

TERMITE Introduction V 2.0

The revised version of *TERMITE* provides a user interface (=UI) for much easier data handling compared to the previous version. The open source software *R* is required to use *TERMITE* (we recommend using *R-Studio* to benefit from all interactive features of the UI, see chapter 2).

1. General remarks

- TERMITE will never change your raw data files. All files will just be read into R's internal memory, and the calculations will be performed within R.
- The files of the reference materials need to have an assigned, unique name for the script to work properly. This name **must** start with the definite text string provided in the reference material section of the script (Figure 5, name in brackets ("NIST612", "GSD", "MACS3", "KL2-G", "BAM-B", "T1-G", "MACS1", "StHs", "ATHO-G" and "NIST610") followed by a sequence of letters or numbers (e.g., NIST610- 01). TERMITE calculates one RSF_{mean} for each isotope from the RSF values of all selected reference materials.
- Experiments consisting of more than 10 spot measurements need one leading zero in the filename (e.g., Spot_01.asc, Spot_02.asc). If the experiment consists of more than 100 spots, the user needs to label the data files using two leading zeroes (e.g., Spot_001, Spot_002).
- The output of the script is written into the "Results" folder.
- If the user needs to update the Reference Material values provided together with the script, the file is found in the directory "TERMITEScriptFolder" named "Standards_GeoReM.csv". This file can be opened and edited with Excel, OpenOffice or any other text editor. The values must be separated by commas, and white space is filled with NA for consistency.

2. Initial preparations

- We recommend using *R-Studio* for best user experience (<https://www.rstudio.com/> - The desktop version is free of charge). The script will also work with base *R* (<https://cran.r-project.org>) in combination with a web-browser. **You must install R first and R-Studio afterwards.**
- Unzip the file *TERMITE_2.zip*. This file contains the script TERMITE and all mandatory files for the script.
- Prepare, if not existing, additional folders in the main directory as shown in Figure 1.

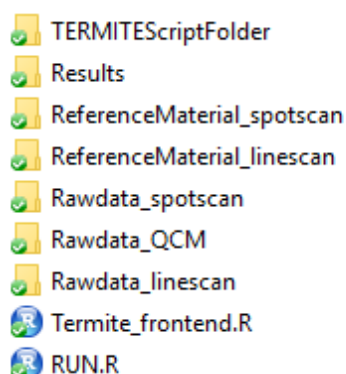


Figure 1: Structure of the directories, which must be created by the user if not existing.

- The main folder “*TERMITE_2*” contains the file “*RUN.R*” which provides all necessary commands to run the script. Additionally all commands are part of this instruction and can be copied from the PDF file. A sample dataset is included and can be used as a tutorial (check chapter 3). The name “*TERMITE_2*” can be changed.
- Copy the raw data sample files of the laser ablation measurements into the corresponding directory (Rawdata_linescan or Rawdata_spotscan). If a line scan is evaluated, the rawdata_linescan directory should only contain one single file. It is important that for all sample files stored in one folder, the same internal standard concentration (e.g., for a CaCO_3 , Calcium: $\sim 400,000 \mu\text{g/g}$) is used. If, for example, a quality control material (QCM), such as MACS-3, is treated as an unknown sample, the concentration of the internal standard is different from the sample (MACS-3, Calcium: $376,900 \mu\text{g/g}$). Therefore, the raw data files of MACS-3 must be stored in a separate folder (e.g., Rawdata_QCM).
- Experiments consisting of more than ten samples need one leading zero in the filename (e.g., Spot_01.asc, Spot_02.asc). If the experiment consists of more than 100 samples, two leading zeroes are required (e.g., Spot_001, Spot_002). This is required to guarantee correct sorting of the samples.
- Copy all Reference Material data files used for calibration into the folder ReferenceMaterial_spotscan/linescan.
- Run “*RUN.R*” to start data evaluation or copy the commands from this PDF file.

3. Data evaluation

- Start *R_Studio* and set working directory to “ \sim /TERMITE_2/” (see Figure 2 & 3).

