

## TERMITE: Step-by-step walkthrough

This guide provides a detailed walkthrough for TERMITE (Mischel et al. (submitted): *TERMITE - an R script for fast reduction of LA-ICPMS data and its application to trace element measurements*). The reader should carefully work through this manual prior to the usage of TERMITE.

### General remarks:

- A section for troubleshooting is provided at the end of this guide.
- Lines starting with # as well as everything in a line following # are commentaries.
- Variables consisting of text (e.g., "your\_sample\_name") must be provided with quotation marks.
- 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 in order 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", also possible are "N610" and "N612") followed by a sequence of letters or numbers (e.g., NIST610-01). When only a single reference material is used, the line RefMat1 must be filled in and all other lines must start with #. If two or more lines without # are indicated, those reference materials will be used for calibration (see example Figure 5). The reference materials must be in chronological order (e.g. Reference Material 1, Reference Material 2), no matter how many reference materials have been analysed during the analytical session. TERMITE calculates for each isotope one  $RSF_{\text{mean}}$  from the RSF values of all selected reference materials.

- It is important to note that even if the files of two reference materials are copied into the ReferenceMaterial\_directory, only the files, which are indicated in the reference material section of the script, are used.
- 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). This practice is mandatory to guarantee correct sorting of the spots on the sample.
- The output of the script is written into the Results\_directory:
  - A pdf file with the raw count rates is plotted together with vertical lines indicating the sections used for the background determination and the sample signal as defined in the HelpValue sections of the script (rawCountrate\_your\_sample\_name.pdf)
  - The results are saved as a csv and a pdf file (Results\_your\_sample\_name.csv and Results\_your\_sample\_name.pdf).
  - The Limit of detection (LoD) is saved as a csv and a pdf file (LoD\_ReferenceMaterial\_your\_sample\_name.csv and LoD\_ReferenceMaterial\_your\_sample\_name.pdf).
  - The RSF values are saved as a csv and a pdf file (RSFused\_your\_sample\_name.csv and RSFused\_your\_sample\_name.pdf).
- 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.

### Initial preparations:

1. Download and install R (<https://cran.r-project.org>). If problems arise during the installation, please use the documentation provided on this website.
2. Unzip the file `your_main_directory.zip`. This file contains the script `TERMITE` and all mandatory files for the script. It also contains example data sets of a line scan and spot measurements.
3. For spot analyses, the file `TERMITE_spotscan.r` is used. For line scan analyses, the file `TERMITE_linescan.r` is used. Technically, these files are identical, but the example data sets include both types of experiments, and the respective file will perform the data reduction of the corresponding data set.
4. Navigate into your main directory.
5. Prepare, if not existing, additional folders in the main directory as shown in Figure 1. The user can decide which names should be assigned to the directories.

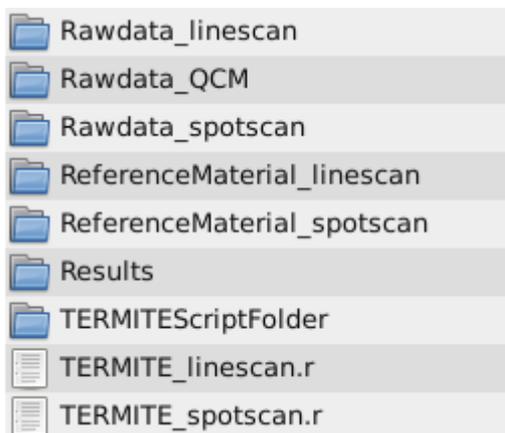


Figure 1: Structure of the directories, which must be created by the user if not existing.

6. 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  $\text{CaCO}_3$ ,

Calcium: ~400,000 µg/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 µg/g). Therefore, the raw data files of MACS-3 have to be stored in a separate folder (e.g., Rawdata\_QCM).

