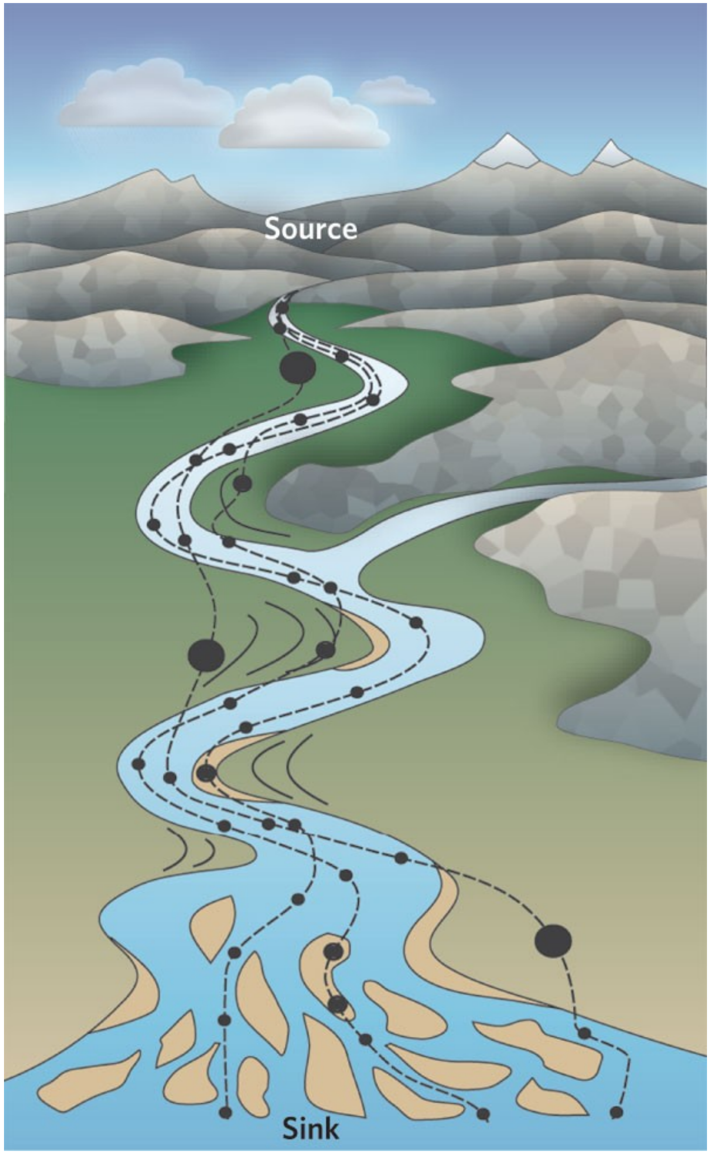
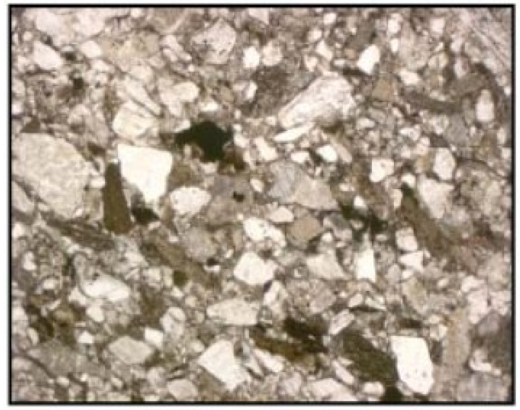
MAKING GEOLOGICAL SENSE OF 'BIG DATA' IN (SEDIMENTARY) GEOLOGY

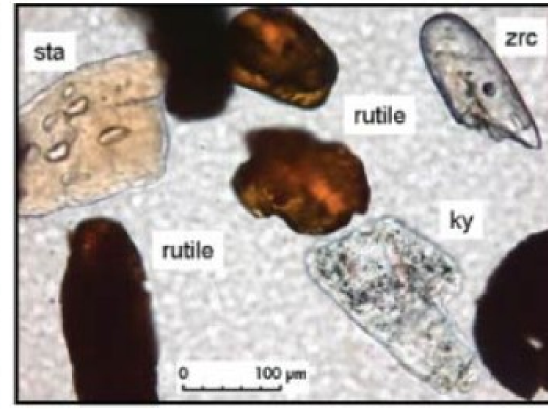
Pieter Vermeesch

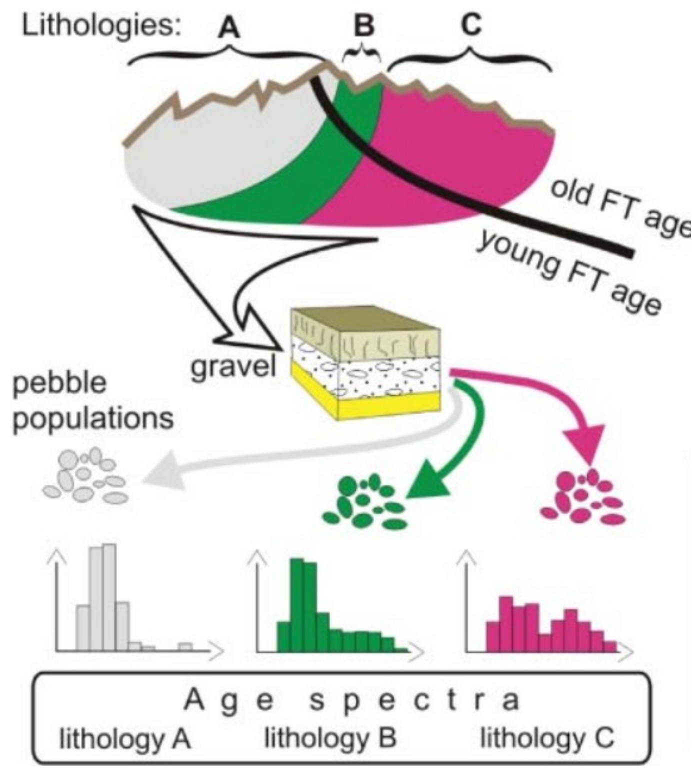
London Geochronology Centre
University College London
p.vermeesch@ucl.ac.uk



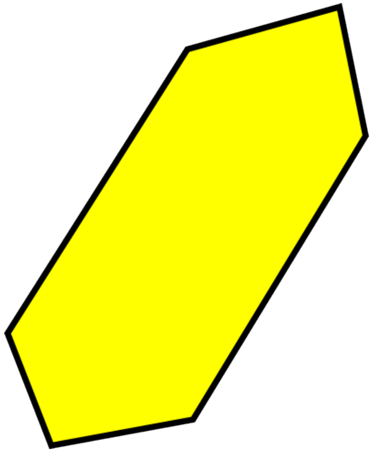
(Figure courtesy of P. Allen)