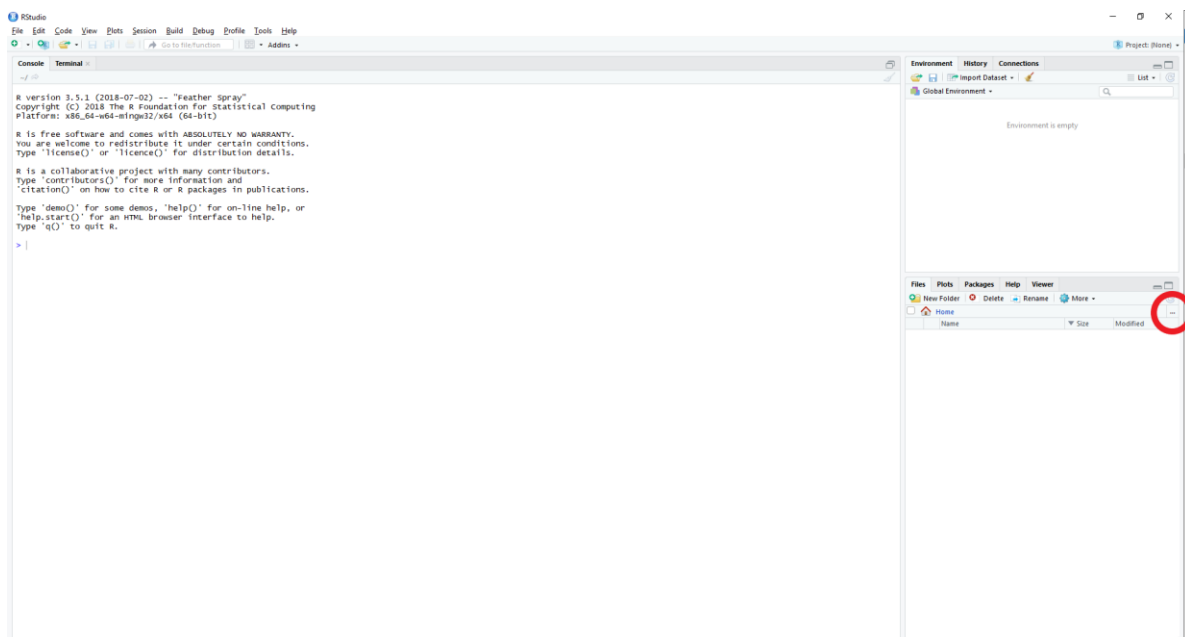


Figure 2: R-Studio command window. By clicking on the “...”symbol (marked with a red circle) a folder browser opens to navigate to your working directory.

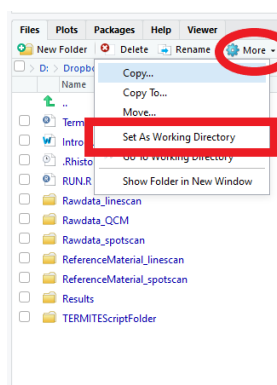


Figure 3: Click on “More” and select “Set As Working Directory”.

