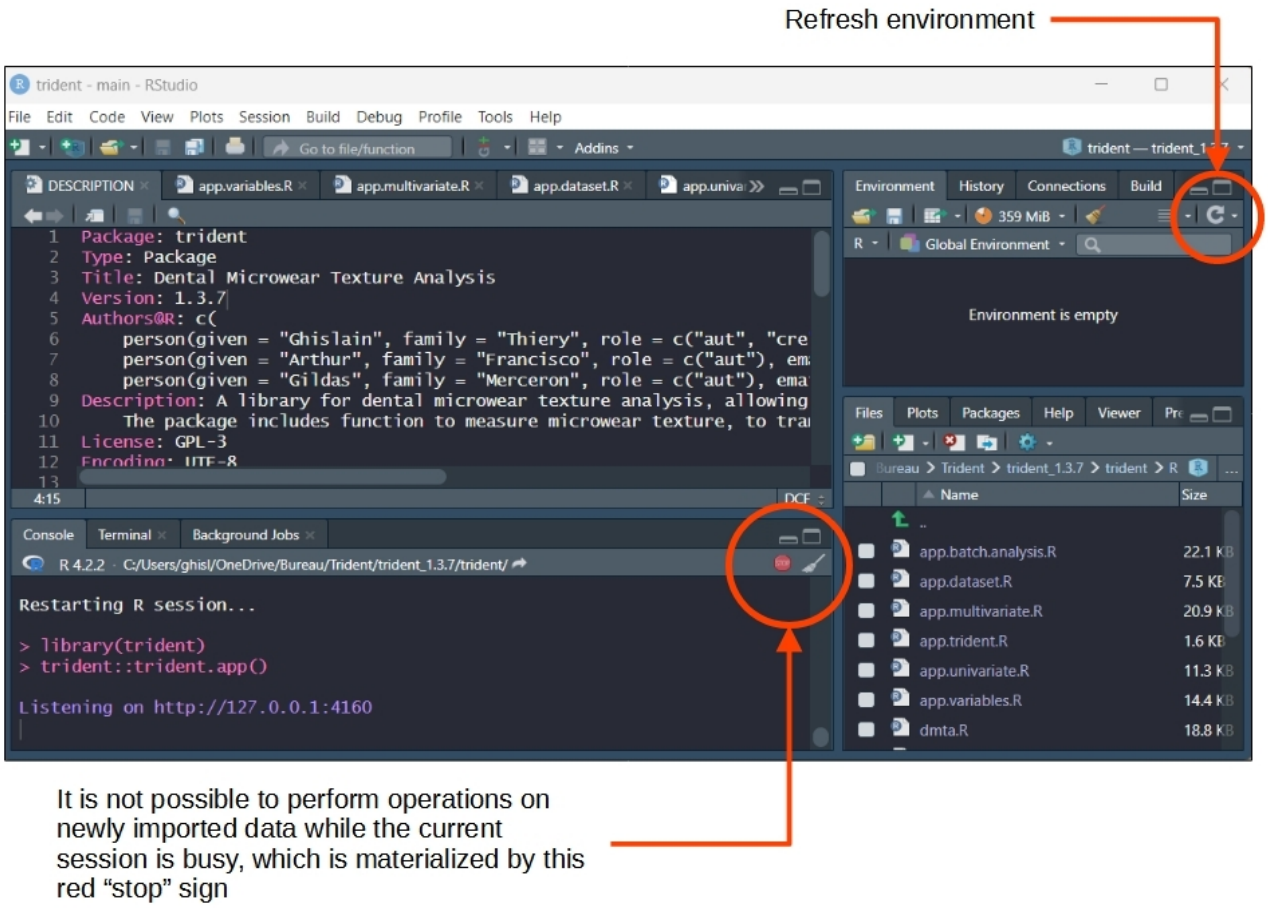


Figure 4.3. Exporting a graph to RStudio might require you to refresh the "Environment" panel

Conclusions

Concluding remarks

The ***trident*** app was developed to help beginners use the functionalities of the ***trident*** package, from the calculations of a broad range of DMTA parameters and their heterogeneity, to the classification of multiple variables according to many criteria. Likewise, the results can be exported in different formats, which will hopefully fit the needs of many users. However, would you like to complement the existing pipeline with one of the many available pakages in R, our functions were also designed as independent R functions. Once the package is installed, you can find more details by using the help documentation (for instance by calling ?trident::<function>).