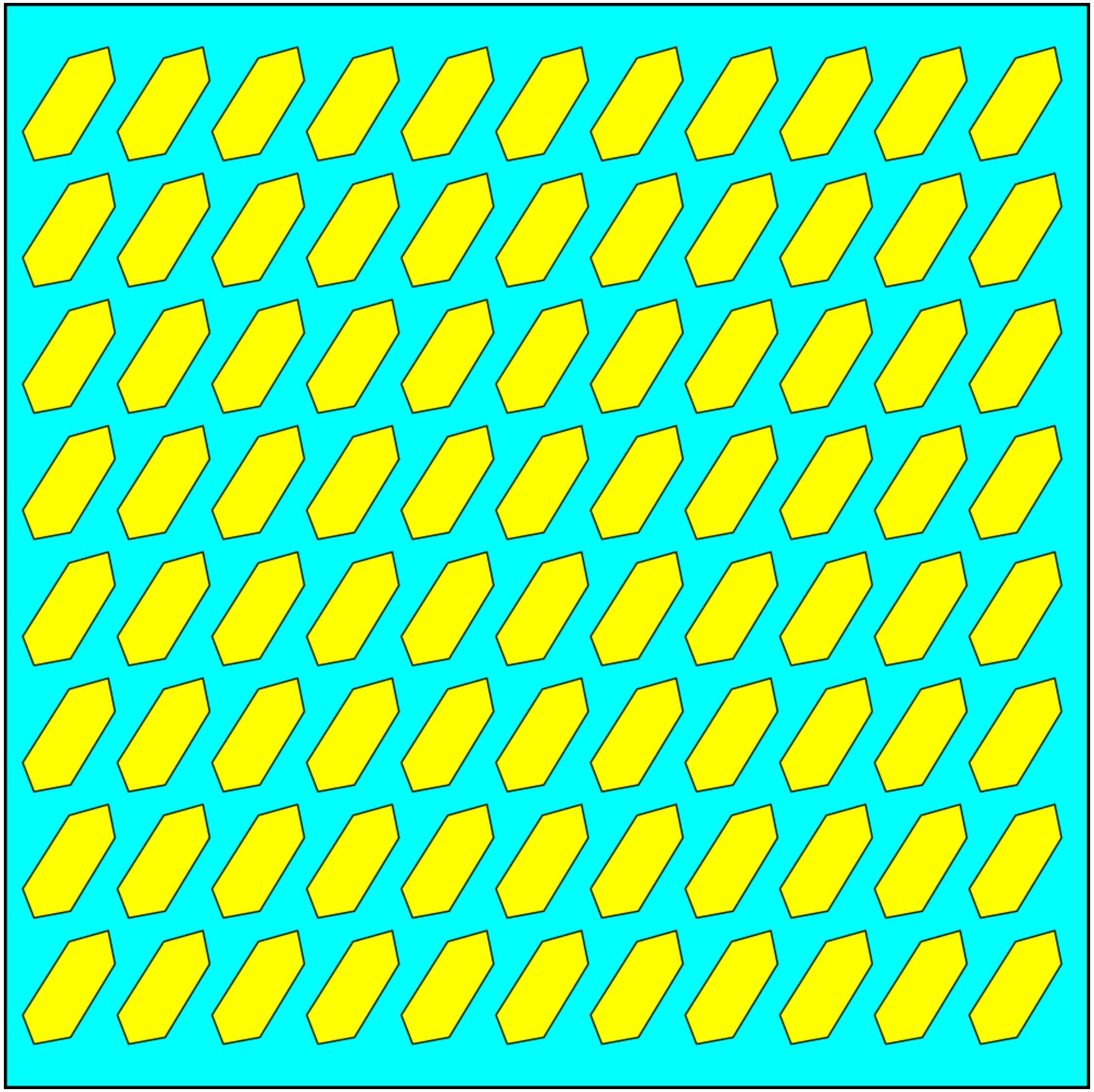




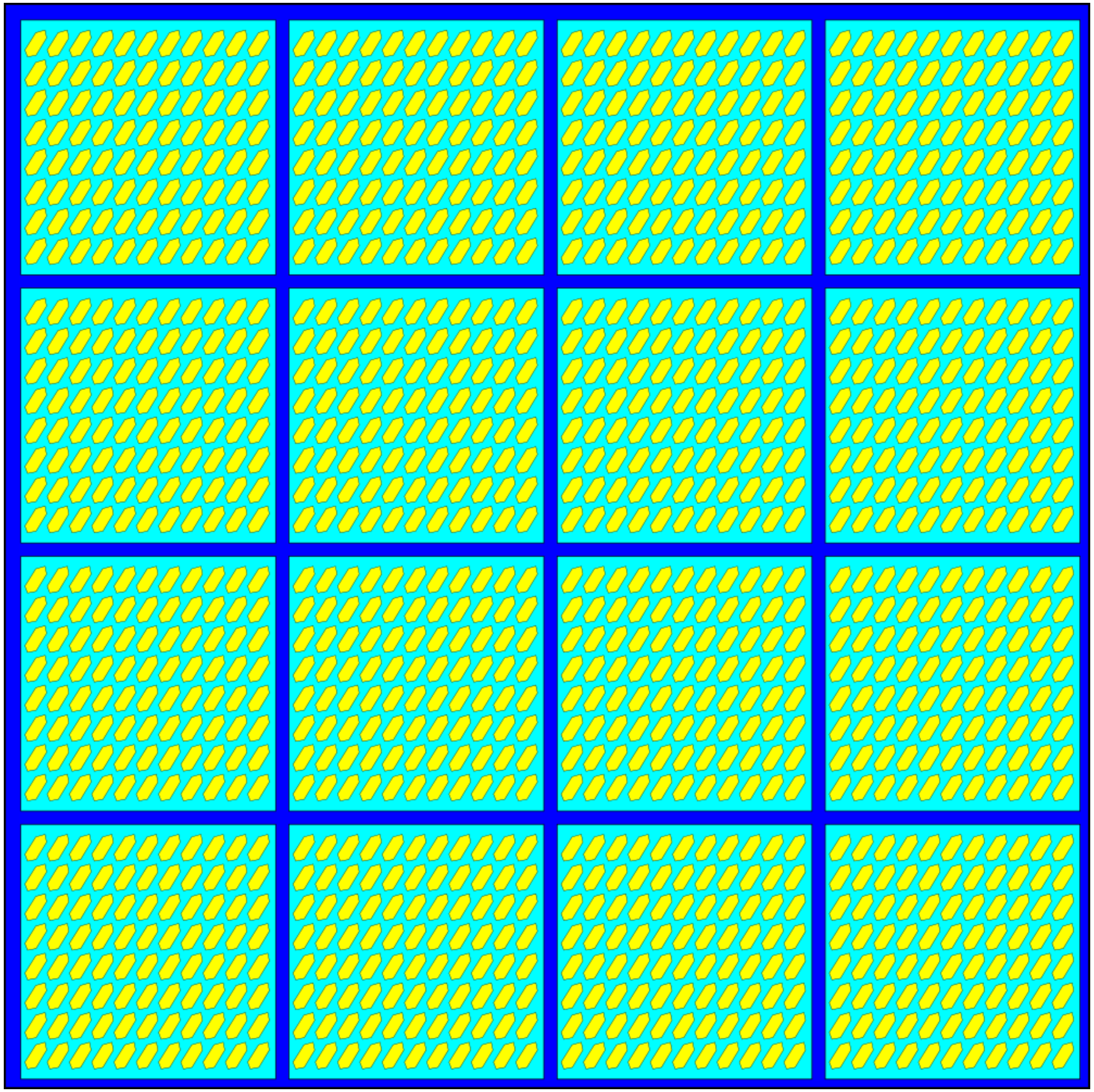


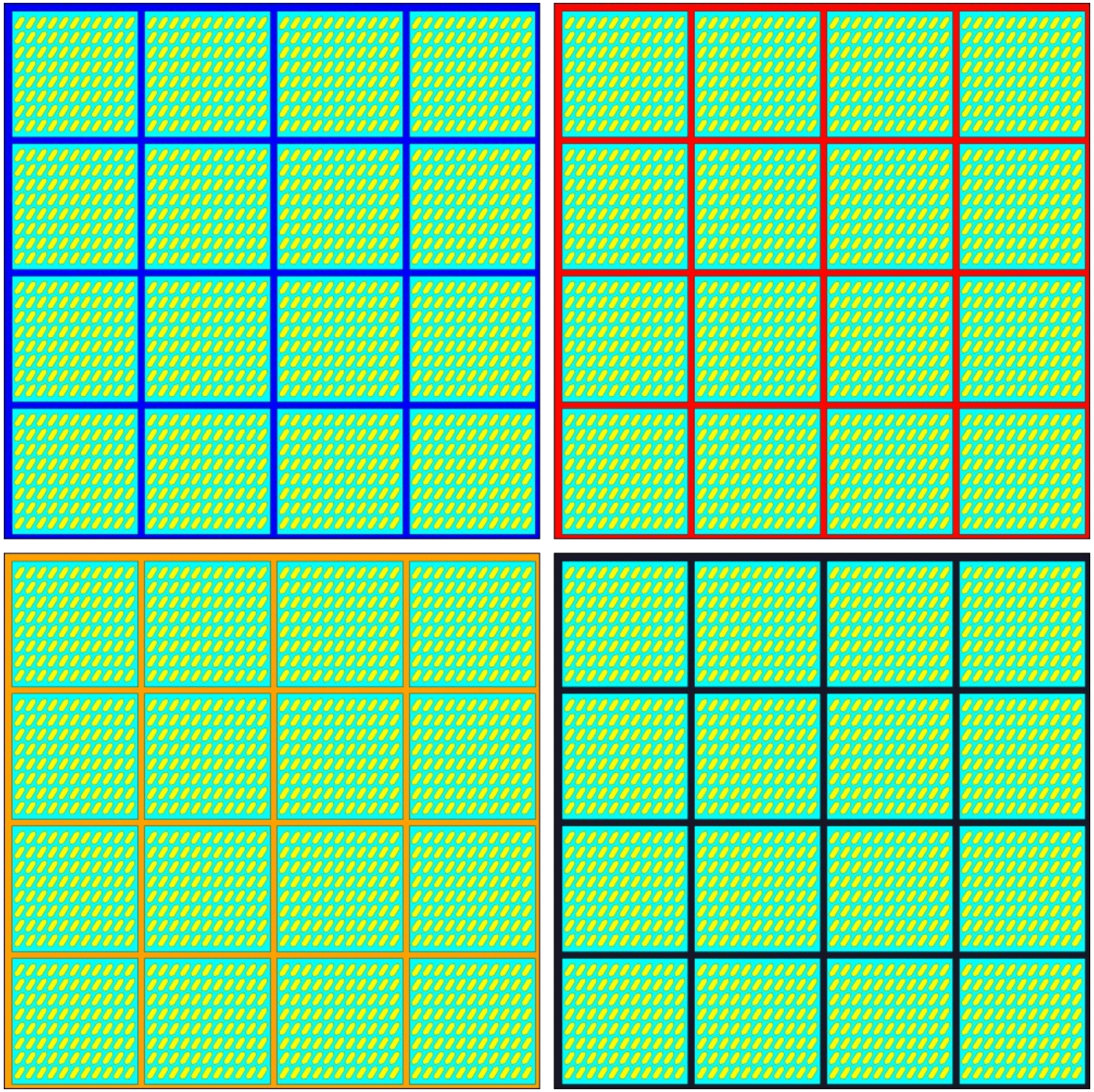
(Figure courtesy of H. von Eynatten)





Multiple samples of multiple grains





7/5	σ(%)	6/8	σ(%)	ρ	t (Ma)	2σ	7/5	σ(%)	6/8	σ(%)	ρ	t (Ma)	2σ	7/5	σ(%1)	6/8	σ(%)	ρ	t (Ma)	2σ
0.89	2.7	0.11	0.83	0.31	645.4	10.1	1.93	2.3	0.18	0.77	0.34	1068.2	14.9	2.06	3.8	0.19	0.88	0.24	1128.5	18.2
0.67	4.5	0.08	1.00	0.22	496.9	9.6	1.96	2.9	0.18	0.82	0.28	1079.0	16.3	1.27	6.8	0.13	1.30	0.19	788.6	19.3
1.75	3.1	0.17	0.82	0.26	1000.3	15.1	1.67	5.0	0.17	1.05	0.21	995.8	19.3	0.75	2.7	0.09	0.81	0.30	559.3	8.6
2.18	3.8	0.20	0.89	0.23	1168.5	18.8	1.88	1.6	0.18	0.74	0.45	1075.4	14.2	2.49	2.1	0.21	0.74	0.35	1246.1	16.5
8.54	4.4	0.42	0.91	0.21	2263.5	33.8	0.97	5.4	0.11	1.13	0.21	664.9	14.3	1.79	2.4	0.16	0.76	0.32	979.2	13.8
5.54	3.6	0.34	0.85	0.24	1878.6	27.1	0.83	2.3	0.10	0.78	0.34	620.9	9.2	0.71	4.1	0.09	0.97	0.24	535.7	10.0
1.13	3.9	0.13	0.91	0.23	769.8	13.1	0.29	1.7	0.04	0.73	0.44	267.1	3.8	0.72	10.0	0.09	1.91	0.19	546.5	20.0
2.23	3.1	0.20	0.82	0.26	1161.7	17.3	0.92	2.4	0.11	0.79	0.32	660.0	9.9	1.27	2.0	0.13	0.74	0.37	777.5	10.7
0.66	7.2	0.08	1.38	0.19	519.4	13.8	0.89	5.8	0.10	1.19	0.21	636.2	14.4	0.94	4.7	0.11	1.02	0.22	686.0	13.3
2.32	1.4	0.21	0.72	0.53	1213.3	15.1	1.91	2.1	0.18	0.75	0.36	1089.8	14.8	1.80	2.8	0.18	0.79	0.28	1040.1	15.1
0.32	3.8	0.04	0.91	0.24	271.3	4.8	0.87	1.8	0.10	0.74	0.41	637.0	8.9	1.73	2.3	0.17	0.76	0.32	1005.8	14.0
1.97	3.4	0.18	0.84	0.25	1065.4	16.5	0.80	3.2	0.10	0.86	0.27	598.8	9.9	0.82	4.9	0.10	1.07	0.22	597.4	12.2
2.15	3.4	0.19	0.85	0.25	1114.6	17.3	4.83	1.8	0.32	0.73	0.40	1794.2	21.4	5.13	2.6	0.32	0.77	0.29	1805.8	23.6
1.72	3.9	0.17	0.90	0.23	998.7	16.5	0.85	1.6	0.10	0.73	0.45	622.6	8.6	2.20	2.4	0.20	0.76	0.32	1157.9	16.0
0.85	4.1	0.10	0.93	0.23	603.3	10.7	0.78	4.4	0.09	1.03	0.24	566.5	11.2	0.69	2.8	0.08	0.81	0.29	520.4	8.1
0.58	9.5	0.07	1.83	0.19	465.9	16.4	0.72	2.7	0.09	0.82	0.31	543.5	8.5	0.86	2.8	0.10	0.80	0.29	612.6	9.4
0.61	3.9	0.08	0.91	0.23	489.4	8.6	1.84	1.5	0.17	0.71	0.48	1004.9	13.0	0.63	2.6	0.08	0.79	0.30	502.8	7.6
1.12	7.5	0.12	1.42	0.19	744.5	20.0	1.75	1.7	0.17	0.72	0.42	1013.3	13.3	0.70	3.3	0.09	0.90	0.27	538.1	9.2
2.50	1.5	0.22	0.73	0.47	1287.4	16.1	0.83	2.0	0.10	0.75	0.37	604.4	8.7	1.33	3.5	0.14	0.89	0.25	832.8	13.8
0.93	4.7	0.10	1.02	0.21	617.0	12.0	0.96	2.8	0.11	0.82	0.29	663.4	10.3	2.22	2.6	0.20	0.78	0.30	1182.3	16.7
4.85	4.9	0.32	0.98	0.20	1802.8	30.4	0.89	1.7	0.11	0.73	0.42	645.1	8.9	2.19	7.3	0.20	1.28	0.18	1187.1	27.7
2.10	2.8	0.20	0.82	0.29	1154.5	17.0	0.58	3.4	0.08	0.91	0.27	475.1	8.4	2.18	2.6	0.19	0.78	0.30	1146.0	16.2
0.66	3.3	0.08	0.90	0.27	511.3	8.8	0.14	2.0	0.02	0.74	0.37	128.9	1.9	2.14	3.3	0.19	0.83	0.25	1142.5	17.3
0.79	2.1	0.09	0.78	0.37	583.8	8.7	0.73	2.3	0.09	0.77	0.34	564.9	8.4	1.97	2.8	0.18	0.79	0.28	1082.8	15.7
0.23	3.2	0.03	0.91	0.28	209.5	3.7	19.65	7.8	0.59	1.15	0.15	2994.5	54.1	0.34	3.1	0.05	0.84	0.27	284.6	4.7
1.60	1.9	0.17	0.75	0.40	987.4	13.4	0.69	2.1	0.08	0.75	0.35	510.2	7.4	0.76	3.1	0.09	0.84	0.27	563.3	9.1
1.91	3.6	0.18	0.89	0.24	1080.5	17.6	0.88	3.5	0.10	0.89	0.25	629.5	10.6	0.67	3.8	0.08	0.91	0.24	516.3	9.0
1.51	1.8	0.16	0.74	0.42	930.2	12.6	1.78	6.0	0.17	1.17	0.20	1031.2	22.2	2.91	5.4	0.24	1.03	0.19	1366.4	25.3
0.68	2.1	0.08	0.78	0.37	518.4	7.7	1.46	2.3	0.15	0.77	0.33	903.2	12.8	1.79	3.1	0.18	0.82	0.27	1041.8	15.6
0.67	1.9	0.08	0.76	0.40	511.0	7.5	0.90	2.4	0.10	0.78	0.32	633.2	9.3	1.69	6.5	0.17	1.22	0.19	1016.8	22.9
2.08	3.2	0.19	0.84	0.27	1109.7	17.1	0.59	3.5	0.08	0.93	0.27	474.5	8.5	1.10	4.8	0.12	1.02	0.21	743.0	14.2
1.79	2.4	0.17	0.78	0.32	1035.7	14.8	1.95	2.7	0.18	0.79	0.29	1073.1	15.4	0.76	3.1	0.09	0.83	0.27	577.2	9.2
1.12	4.0	0.12	0.95	0.24	748.4	13.4	0.74	3.2	0.09	0.87	0.27	549.2	9.1	2.04	3.0	0.19	0.81	0.27	1110.6	16.4

Grain-age histogram (e.g., Kober *et al.*, EPSL, 1989)

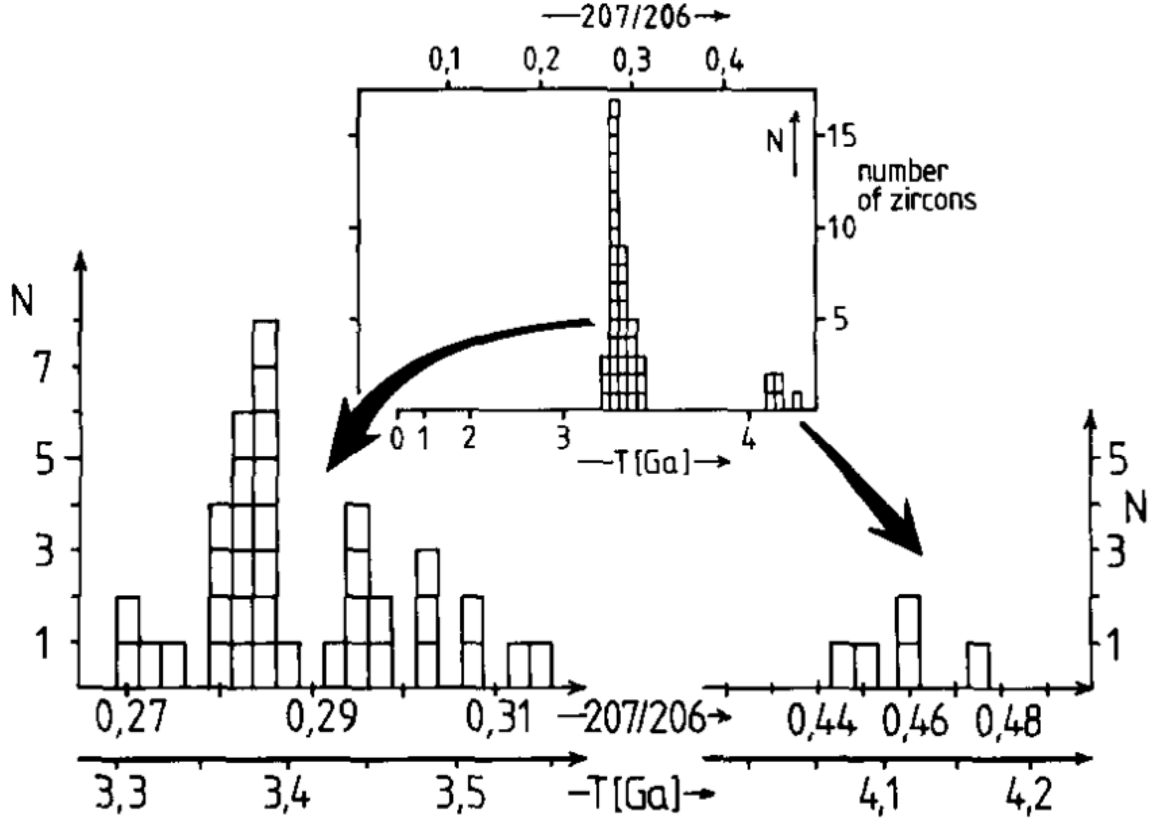


Fig. 4. Frequency histogram of 207/206 ages determined for 42 zircons by the evaporation method. The variations in the age distributions of both the Archean zircon subpopulations are far outside the routine data reproducibility of 1%. Widths of the apparent ages classes: $\Delta(207/206) = 0.01$ (inset), $\Delta(207/206) = 0.0025$ (expanded presentation).