- Open “Run.R” in R-Studio. The file contains the needed **commands**, which must be executed stepwise (Figure 4).

```

1 #####
2 ###
3 ### TERMITE Commands
4 ###
5 ### An R-Script for fast reduction of LA_ICPMS data and its application to trace element measurements
6 ###
7 ### Version 2.0
8 ### TERMITE originally developed by Simon Mischel
9 ### Revised by Dennis Rupprecht
10 ###
11 ### Speleothem research group, Johannes-Gutenberg-University Mainz
12 ###
13 #####
14
15 ### Additional packages will be installed during the first usage of the script.
16 ### regarding admin access following scripts have to be installed manually:
17 Commands: miscTools", "matrixStats", "ggplot2", "reshape2", "shiny", "shinyjs", "DT", "Hmisc"))
18
19 has to be executed stepwise, e.g. line by line
20
21 rm(list=ls(all=TRUE)) # Deletes all active variables to avoid duplicates
22
23 ##### !!!!!
24 ###
25 ### Edit your working directory, i.e. the directory which contains the 'Termite_frontend' script and all subfolders
26 ### !!!!!
27
28 setwd("~/TERMITE/your_main_directory/") # Defining the working-directory of the script
29
30 ### Do not edit
31 source(paste(getwd(),"Termite_frontend.R", sep ="/")) # Defining your source file for the script
32
33 ### Data processing. Execute stepwise
34 # Step 1 requires additional editing in the popup window
35 Data(Data_input) # Runs the editable input script
36 # Step2 generates the output files
37 Run_script(Start_script) # Starts data processing and generates the output
38
39
40 29:1 (Untitled)
41
42 Console Terminal
43
44 R version 3.5.2 (2018-07-02) -- "Feather Spray"
45 Copyright (C) 2018 The R Foundation for Statistical Computing
46 Platform: x86_64-w64-mingw32/x64 (64-bit)
47
48 R is free software and comes with ABSOLUTELY NO WARRANTY.
49 You are welcome to redistribute it under certain conditions.
50 Type 'license()' or 'licence()' for distribution details.
51
52 R is a collaborative project with many contributors.
53 Type 'contributors()' for more information and
54 'citation()' on how to cite R or R packages in publications.
55
56 Type 'demo()' for some demos, 'help()' for on-line help, or
57 'help.start()' for an HTML browser interface to help.
58 Type 'q()' to quit R.
59
60 > |

```

Figure 4: “Run.R” in R-Studio command window. All lines indicated in green starting with “#” represent comments and do not affect the script. The needed command-lines to operate the script can also be found in this manual, chapter 3. The console window is displayed in the lower part of the program.

- Copy the commands into R-Studios’ console window step-by-step and execute each command by pressing *enter* (see Figure 4). Code is highlighted in red.

- Clear the workspace of previous information:
rm(list=ls(all=TRUE))
- Pass the information of the *TERMITE* source file to R:
source(paste(getwd(),"Termite_frontend.R", sep ="/"))
- If the script is executed for the first time several additional R packages will be installed. If any error occurs based on admin rights or server errors, try installing the packages by hand with the following command. (Information about old versions may be ignored):
install.packages(c("miscTools", "matrixStats", "ggplot2", "reshape2", "shiny", "shinyjs", "DT", "Hmisc"))
- Start the script and begin editing your data in the pop-up window. (Data editing is shown in Figures 5-7):
Data(Data_input)
- After finishing the editing of your data, the “GO!” button must be pressed, and the pop-up window must be closed. Afterwards TERMITE will perform all necessary calculations:
Run_script(Start_script)

- Data editing and visualization takes place within the UI generated by *TERMITE* (Figure 5). The meaning of each value listed below “Data Input” is explained in detail in the subsequent part of this manual. (The tutorial files stored in “Rawdata_linescan” or “Rawdata_spotsan” might be opened with excel or similar software to compare the file to the input from the UI):

TERMITE

Input Data Plot

Data Input

☒ Save a summary of your inputs

Sample name
your_sample_name

ICP-MS used for analysis
Agilent

Resolution (Thermo Element 2 only)
(LR)

Background correction mode
median

Line of Header in raw data file
3

First line of data
4

Number of measured isotopes
12

Column of internal standard in the raw datafile
5

Concentration of internal standard element (µg/g)
400003.81

Summary

Name	Value
Save summary	TRUE
Name of sample	your_sample_name
Machine used	Agilent
Resolution (Thermo Element 2 only)	(LR)
Background correction mode	median
Header (isotopes measures), line	3
Start of data, line	4
Number of isotopes measured	12
Column number of internal standard	5
Concentration of internal standard (µg/g)	400003.81
Perform outlier test	Yes
Range for outlier test [%]	30
Scanmode	line scan
Overall number of rows (line scan)	6996
First row number used for integration interval of background-correction	5
Last row number used for integration interval of background-correction	124
First row number used for integration interval of sample analysis	143
Last row number used for integration interval of sample analysis	6890
Laser speed (µm/s)	5
Overall number of rows (spot scan: sample & reference material)	550
First row number used for integration interval of background-correction	5
Last row number used for integration interval of background-correction	124
First row number used for integration interval of sample analysis	130

Figure 5: Pop-up window generated by *TERMITE*. The user interface consists of 3 tabs: “Input”, “Data” and “Plot”. In the first tab, “Input”, the values needed for evaluation are provided by the user. The second tab, “Data”, generates a table based on the inputs and the third tab, “Plot”, visualizes the data handed to the algorithm.