Note: Visualization of 2D and 3D height maps

Although it did not make it to the app, the ***trident*** package is able to display 2D and 3D height maps of the surfaces. Here is an example of how to use it:

```
#2D height map

trident::sur.map(mySURfile, col.levels = 256, col.type = "temperature",
is3d = FALSE)

#3D height map

trident::sur.map(mySURfile, col.levels = 20, col.type = "colorblind",
is3d = TRUE)
```

It is possible to select up to 256 color levels, and three color hue types are available (“grayscale”, “colorblind” and “temperature”). You can take a snapshot and change some aspects of the window using options from the ‘rgl’ R package (Murdoch and Adler, 2021). For more information, use the help of the function:

```
?trident::sur.map
```

With all that being said, good luck and have fun using trident!

Acknowledgments

We would like to acknowledge our funders: the French National Agency for Research (ANR-13-JSV7-0008-01 Trident, PI: Gildas Merceron; ANR-17-CE27-0002-02 DIET-Scratches, PIs: Gildas Merceron, Stéphane Ferchaud); the Region of Nouvelle Aquitaine (ALIHOM #210389); the University of Poitiers; the National Center of Scientific Research (CNRS); National Research Institute for Agriculture, Food and Environment (INRAe).

We are also very thankful to many colleagues for their help and advice at different stages of this program's development, most notably: Emilie Berlioz, Cecile Blondel, Noël Brunetière, Ivan Calandra, Stéphane Ferchaud, Denis Gautier, Franck Guy, Anusha Ramdarshan, Antoine Souron, Axelle E.C. Walker.

References

Francisco, A. *et al.* (2018) ‘Enamel surface topography analysis for diet discrimination. A methodology to enhance and select discriminative parameters’, *Surface Topography: Metrology and Properties*, 6(1), p. 015002. Available at: https://doi.org/10.1088/2051-672X/aa9dd3.

Francisco, A., Brunetière, N. and Merceron, G. (2018) ‘Gathering and Analyzing Surface Parameters for Diet Identification Purposes’, *Technologies*, 6(3), p. 75. Available at: https://doi.org/10.3390/technologies6030075.

Kassambara, A. and Mundt, F. (2020) ‘factoextra: Extract and Visualize the Results of Multivariate Data Analyses’. Available at: https://CRAN.R-project.org/package=factoextra (Accessed: 30 March 2023).

Kembel, S.W. *et al.* (2010) ‘Picante: R tools for integrating phylogenies and ecology’, *Bioinformatics*, 26(11), pp. 1463–1464. Available at: https://doi.org/10.1093/bioinformatics/btq166.

Louail, M. *et al.* (2021) ‘Dental microwear textures differ in pigs with overall similar diets but fed with different seeds’, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 572, p. 110415. Available at: https://doi.org/10.1016/j.palaeo.2021.110415.

Merceron, G. *et al.* (2021) ‘Further away with dental microwear analysis: Food resource partitioning among Plio-Pleistocene monkeys from the Shungura Formation, Ethiopia’, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 572, p. 110414. Available at: https://doi.org/10.1016/j.palaeo.2021.110414.

Murdoch, D. and Adler, D. (2021) ‘3D Visualization Using OpenGL [R package rgl version 0.106.8]’. Comprehensive R Archive Network (CRAN). Available at: https://CRAN.R-project.org/package=rgl (Accessed: 10 June 2021).

R Core Team (2021) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Scott, R.S. *et al.* (2006) ‘Dental microwear texture analysis: technical considerations’, *Journal of Human Evolution*, 51(4), pp. 339–349. Available at: https://doi.org/10.1016/j.jhevol.2006.04.006.

Scott, R.S., Teaford, M.F. and Ungar, P.S. (2012) ‘Dental microwear texture and anthropoid diets’, *American Journal of Physical Anthropology*, 147(4), pp. 551–579. Available at: https://doi.org/10.1002/ajpa.22007.

Wickham, H. *et al.* (2023) ‘ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics’. Available at: https://cran.r-project.org/web/packages/ggplot2/index.html (Accessed: 20 June 2023).