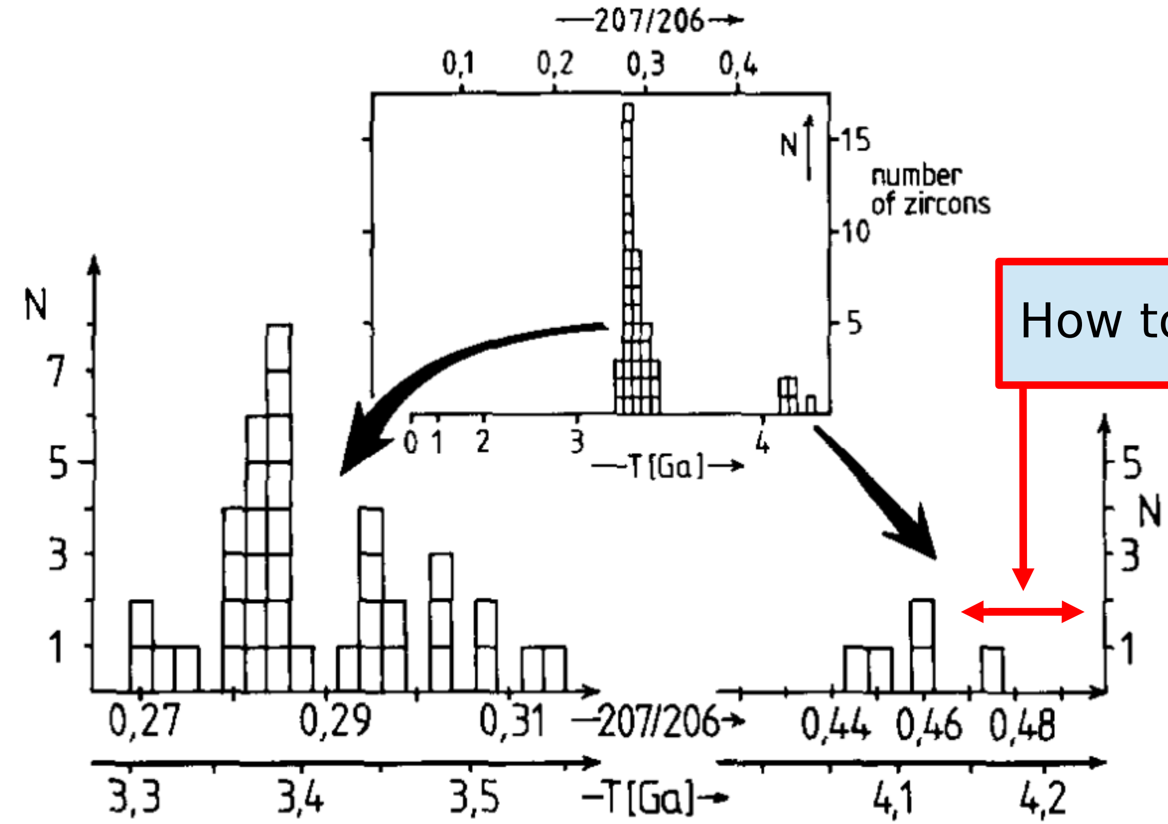


Fig. 4. Frequency histogram of 207/206 ages determined for 42 zircons by the evaporation method. The variations in the age distributions of both the Archean zircon subpopulations are far outside the routine data reproducibility of 1%. Widths of the apparent ages classes: Δ(207/206) = 0.01 (inset), Δ(207/206) = 0.0025 (expanded presentation).

How to place the bins?

$$k=[\log_2 n+1] \; (Sturges' \; Rule)$$

$$k=\sqrt{n} \; (Excel)$$

$$h=2\frac{IQR(x)}{n^{1/3}} \; (Freedman-Diaconis)$$

...

A search for ancient detrital zircons in Zimbabwean sediments

M. H. DODSON,[1*] W. COMPSTON,[1] I. S. WILLIAMS[1] & J. F. WILSON[2]

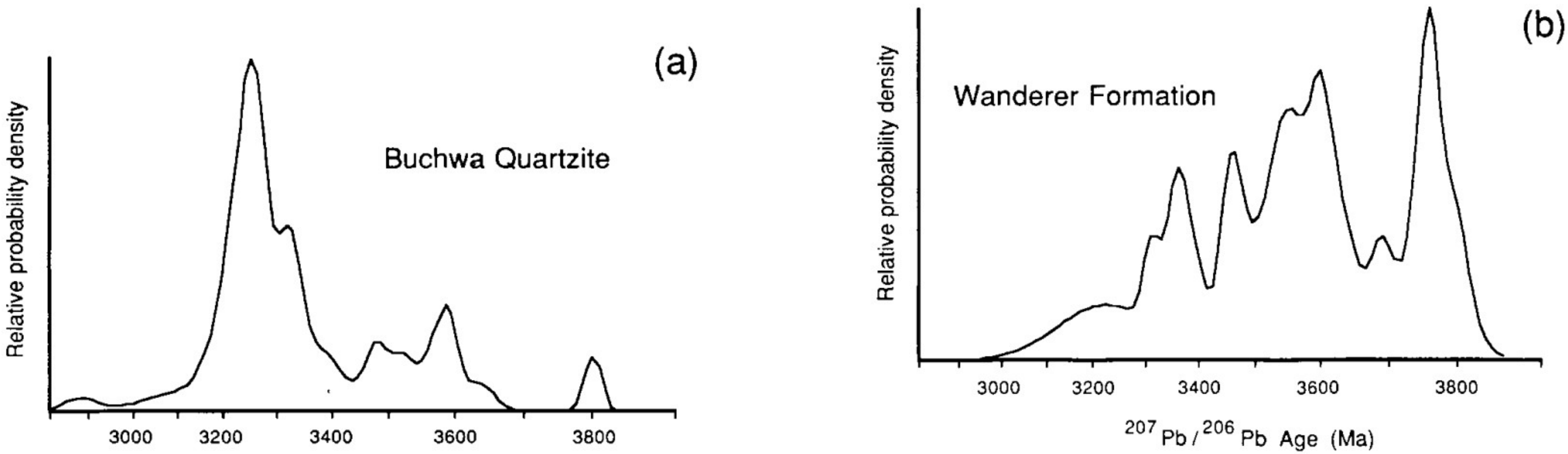
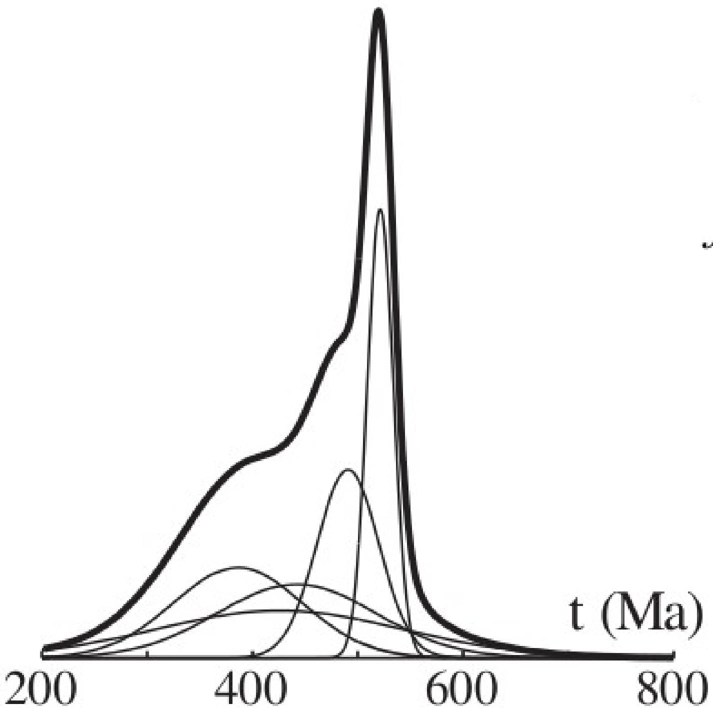


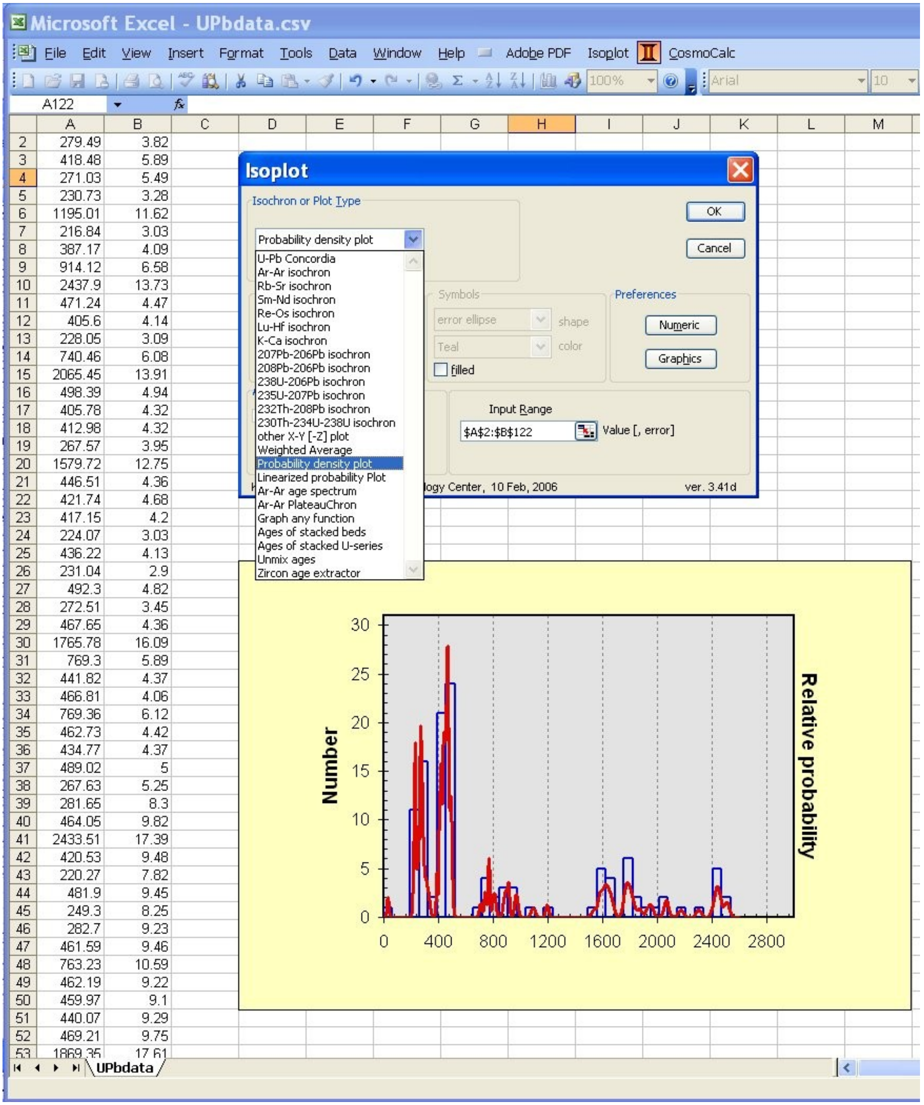
Fig. 3. 'Histograms' of $^{207}Pb/^{206}Pb$ ages of zircons extracted from (**a**) Buchwa Quartzite and (**b**) Wanderer Formation. The curves are obtained by summing Gaussian distributions of constant area, one for each age with standard deviation equal to the calculated uncertainty in the age. (Technique and program due to Dr P. Zeitler, ANU.)