7. 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.
8. Copy all Reference Material data files used for calibration into the folder ReferenceMaterial\_directory.
9. Open the file "TERMITE\_script\_linescan.r" or "TERMITE\_script\_spotscan.r" depending on the type of your experiment (spot or line scan analysis) and work through the script prior to data evaluation. The corresponding steps are explained in the following (Figures 3, 5 and 6).
10. The packages "miscTools" and "matrixStats" will be installed automatically if the corresponding lines are run (the three lines after the section "installs all required packages", Figure 2). In case of any problems, try to choose a different download mirror or try using "http" instead of "https". In case of errors, these packages can be installed manually by executing the commands `install.packages("matrixStats")` and `install.packages("miscTools")`.
11. As a first step, the path and names of the folders containing the raw data files (reference materials and sample) must be filled in the Directory section of the script (if not already existing, Figure 2). It is important that the names are identical to the names assigned in step 5. Note that R is case sensitive and distinguishes between lower and upper case letters. In addition, all directories must be provided using slash (/) instead of backslash (\).

```

#####
#
#
#          TERMITE
#
#   An R-Skript for fast reduction of LA_ICPMS data and
#   its application to trace element measurements
#
#          Version 1
#
#          Simon Mischel
#
#   Speleothem Research Group, University Mainz
#
#   eMail: simon.mischel@uni-mainz.de
#
#####
### find . -name "*.csv" -type f -exec cp {} ~/your_folder \;

##### installs all required packages
rm(list=ls(all=TRUE))
if (!require("miscTools")) install.packages("miscTools", dependencies = TRUE) # delete the internal R memory to avoid duplicates
if (!require("matrixStats")) install.packages("matrixStats", dependencies = TRUE) # on first run package will be installed if not available

##### Directory section #####
path <- "~/your_main_directory/" # defining the working-directory of the script
path.rawData.samples <- "Rawdata_linescan" # defining the directory of the raw-data-files
path.rawData.RefMat <- "ReferenceMaterial_linescan" # defining the directory of the reference material
path.corrData.results <- "Results/" # defining the directory for the overview-plots

sample.name <- "your_sample_name_linescan" # is printed in the file-names

```

Figure 2: Directory section of the script. Slash (/) at the end of path is mandatory (black arrow).

12. The slash (/) in Figure 2 at the end in the object path is mandatory.

13. The name of the sample (sample.name, e.g. "your\_sample\_name") should be assigned. This helps the user to identify the sample after data reduction because this text string is used in all file names in the results folder.

14. Prior to data reduction, the user should work through the section "HelpValues" and provide all important parameters for data reduction in the script (Figure 3).

```

##### general HelpValue Section #####
Mode.of.Scan <- "line.scan" # spot.scan or line.scan
Line.of.Header <- 31 # the line where the measured isotopes are listed
Line.of.Signal <- 42 # the line where the raw data start
measured.isotopes <- 12 # number of analysed isotopes
column.IS <- 5 # column of internal standard in the rawdata-file
IS <- 400003.81 # [µg/g] internal standard
outlier.Test <- "Y" # "Y" or "N"
m <- 30 # percentage for the range of the outlier test
number.sweeps.sample <- 6996 # No. of sweeps (e.g. overall number of rows)
number.sweeps.Refs <- 550 # No. of sweeps (e.g. overall number of rows)
resolution <- "(LR)" # Resolution "(LR)" "(MR)" "(HR)" (needed for Thermo Element2 ICP-MS)
machine <- "ThermoFosher" # ThermoFisher or Agilent

background.correction <- "median" # method for background.correction (mean <-> median)

##### HelpValue for Linescans Section #####
first.blankValue.linescan <- 43 # first Blank-Value (e.g. line number) used for background-correction
last.blankValue.linescan <- 64 # last Blank-Value (e.g. line number) used for background-correction

first.sampleValue.linescan <- 70 # first Laser-on-sample-Value (e.g. line number) used for analysis
last.sampleValue.linescan <- 1520 # last Laser-on-sample-Value (e.g. line number) used for analysis

laser.speed <- 5 # laser scan speed in µm/s

##### HelpValues for spot scans and/or reference material Section #####
first.blankValue <- 43 # first Blank-Value (e.g. line number) used for background-correction
last.blankValue <- 64 # last Blank-Value (e.g. line number) used for background-correction

first.sampleValue <- 70 # first Laser-on-sample-Value (e.g. line number) used for analysis
last.sampleValue <- 1520 # last Laser-on-sample-Value (e.g. line number) used for analysis

```

Figure 3: Section of the script, where all important parameters regarding structure of raw data are filled in. Please change only blue numbers or green text strings with quotation marks.