- **Sample name:** The output file name of your data
- **ICP-MS used for analysis:** The machine used for laser ablation
- **Resolution:** The resolution chosen (low “LR”, mid “MR”, high “HR”) for Thermo Element 2
- **Background correction mode:** The mean or median can be used for background correction
- **Line of Header in raw data file:** The row in which the names of the isotopes are stored
- **First line of data:** The row in which the signal for the ablation run starts including the background signal
- **Number of measured isotopes:** Number of all measured isotopes including the internal standard. The measured isotopes in sample material and reference material have to be the same.
- **Column of internal standard in the raw datafile:** Number of the column where the internal standard isotope is located
- **Concentration of the internal standard element in µg/g:** e.g. 400003 for Ca in CaCO₃
- **Perform outlier test:** The algorithm can perform an outlier test for reference materials and spot scans. This option can be switched on/off.
- **Range of outlier test:** The percentage of the range used for the outlier test. (This feature can be seen in the tab “Plot”).
- **Scan mode:** The mode of scan used for the experimental setup. The “Line scan section” will be switched off, if “spot scan” is selected.
- **Integration intervals** need to be defined. In case of line scan analyses this is done in the “Line Scan section” for the sample measurement and in the “Spot scan/reference material section” for the reference materials. In case of spot scan analyses, the integration intervals are only defined in the “Spot scan/reference material section”, since spots on the samples and reference materials are treated identically:

- **No. of last row in raw datafile:** The overall number of rows in your raw data file
- **First row number used for integration interval of background-correction:** The first value used for the background calculation. Typically, the first and the last one or two values are cut off to ensure consistency of the data. This is done in the section for the line scans as well as the spots.
- **Last row number used for integration interval of background-correction:** The last value used for the background calculation. Typically, the first and the last one or two values are cut off to ensure consistency of the data. This is done in the section for the line scans as well as the spots.
- **First row number used for integration interval of sample analysis:** The first value used for the calculation of the sample concentration need to be provided in the sections for the line scans as well as the spot and reference material measurements.
- **Last row number used for integration interval of sample analysis:** The last value used for the calculation of the sample concentration need to be provided in the sections for the line scans as well as the spot and reference material measurements.
- **Laser scan speed in $\mu\text{m/s}$:** The scan speed of the laser needs to be provided in the line scan section because this value allows the calculation of the length of the measurements along the sample surface using the time information of the raw data file.
- **Reference Materials used for calibration:** The user can decide which and how many reference materials are used for calibration. The script will only work properly if at least **two** reference material files from one material are present in the directory. At least one measurement of the selected reference material(s) should be performed before and after a set of spots (e.g., 30-45 spots). This enables to correct for the drift of the machine during the analytical session. For each reference material, a Relative Sensitivity Factor (see Mischel et al. (2017) for details) is calculated. Finally, to correct the measured element concentrations of the unknown sample, *TERMITE* applies the mean of the individual RSF values determined from the measurements of the selected reference materials. All materials with a set tick mark will be used.

TERMITE

Input Data Plot

Data

Generate data tables

Spotscan: The first file is automatically chosen from your input folder

Table containing all values

Show 10 entries

	Time [Sec]	Mg25	Al27	P31	Ca43	Sr88	Y89	Ba137	La139	Ce140	Pb208	Th232	U238
1	0.156	0	1000.05	2700.34	100	0	0	0	0	0	0	0	0
2	0.302	100	900.04	3400.54	100	0	0	0	0	0	100	0	0
3	0.448	0	1100.06	2600.32	0	0	0	100	0	0	0	0	0
4	0.594	0	600.02	3600.61	0	0	0	0	0	0	100	0	0
5	0.74	100	900.04	3300.51	0	0	100	0	0	0	0	0	0
6	0.886	0	700.02	3000.42	0	100	0	0	0	0	0	0	0
7	1.032	0	800.03	3300.51	100	0	0	0	0	0	100	0	0
8	1.178	100	700.02	3200.48	100	0	0	100	100	0	0	0	0
9	1.324	0	700.02	3000.42	0	0	0	0	0	0	0	0	0
10	1.47	0	1000.05	2300.25	0	200	0	0	0	0	0	0	0