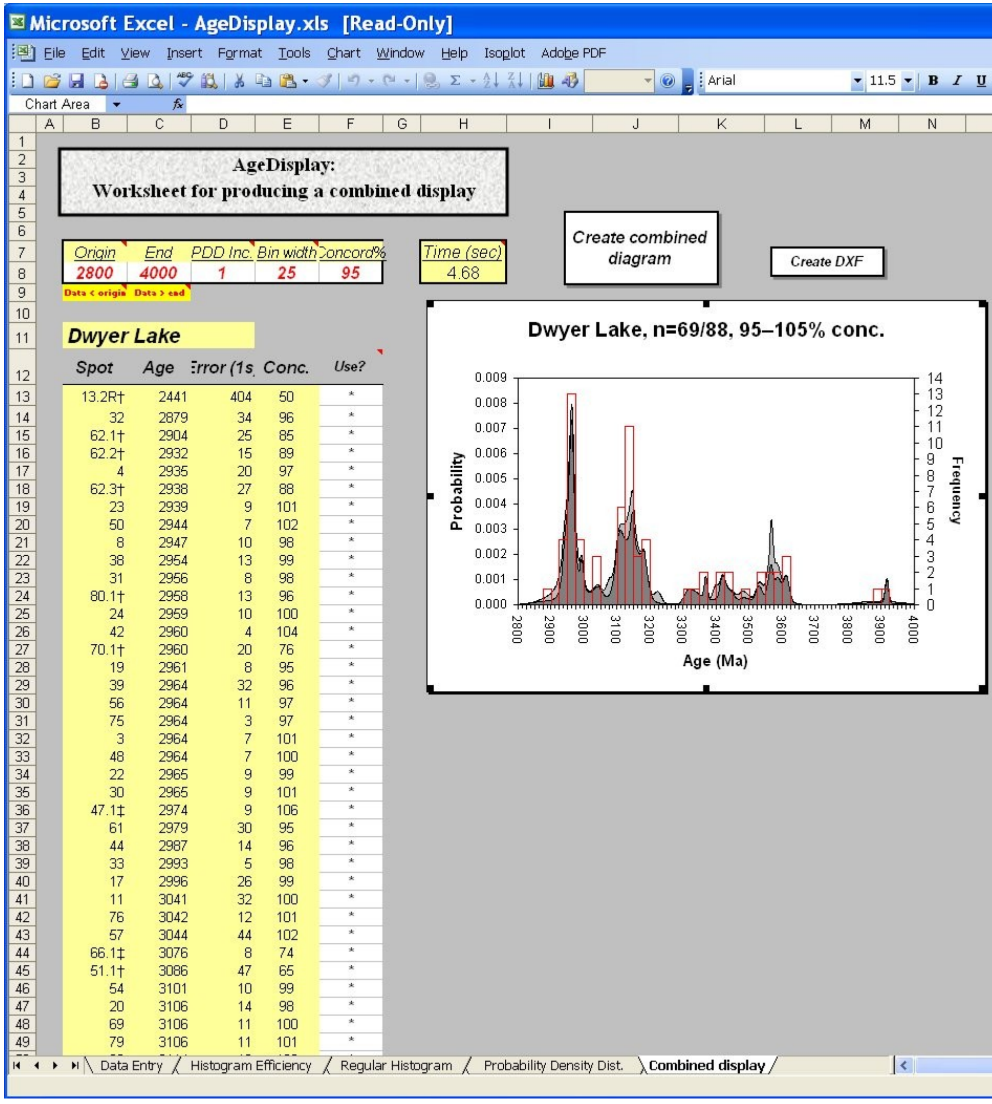
$$f(t) = \sum_{i=1}^{N} \frac{1}{e_i\sqrt{2\pi}} \exp^{-(t-x_i)^2/2e_i^2}$$

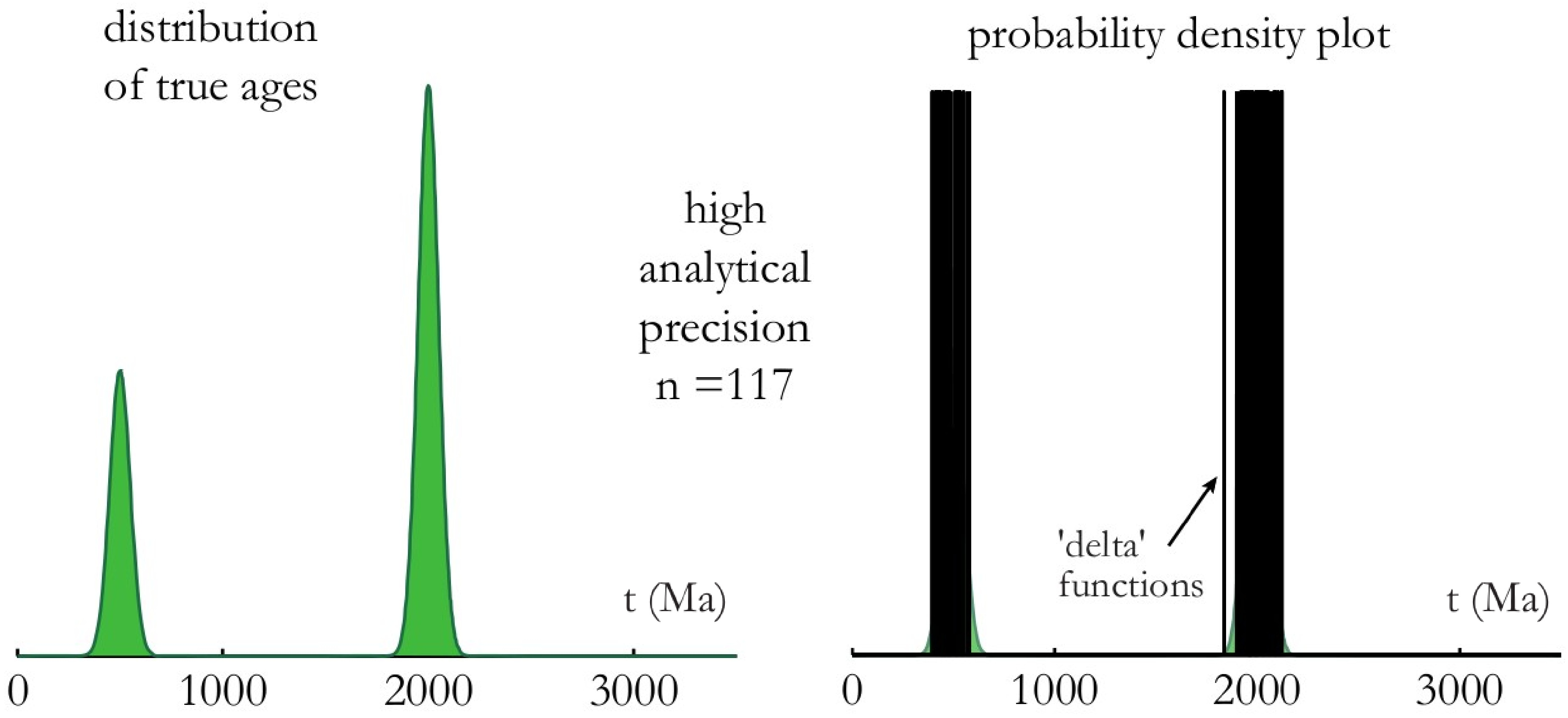
The PDP is implemented in user-friendly software

Isoplot (Ludwig, 2003)

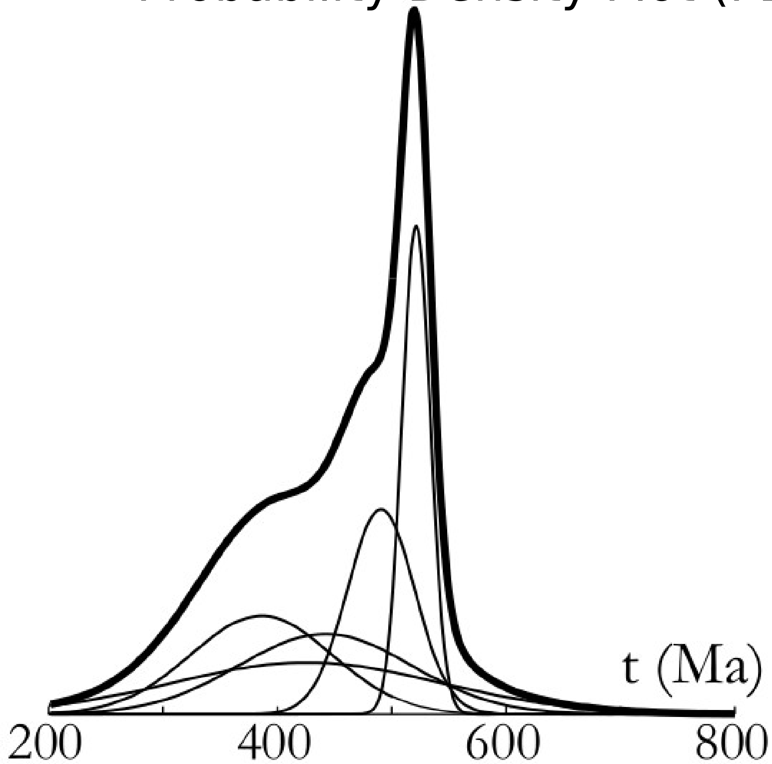


AgeDisplay (Sircombe, 2004)



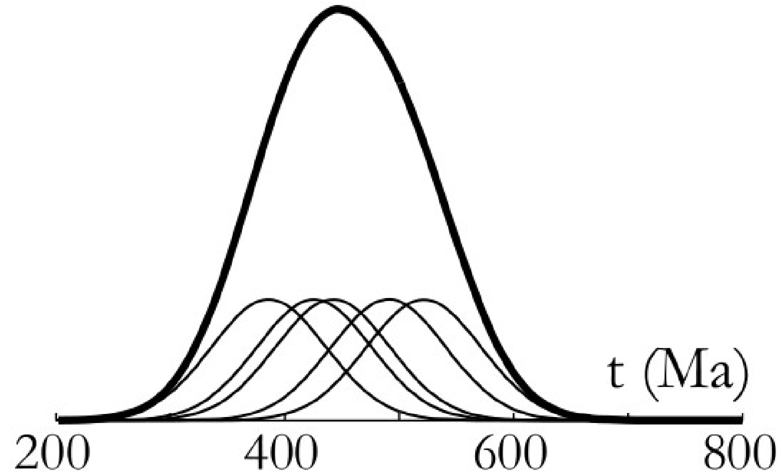

distribution
of true ages
t (Ma)
0
1000
2000
3000
high
analytical
precision
n =117
probability density plot
'delta'
functions
t (Ma)
0
1000
2000
3000

Probability Density Plot (PDP)



$$PDP(x) = \frac{1}{nh_i}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_i}\right)$$

Kernel Density Estimate (KDE)



$$KDE(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)$$

W Kernel density estimati ×

en.wikipedia.org/wiki/Kernel_density_estimation

Create account Log in

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox

Print/export
Create a book
Download as PDF
Printable version

Languages
Deutsch
Français
日本語
Українська
中文

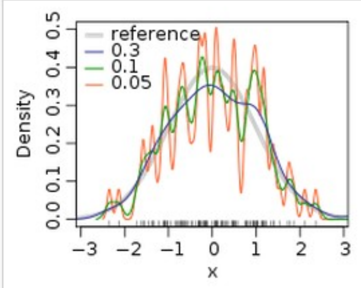
Article Talk Read Edit View history Search

Kernel density estimation

From Wikipedia, the free encyclopedia

It has been suggested that *Multivariate kernel density estimation* be merged into this article or section. (Discuss)
Proposed since September 2010.

In statistics, **kernel density estimation (KDE)** is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. In some fields such as signal processing and econometrics it is also termed the *Parzen–Rosenblatt window* method, after Emanuel Parzen and Murray Rosenblatt, who are usually credited with independently creating it in its current form.[1][2]



Kernel density estimation of 100 normally distributed random numbers using different smoothing bandwidths.