15. In the following, the HelpValue section is explained step by step:

- `Mode.of.Scan`: insert "spot.scan" or "line.scan" depending on your experimental setup.
- `Line.of.Header` is the line in the original raw data file, where the names of the isotopes are found in the data files (Figure 4).
- `Line.of.Signal` is the line, where the signal of the ablation run starts, normally starting with the background signal (Figure 4).
- `Measured.isotopes` is the number of all measured isotopes including the Internal Standard (IS).
- `Column.IS` is the number of the column in the raw data file, where the internal standard isotope (IS) is located (Figure 4).
- `IS` is the concentration of the internal standard in the sample (e.g., 400,000 µg/g for Ca in case of CaCO<sub>3</sub>)
- The outlier test (for spot scan measurements and for reference materials) can be switched on and off. Please note that in case of a line scan, the outlier test will only be performed for the reference materials.
- `m` is the percentage of the range used for the outlier test.
- `number.sweeps.sample` is the number of total sweeps recorded in the raw data file of the sample (Figure 4).
- `number.sweeps.Refs` is the number of total sweeps recorded in the raw data files of the reference materials. Typically, spot scans are measured with the same experimental setup (e.g., background and ablation time, etc.) as the reference materials. In this case, `number.sweeps.sample` and `number.sweeps.Refs` contain the same value.
- The string machine allows the user to specify, which ICPMS instrument was used (provide "ThermoFischer" or "Agilent").
- Using the ThermoFisher Element2 mass spectrometer, the mode of resolution during ablation can be set to low ("LR"), mid ("MR") and high ("HR"). This text string must be

provided to enable TERMITE to delete this text string from the header to obtain the names of the isotopes measured during ablation.

- The user can decide, which `background.correction` for the background will be applied (“mean” or “median”).
- The section “HelpValues for Linescans” needs to be completed if line scans are evaluated. In case of spot measurements, the values in this section will not be used during calculation and can be left unchanged.
- The values in the “HelpValues for spot scans and/or reference material Section” must always be provided.