Showing 1 to 10 of 6,993 entries

Previous 1 2 3 4 5 ... 700 Next

Table containing values for integration interval only

Show 10 entries

	Time [Sec]	Mg25	Al27	P31	Ca43	Sr88	Y89	Ba137	La139	Ce140	Pb208	Th232	U238
1	20.453	77882.53	14309.58	169636.57	2968685.42	163579.67	200	11306.3	0	200	4200.9	0	1700.15
2	20.599	73854.1	19617.99	156540.54	2841356.51	155963.72	0	9204.18	0	100	2900.43	0	1900.19
3	20.745	87355.26	15611.4	165067.66	2836979.96	167745.39	400.01	9004	0	0	2800.4	0	1500.12
4	20.891	82315.54	32850.42	161412.36	2825327.6	156065.23	300	12307.47	100	100	2800.4	0	1700.15

Figure 6: The upper table contains all values provided to the algorithm including the background corrections. The lower table includes the values for the integration interval of the analysis. Values are displayed after clicking “Generate data tables” marked with a red circle. No further editing is needed in this tab.

- A summary of the data input will be displayed on the right (see Figure 5). This information will also be stored in the Result folder as “Summary.txt”.
- If all values are edited in the tab “Input” you might proceed to the “Data” tab. The tab “Data” provides 2 tables generated based on your input values for one sample file. By clicking “Generate data tables” the algorithm creates the tables needed for visualization on the tab “Plot” (see Figure 6).
- The tab “Plot” generates the visualization of your data (see Figure 7) listed in the tab “Data”. The plot is reactive, which means that you can edit your input values in the tab “Input”, then click on “Generate data tables” from the tab “Data” again and then look at your updated plots in the “Plot” tab.