Contents [hide]

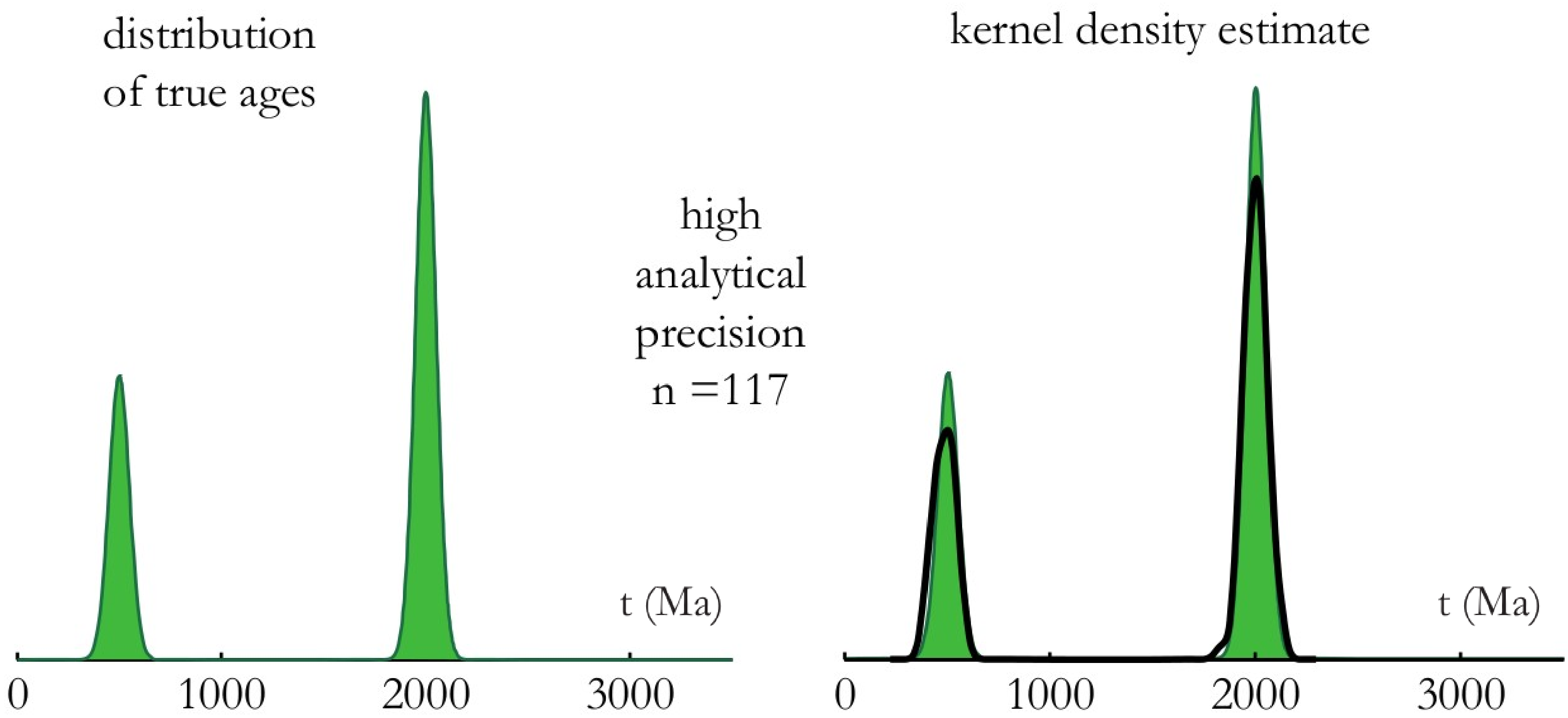
Definition

[edit]

Let $(x_1, x_2, \ldots, x_n)$ be an iid sample drawn from some distribution with an unknown density f. We are interested in estimating the shape of this function f. Its *kernel density estimator* is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \quad = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big),$$

where $K(\bullet)$ is the kernel — a symmetric but not necessarily positive function that integrates to one — and $h > 0$ is a smoothing parameter called the *bandwidth*. A


distribution
of true ages
kernel density estimate
high
analytical
precision
n =117
t (Ma)
0
1000
2000
3000
t (Ma)
0
1000
2000
3000

histograms

$$k = [\log_2 n + 1] (Sturges' \ Rule)$$

$$k = \sqrt{n} (Excel)$$

$$h = 2\frac{IQR(x)}{n^{1/3}} (Freedman - Diaconis)$$

...

KDEs

$$h = 1.06 \sigma n^{-1/5} (Silverman)$$

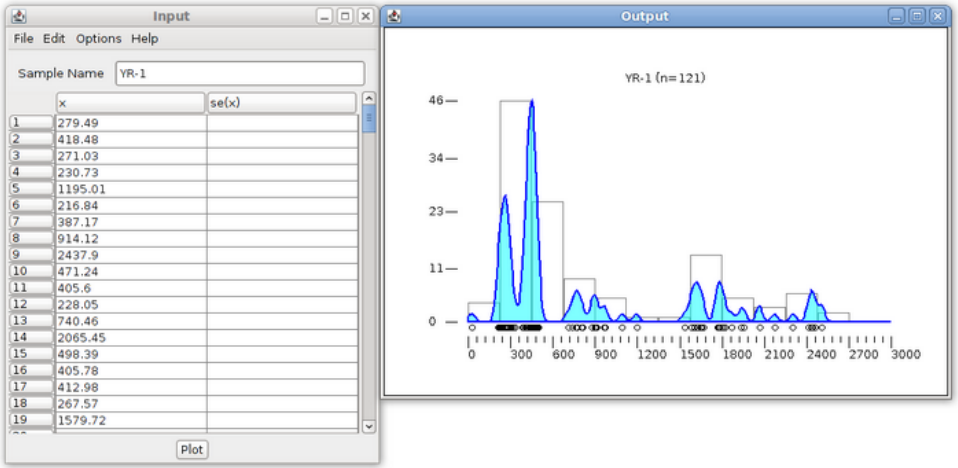
$$h_{AMISE} = \frac{R(k)^{1/5}}{m_2(K)^{2/5} R(f'')^{1/5} n^{1/5}}$$

adaptive KDE methods

...

DensityPlotter ×
www.ucl.ac.uk/~ucfbpve/densityplotter/

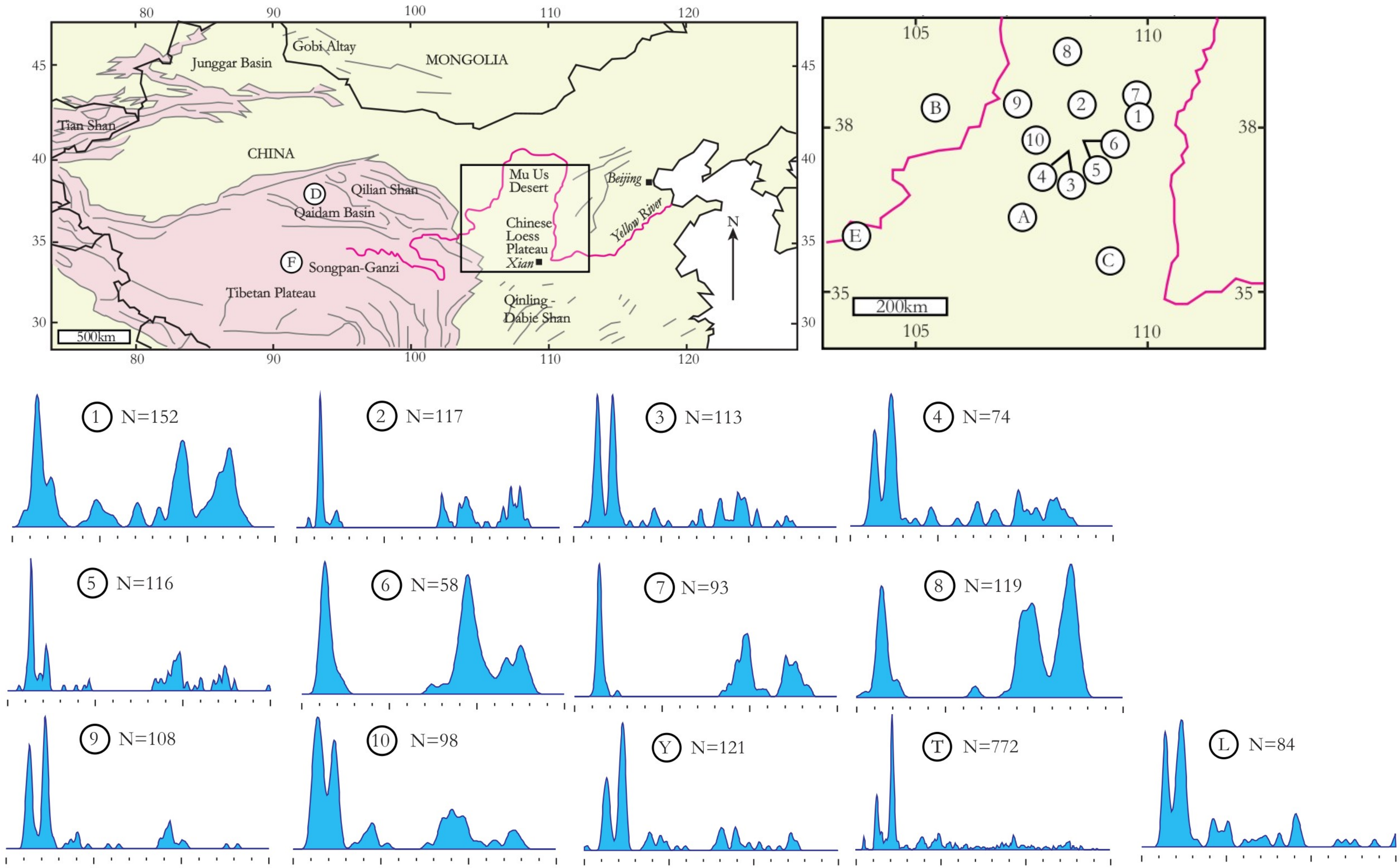
DensityPlotter - a Java application for Kernel Density Estimation



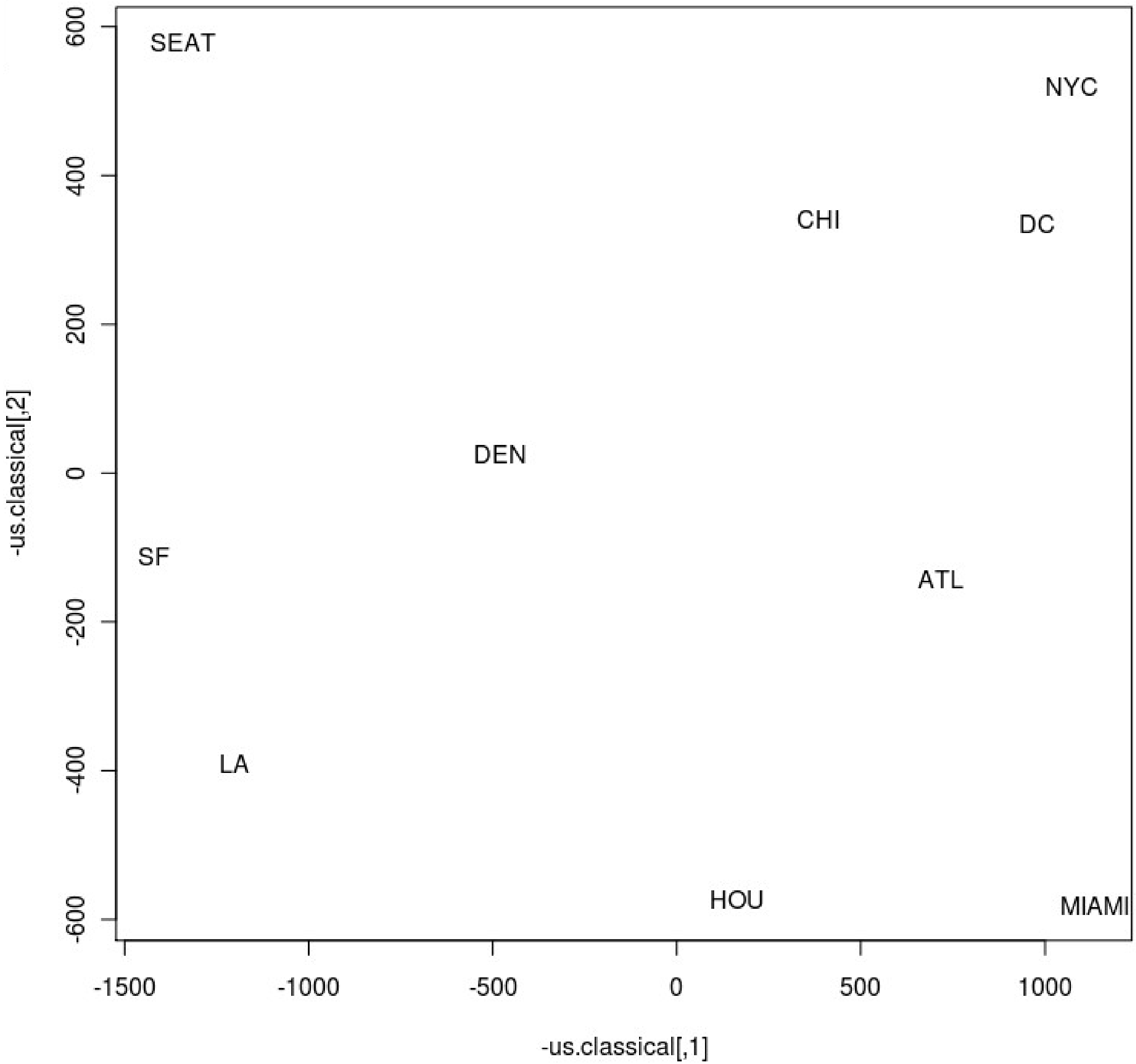
downloads	
DensityPlotter.jar Version 2.0 of the program earlier versions	Example input files FTdata.csv: Fission Tracks UPbdata.csv: U-Pb ages

`DensityPlotter` produces publication-ready kernel density estimates, probability density plots, histograms, radial plots and mixture models of (detrital) age distributions. The program is based on, and in fact offers exactly the same functionality as RadialPlotter albeit with a different set of pre-loaded preferences.