	A	B	C	D	E	F
1	Acquisition Parameters				column.IS	
2						
3	Data File :		C:\Element\user\Element2\Data\Simon-07_10_15-ine1\HLK2-1.dat			
4	Error :					
5	Analysis Date :	Wed		07-Oct-2015 11:40:50		
6	Sample Name :					
7	Tune Parameters :		C:\Element\user\Element2\Tune Parameters\213-Laser-28-07-15.tpf			
8	Method File :		C:\Element\user\Element2\Methods\Simon-07_10_15-HLK2-1.met			
9	Runs/Passes (Meas.) :		1490 * 1 + 0 * 0 + 0 * 0			
10	Res. Switch Delay [s] :			2		
11	Washline [min] :			0		
12	Take-up Time [min] :			0		
13	Deadtime [ns] :			20		
14	Evaluation Parameters					
15						
16	Evaluation Date :	Wed		07-Oct-2015 11:58:10		
17	User :	Element2				
18	Analysis Type :	SMP				
19	Standard File :					
20	Int. Standard File :					
21	Blank File :					
22	Runs/Passes (Eval.) :		1490 * 1 + 0 * 0 + 0 * 0			
23	Quantification Type :		Intensities			
24	Calibration File :					
25	Response File :					
26	Dilution Factor :					
27	Sample Amount :					
28	Spike Amount :					
29	Final Volume :					
30	Int. Std. active :	No				
31	IS before BS :	Mg25(LR)	Al27(LR)	P31(LR)	Sr86(LR)	Sr88(LR)
32						
33		Intensity AVG	Intensity AVG	Intensity AVG	Intensity AVG	Intensity AVG
34		[cps]	[cps]	[cps]	[cps]	[cps]
35						
36	Average	558139.6		59438.4	1067458.7	3216367.3
37	Std.Dev.	142134.9		301641.2	301428.3	5400652
38	RSD [%]	25.47		507.49	28.24	16.79
39						
40	Time [sec]					
41						
42		0.7	353.5	9496.5	45950.9	2244.6
43		1.4	101	7526.2	43674.3	2424
44		2.1	189.4	7323.9	41853.5	1907.8
45		4.8	448.9	7879.6	47773	1515
46	first.blankValue.linescan	144.3	9294.4	47317.5	1767.5	12730.2
47		...	0	6717.6	47773.5	2188.4
48		4.9	168.3	8334.5	46102.9	1464.5
49		5.6	189.4	8081.7	43573.1	1010
50		6.3	63.1	9142.6	45698.2	1683.3
51		7	151.5	8283.7	44028.5	1571.1
52		7.7	568.1	9041.8	46457	1717
53		8.4	454.5	8081.7	42662.6	1402.8
54		9.1	56.1	8536.4	48380	1414
55		9.8	252.5	8530.9	44180.5	2076.1
56		10.5	280.6	8283.7	45242.6	1161.5
57		11.2	224.4	8233.3	41397.8	1767.5
58		11.8	202	9193.3	44686.2	1963.9
59		12.5	112.2	8587	46760.2	1739.4
60		13.2	168.3	9041.6	41802.5	2020
61			101	8839.6	45698.2	2076.1
62	last.blankValue.linescan		168.3	9294.3	42259	1717
63			168.3	8637.5	45141.5	1851.7
64		18	202	8940.8	44493.3	1969.5
65		16.7	84.2	8586.9	48126.5	1111
66		17.4	221028	560708	642158.4	16272081.6
67		18.1	503582.4	814980.1	1231744	32426720.6
68		18.8	431705.6	4694470.9	1132627.2	30091390
69		19.5	469373.6	70404.7	1270345.6	28409467.6
70		20.2	429149.6	196346.1	1198758.4	26278938
71		20.9	421131.2	128177.9	1176188.8	31093709.6
72			7904.8	59812.7	1170502.4	28683005.4
73	first.sampleValue.linescan		2620.8	82989.3	1069083.2	29852783.8
74			8543.2	26782.5	1099766.4	29012144.2
75		23.7	353193.6	27743.2	1095865.6	28055164.6
76		24.4	389300	99033.4	1230832	26004250.2
77		25.1	413461.6	29009.6	1166097.6	29123816.2
78		25.8	369791.2	40019.9	1167104	29096197.4
79		26.5	409515.2	36600.5	1225414.4	29251566.4
80		27.2	390862.4	25063.1	1274144	30288647
1516		1027.8	734971.2	23293.4	1252499.2	36756310
1517			2763.2	24910.5	1150937.6	35421470
1518	last.sampleValue.linescan		578336	26329	1285363.2	36262152
1519			849392	24557.8	1253964.8	36459804
1520		1030.6	639203.2	37427.4	1243395.6	36454318
1521		1031.3	717182.4	24456.1	1367244.8	39085718
1522		1032	668681.6	21271.5	1277222.4	36578223.4
1523		1032.7	658579.2	21776.9	1286704	40920177.2
1524		1033.4	779214.4	32451.1	1460086.4	40968632.8
1525		1034.1	623105.6	30336.3	1119900.8	30969663.6
1526		1034.8	209322	10608.2	370260.8	8916298.2
1527		1035.5	30271.2	5101.3	77394.8	1507732
1528		1036.2	7425.2	4545.5	39627.3	232340
1529		1036.9	1666.5	4949.6	33457.1	49291.8
1530		1037.6	555.5	4191.9	32647.6	8789.2
1531		1038.3	63.1	4040.2	31585.4	6117.6
1532	number.sweeps.sample					
1533						
1534						
1535	Flags:	S=Amplifier Skipped	D=Intensity Defocused	O=Overflow		

Figure 4: Example of a raw data file obtained using the Thermo Scientific Element2 ICPMS opened in Microsoft Excel or OpenOffice. Highlighted are the lines of the “header”, where names of measured isotopes are found (“Line.of.Header”), and the beginning of the recorded signal (“Line.of.Signal”), which need to be specified in the script. Also highlighted is the column containing data of the isotope used as internal standard (“column.IS”, in this example, <sup>43</sup>Ca in the fifth column). Additionally, the lines of the first and last blank value used for the calculations are highlighted

(`first.blankValue.linescan` and `last.blankValue.linescan`), which need to be filled in as first and last value used for sample (`first.sampleValue.linescan` and `last.sampleValue.linescan`). The total number of lines recorded (`number.sweeps.sample`) is also indicated.