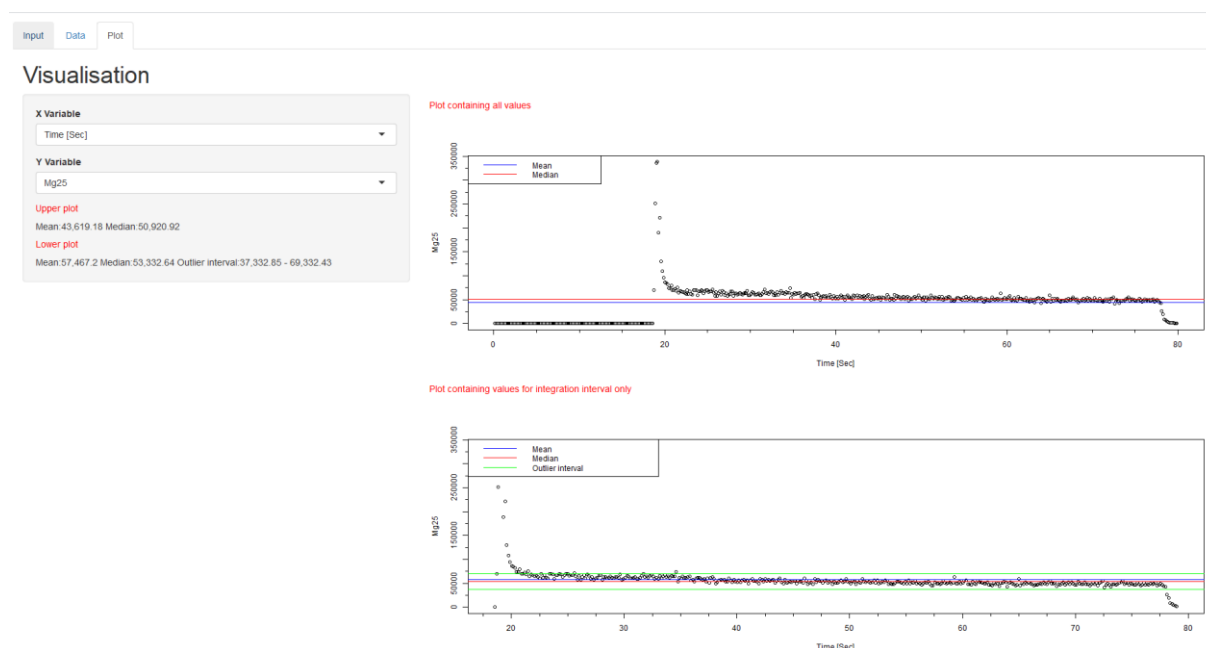


Figure 7: On the left panel you can choose the values for the x- and y-axis. It is recommended to choose the time as the x-axis and the isotope of your choice as the y-axis. The upper plot contains all values (e.g. from “First row number used for integration interval of background-correction” to “Last row number used for integration interval of sample analysis”) and a calculated mean (blue) and median (red) (values are shown on the left sidebar). The lower plot shows the integration interval of your sample (“First row number used for integration interval of sample analysis” to “Last row number used for integration interval of sample analysis”). The mean (blue) and median (red) are shown as well as the Outlier interval (green) which was previously chosen in the “Input” tab (“Range of outlier test (Deviation from median) [%]”). Values can be changed, and the plot updated by generating the tables (“Data” tab) again.

- After all values have been adjusted in the tab “Input”, press the “Go!” button.
- Close the GUI window to pass the edited information to R. Afterwards TERMITE will perform all necessary calculations by running the command (copy the commands into R-Studios’ console window):

Run_script(Start_script)

The data output (raw count rates, element concentrations, LoD, RSF, all as csv and pdf) will be written to the “Results” folder.

- If any values in the Input tab are updated and the changes shall be used for the final data output the “Go!” button in the “Input” tab must be pressed again before closing the pop-up window.
- Previous results will be overwritten in a subsequent application of the algorithm. Therefore, your results should be copied to another folder once the evaluation is done.

4. Troubleshooting:

General considerations: Warnings will not stop the script from execution, and the script should finish. Any error that occurs during data reduction will terminate the script and troubleshooting should start. The user is advised to try to understand the error messages provided by R either by reading the provided troubleshooting section or by using a search engine on the internet.

- If the raw data contain any E-values, which are errors introduced during the data saving process by the software of the ICP-MS, these values will be set to NA prior to data reduction. In R, this will produce warnings, which can be displayed using the command `warnings()`. The message will be (in short): *Warning message: In FUN(x[[i]], ...): NAs introduced by coercion.*
- An “error in file(file, “rt”) : invalid ‘description’ argument” may occur when more than ONE file is present in the directory “Rawdata_linescan”.
- In rare cases, R produces a warning
`Warning: Error in exprFunc: object 'tabelle' not found`
[No stack trace available]
`Warning: Error in exprFunc: object 'tabelle2' not found`
[No stack trace available],
this can be avoided by generating the data tables in the tab “Data”, sometimes it is required to generate them twice.
- If other errors occur, please check again all values in the different sections of the script for correctness.
- Huge outliers within your blank (background correction) may lead to program crashes in rare cases and must be dealt with manually in your raw data file.
- If, in rare cases, R will not stop the calculation for a long time, please consider the amount of data which is evaluated. For instance, 24,000 spot measurement files will result in TERMITE to work for about 3 h. This will produce a raw count rate file with several GBs in size.

TERMITE was originally developed by Simon Mischel.
The GUI was added Dennis Rupprecht.

Mischel SA, Mertz-Kraus R, Jochum KP, & Scholz D (2017). TERMITE - An R script for fast reduction of LA-ICPMS data and its application to trace element measurements. *Rapid Communications in Mass Spectrometry* 31, 1079-1087.

R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rupprecht D, Mertz-Kraus R, Mischel S, Budsky A, Jochum KP & Scholz D (2019). TERMITE 2.0 – An R Script for Data Reduction of LA-ICP-MS Trace Element Measurements. *Goldschmidt Abstracts*, 2019

Speleothem research group, Johannes Gutenberg University Mainz
Contact: Prof. Dr. Denis Scholz, scholz@uni-mainz.de