Citeable reference:
Vermeesch, P., 2012. On the visualisation of detrital age distributions. Chemical Geology, v.312-313, 19 0

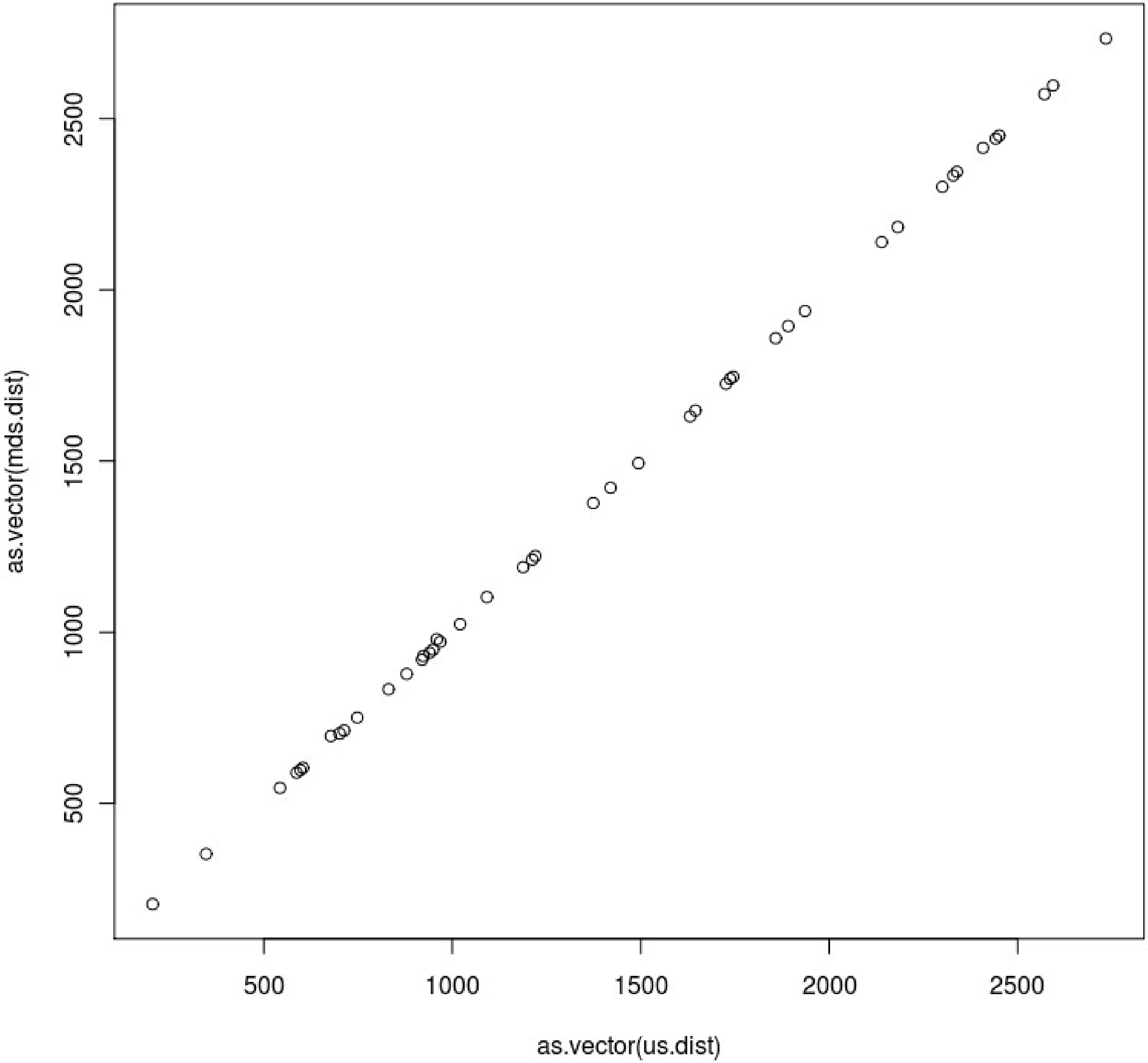

Junggar Basin
Gobi Altay
MONGOLIA
Tian Shan
CHINA
Qilian Shan
Qaidam Basin
Mu Us Desert
Beijing
Chinese Loess Plateau
Xian
Yellow River
Songpan-Ganzi
Tibetan Plateau
Qinling - Dabie Shan
N
500km
200km
1 N=152
2 N=117
3 N=113
4 N=74
5 N=116
6 N=58
7 N=93
8 N=119
9 N=108
10 N=98
Y N=121
T N=772
L N=84

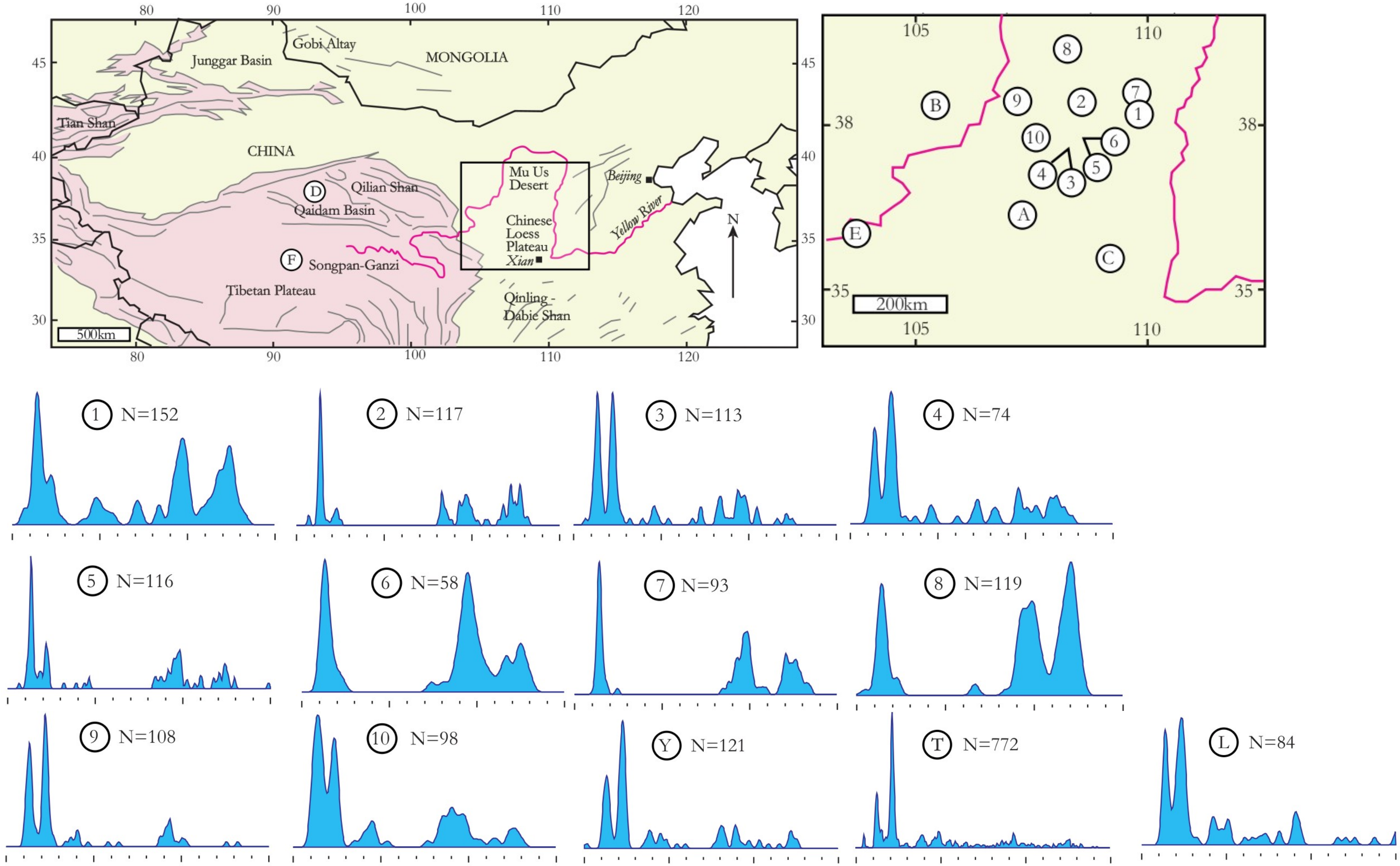
	ATL	CHI	DEN	HOU	LA	MIAMI	NYC	SF	SEAT	DC
ATL	0	587	1212	701	1936	604	748	2139	2182	543
CHI	587	0	920	940	1745	1188	713	1858	1737	597
DEN	1212	920	0	879	831	1726	1631	949	1021	1494
HOU	701	940	879	0	1374	968	1420	1645	1891	1220
LA	1936	1745	831	1374	0	2339	2451	347	959	2300
MIAMI	604	1188	1726	968	2339	0	1092	2594	2734	923
NYC	748	713	1631	1420	2451	1092	0	2571	2408	205
SF	2139	1858	949	1645	347	2594	2571	0	678	2442
SEAT	2182	1737	1021	1891	959	2734	2408	678	0	2329
DC	543	597	1494	1220	2300	923	205	2442	2329	0

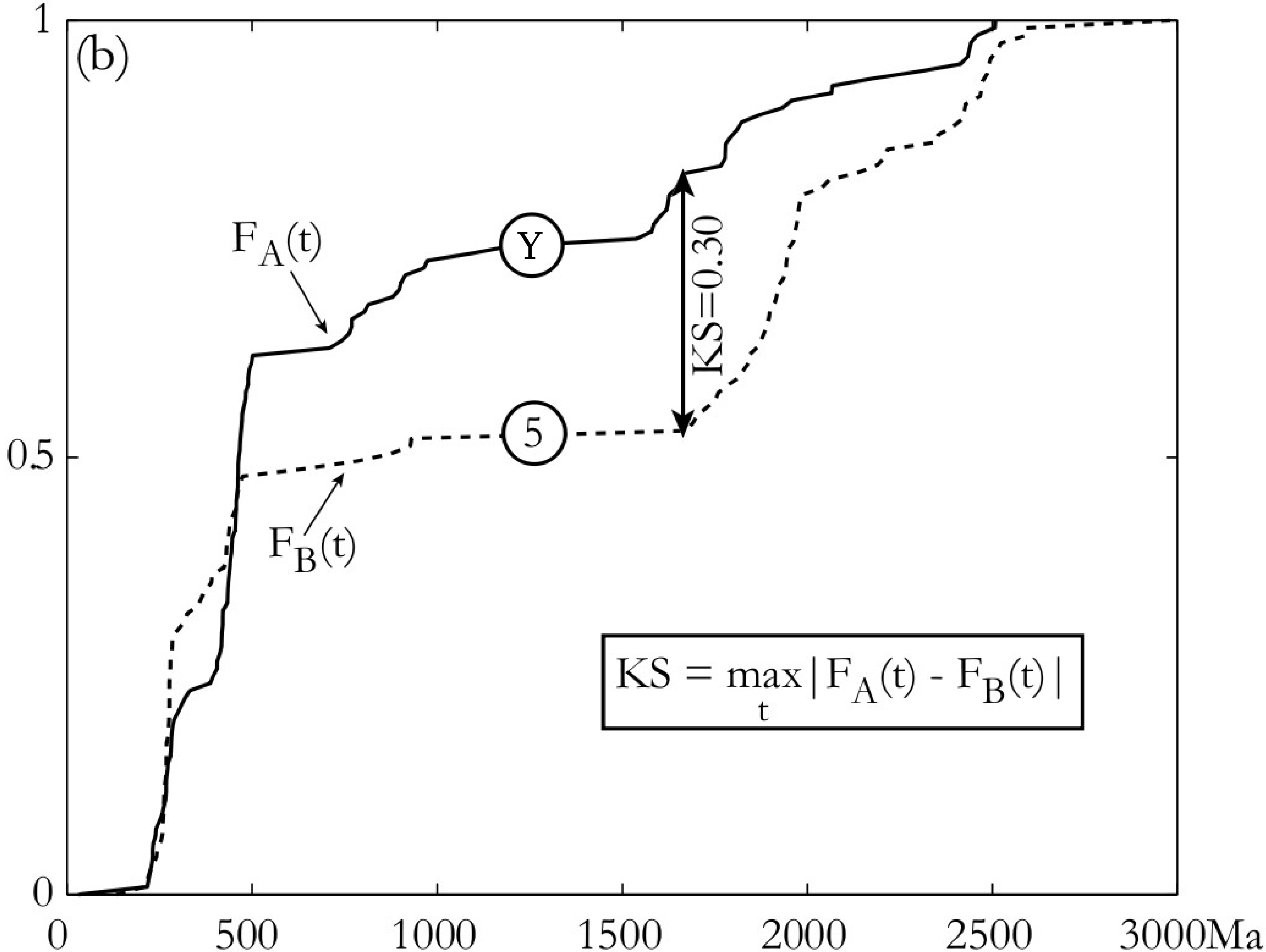

SEAT
NYC
CHI
DC
DEN
SF
ATL
LA
HOU
MIAMI
-1500
-1000
-500
0
500
1000
600
400
200
0
-200
-400
-600
-us.classical[,1]
-us.classical[,2]

The "Shepard plot" is obtained by plotting fitted distances vs. dissimilarities

$$d_{i,j} = \sqrt{(x_i^1 - x_j^1)^2 + (x_i^2 - x_j^2)^2 + \ldots + (x_i^R - x_j^R)^2}$$