- The first and last value (e.g., the line number) used for the blank calculation (`first.blankValue.linescan` & `last.blankValue.linescan`) need to be filled in as well as the first and last value (e.g. the line number), which will be used for the calculation of the sample (`first.sampleValue.linescan` & `last.sampleValue.linescan`) (Figure 4). 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.
- The first and the last values (e.g., the line number) 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.
- 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.
- In the “Reference Material section” 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 Manuscript 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.
- Figure 5 shows an example, where two reference materials are used for calibration (e.g., NIST SRM 612 and USGS MACS-3). If only one reference material (for example only NIST612)

is used, RefMat line numbers 2 to 10 should start with a #. All reference materials used for calibration should be in directly successive lines, i.e., if 3 reference materials are evaluated (e.g. NIST612, MACS3 and NIST610), the # before RefMat3 must be deleted and “GSD” must be replaced by “NIST610”.

- If no reference material file containing the text string (e.g. “NIST610”) is placed in the “ReferenceMaterial\_directory”, an error will be produced (“Error in file... No such file or directory”). If the correct files are in the directory, but the wrong line is used, an error is produced (e.g., “Error in eval: object “RefMat3” not found”). If these errors occur, the user should check the content of the directory as well as the Reference material Section of the script.

```
##### Reference material Section

##### Values from the GeoReM database (www.http://georem.mpch-mainz.gwdg.de)
##### MPI-DING ATH0-G (ATH0), MPI-DING KL2-G (KL2-G)
##### MPI-DING StHs6/80-G (StHs), MPI-DING T1-G (T1-G)
##### BAM-S005B (BAM-B)
##### USGS MACS-1 (MACS1), USGS MACS-3 (MACS3), USGS GSD-G1 (GSD)
##### NIST SRM 610 (NIST610), NIST SRM 612 (NIST612)

RefMat1 <- "NIST612" ; Reference.Material <- c(RefMat1) # 1. referencematerial
RefMat2 <- "MACS3" ; Reference.Material <- c(Reference.Material,RefMat2) # 2. referencematerial
#RefMat3 <- "GSD" ; Reference.Material <- c(Reference.Material,RefMat3) # 3. referencematerial
#RefMat4 <- "KL2-G" ; Reference.Material <- c(Reference.Material,RefMat4) # 4. referencematerial
#RefMat5 <- "BAM-B" ; Reference.Material <- c(Reference.Material,RefMat5) # 5. referencematerial
#RefMat6 <- "T1-G" ; Reference.Material <- c(Reference.Material,RefMat6) # 6. referencematerial
#RefMat7 <- "MACS1" ; Reference.Material <- c(Reference.Material,RefMat7) # 7. referencematerial
#RefMat8 <- "StHs" ; Reference.Material <- c(Reference.Material,RefMat8) # 8. referencematerial
#RefMat9 <- "ATH0-G" ; Reference.Material <- c(Reference.Material,RefMat9) # 9. referencematerial
#RefMat10 <- "NIST610" | ; Reference.Material <- c(Reference.Material,RefMat10) # 10. referencematerial

##### End HelpValues #####
```

Figure 5: Reference material section with reference materials provided by TERMITE.

16. As a last step, the user should check the complete script again if all values are inserted correctly and then run the script either by executing the whole sequence or every single line individually. Running an R script can be performed either via a Terminal by executing the command ‘Rscript TERMITE\_spotscan.r’ or ‘R CMD BATCH TERMITE\_spotscan.r’ or by marking the whole script and pressing the RUN button (Rstudio).

### 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**.
- If you receive an **"Error in file: cannot open the connection"** together with an addition: **Warning message: cannot open file**, the user should first check if the spelling of the path and file names in the directory section is correct (Figure 2).
- If an **"Error in file: cannot open the connection with addition: Warning message: In file: cannot open file 'NA': No such file or directory"** occurs, the most likely case is a wrong spelling in the Directory section (Figure 2). Please also check the spelling of the directory on the hard-drive of your PC. Another reason for this error could be the spelling of the reference material files.
- An **"error in pdf cannot open file"** is most likely the result of a misspelled directory name. Therefore, the pdf cannot be written.
- An **"error in file(file, "rt") : invalid 'description' argument"** may occur when more than ONE file is present in the directory `Rawdata_linescan` (Figure 1).
- If other errors occur, please check again all values in the different sections of the script for correctness.
- 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 a size of several GB in size.