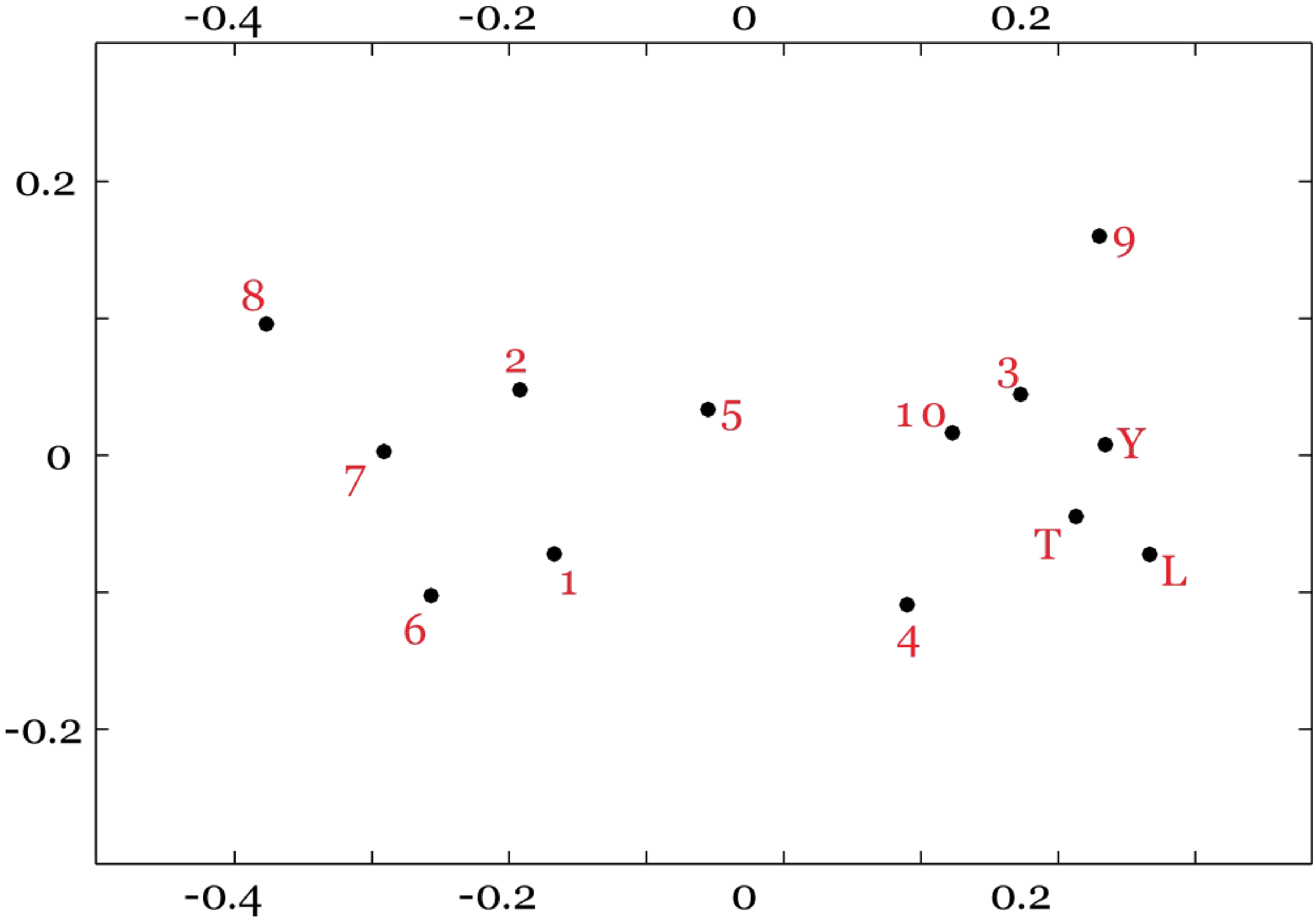



Gobi Altay
Junggar Basin
MONGOLIA
Tian Shan
CHINA
Mu Us Desert
Beijing
Qilian Shan
D
Qaidam Basin
Chinese Loess Plateau
Xian
Yellow River
N
F
Songpan-Ganzi
Tibetan Plateau
Qinling - Dabie Shan
500km
200km
1 N=152
2 N=117
3 N=113
4 N=74
5 N=116
6 N=58
7 N=93
8 N=119
9 N=108
10 N=98
Y N=121
T N=772
L N=84



(b)
$F_A(t)$
Y
KS=0.30
5
$F_B(t)$
$KS = \max_t |F_A(t) - F_B(t)|$
1
0.5
0
0
500
1000
1500
2000
2500
3000Ma

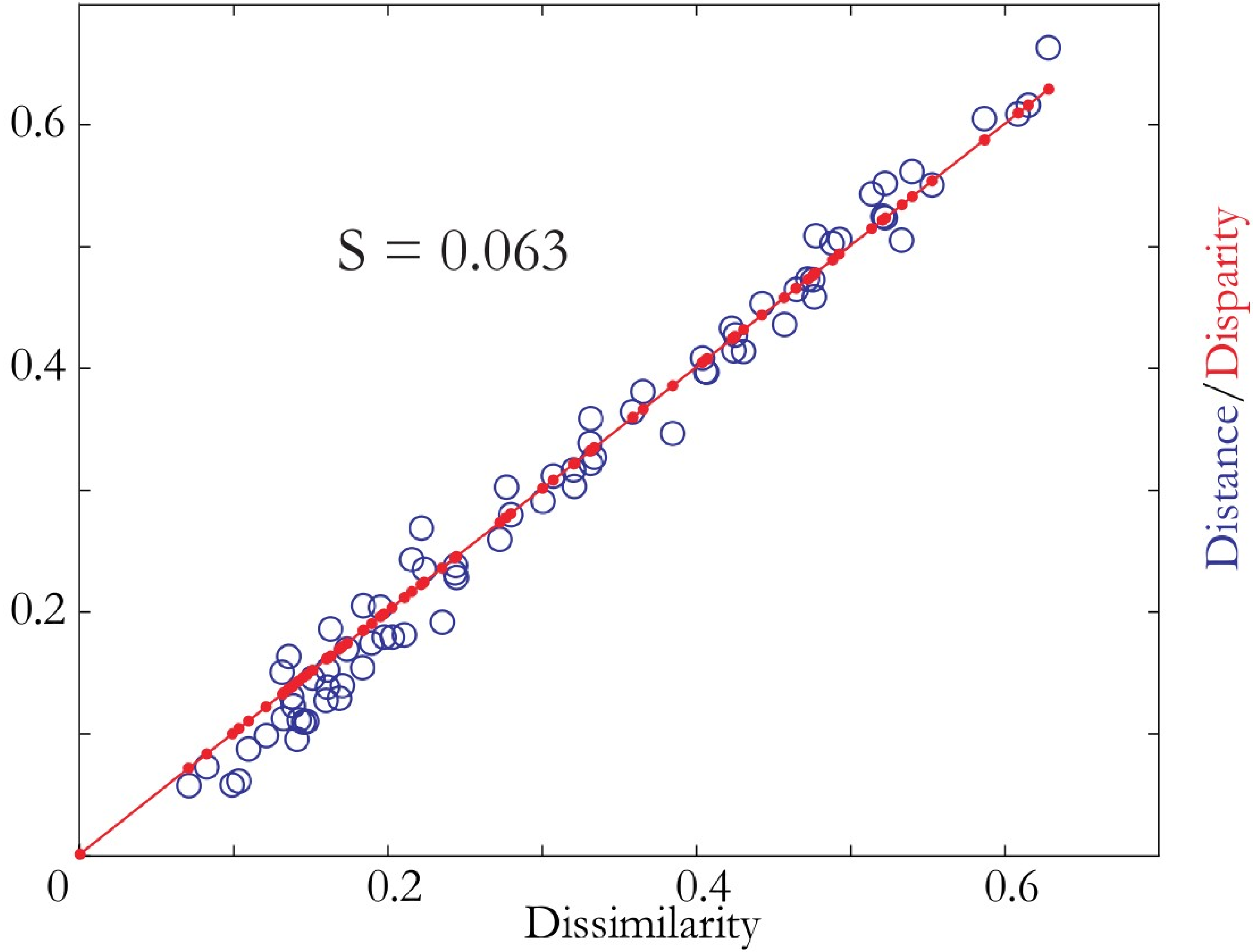
$$\delta = \begin{array}{cc}
 & \begin{array}{ccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & L & T & Y \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ L \\ T \\ Y \end{array} &
\begin{pmatrix}
0 & 14 & 33 & 27 & 18 & 14 & 15 & 22 & 48 & 32 & 42 & 37 & 40 \\
14 & 0 & 36 & 33 & 16 & 14 & 15 & 24 & 46 & 32 & 47 & 42 & 43 \\
33 & 36 & 0 & 19 & 24 & 44 & 47 & 55 & 17 & 10 & 13 & 12 & 8 \\
27 & 33 & 19 & 0 & 20 & 38 & 41 & 48 & 28 & 14 & 21 & 17 & 16 \\
18 & 16 & 24 & 20 & 0 & 22 & 24 & 33 & 31 & 20 & 33 & 28 & 30 \\
14 & 14 & 44 & 38 & 22 & 0 & 14 & 24 & 52 & 41 & 52 & 48 & 49 \\
15 & 15 & 47 & 41 & 24 & 14 & 0 & 16 & 51 & 43 & 54 & 49 & 52 \\
22 & 24 & 55 & 48 & 33 & 24 & 16 & 0 & 61 & 53 & 63 & 59 & 62 \\
48 & 46 & 17 & 28 & 31 & 52 & 51 & 61 & 0 & 20 & 22 & 18 & 16 \\
32 & 32 & 10 & 14 & 20 & 41 & 43 & 53 & 20 & 0 & 17 & 15 & 13 \\
42 & 47 & 13 & 21 & 33 & 52 & 54 & \boxed{63} & 22 & 17 & 0 & 10 & 11 \\
37 & 42 & 12 & 17 & 28 & 48 & 49 & 59 & 18 & 15 & 10 & 0 & \boxed{7} \\
40 & 43 & 8 & 16 & 30 & 49 & 52 & 62 & 16 & 13 & 11 & 7 & 0
\end{pmatrix}
\end{array}$$


-0.4
-0.2
0
0.2
0.2
0
-0.2
8
2
5
1 0
3
9
Y
7
T
L
1
6
4

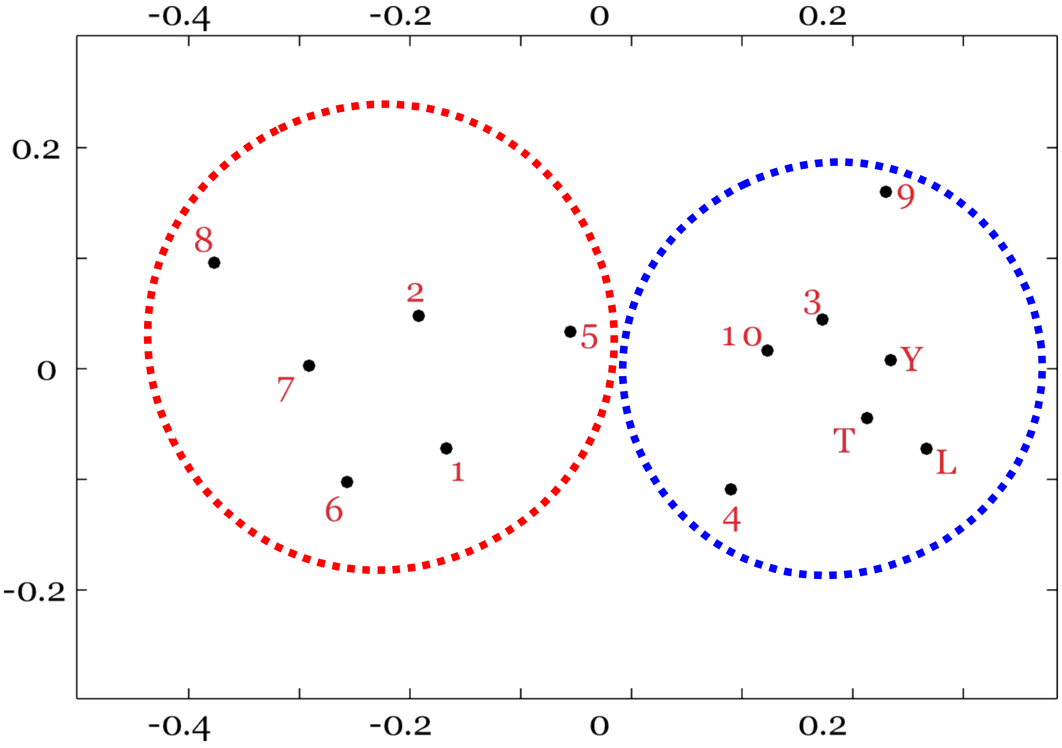
The fitted distances are in good agreement with the dissimilarities

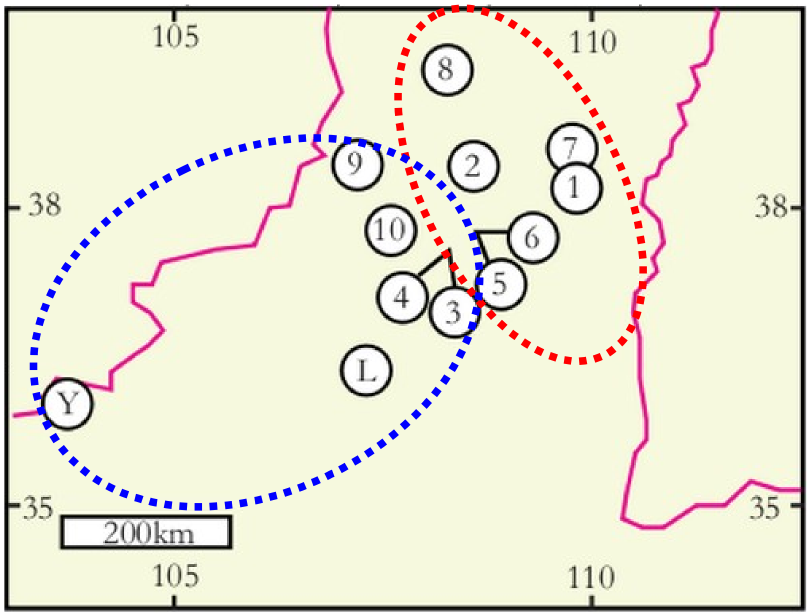
$$
d = \begin{array}{cc}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & L & T & Y \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ L \\ T \\ Y \end{matrix} &
\begin{pmatrix}
0 & 12 & 35 & 26 & 18 & 11 & 18 & 23 & 48 & 32 & 42 & 38 & 41 \\
12 & 0 & 36 & 30 & 14 & 7 & 9 & 18 & 44 & 33 & 45 & 40 & 43 \\
35 & 36 & 0 & 10 & 23 & 42 & 45 & 54 & 17 & 3 & 9 & 5 & 7 \\
26 & 30 & 10 & 0 & 18 & 35 & 39 & 47 & 27 & 8 & 16 & 12 & 15 \\
18 & 14 & 23 & 18 & 0 & 21 & 22 & 31 & 31 & 20 & 32 & 27 & 29 \\
11 & 7 & 42 & 35 & 21 & 0 & 7 & 13 & 51 & 39 & 50 & 46 & 48 \\
18 & 9 & 45 & 39 & 22 & 7 & 0 & 9 & 52 & 42 & 54 & 49 & 51 \\
23 & 18 & 54 & 47 & 31 & 13 & 9 & 0 & 61 & 50 & 62 & 58 & 60 \\
48 & 44 & 17 & 27 & 31 & 51 & 52 & 61 & 0 & 19 & 21 & 20 & 18 \\
32 & 33 & 3 & 8 & 20 & 39 & 42 & 50 & 19 & 0 & 12 & 8 & 10 \\
42 & 45 & 9 & 16 & 32 & 50 & 54 & \boxed{62} & 21 & 12 & 0 & 5 & 4 \\
38 & 40 & 5 & 12 & 27 & 46 & 49 & 58 & 20 & 8 & 5 & 0 & \boxed{3} \\
41 & 43 & 7 & 15 & 29 & 48 & 51 & 60 & 18 & 10 & 4 & 3 & 0
\end{pmatrix}
\end{array}
$$

The Shepard plot and stress value indicate a 'good' fit for the metric MDS



MDS analysis reveals a pronounced East-West dichotomy of the field area





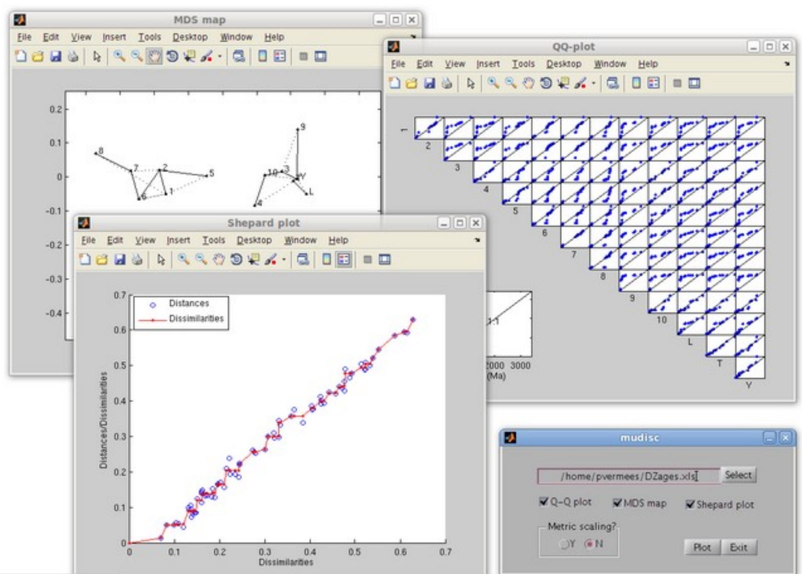
www.ucl.ac.uk/~ucfbpve/mudisc/

MuDiSc: Multi-Dimensional Scaling with Matlab and R

An increasing number of detrital zircon provenance studies are based on not just a few but many samples. This trend is likely to continue as the price of zircon U-Pb analyses continues to drop. The large datasets resulting from such studies call for a dimension-reducing technique such as Multi-Dimensional Scaling (MDS). Given a dissimilarity matrix (i.e., a table of pairwise distances), MDS constructs a 'map' on which 'similar' samples cluster closely together and 'dissimilar' samples plot far apart. This website presents some software tools for MDS analysis in the context of detrital geochronology, using the *effect size* of the two-sample Kolmogorov-Smirnov statistic as a dissimilarity metric. Two alternative sets of tools are presented here, written in Matlab (Section 1) and R (Section 2). MDS is closely related to, and in fact is a superset of, Principal Component Analysis (PCA), which is an established technique for conventional petrographic provenance studies. An example of this (written in R) is given at the end of this page, in Section 3. Further detail about these methods is provided in an accompanying paper:

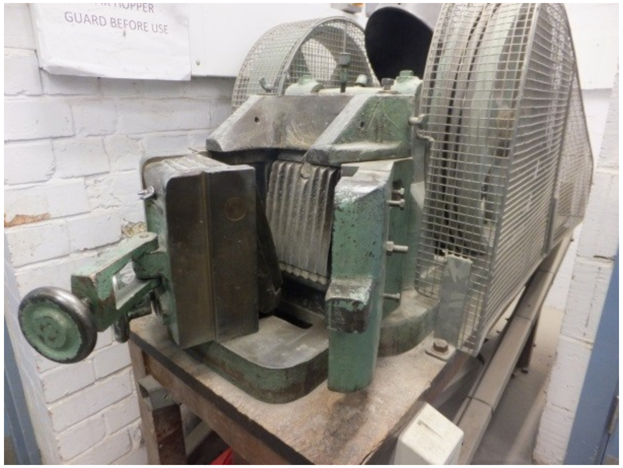
Vermeesch, P., 2013, Multi-sample comparison of detrital age distributions. Chemical Geology, doi:10.1016/j.chemgeo.2013.01.010.

1. A user-friendly Matlab-GUI:

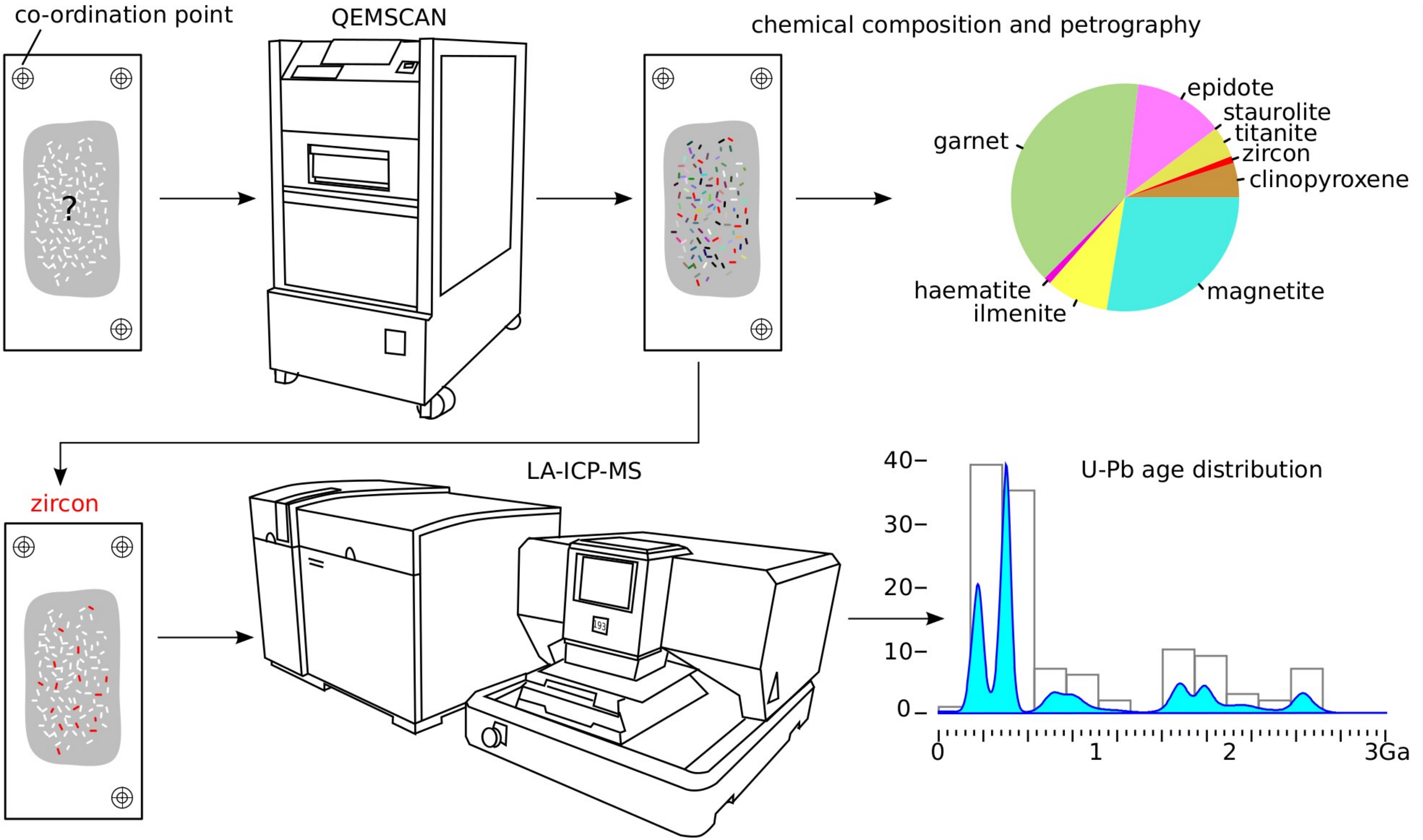


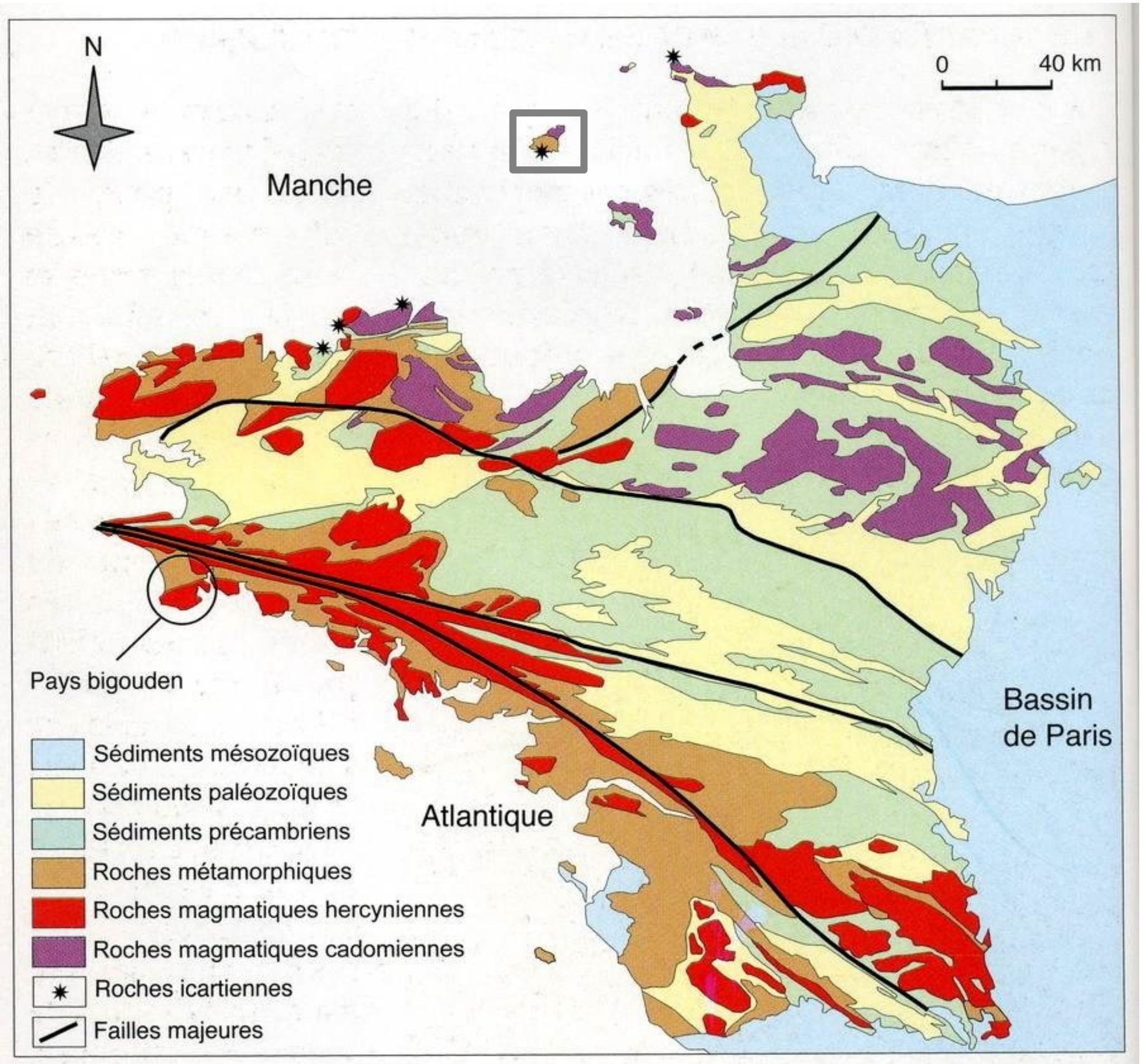
www.ucl.ac.uk/~ucfbpve/pictures/MuDiSc.png

Generating 'Big Data' is tedious and time consuming



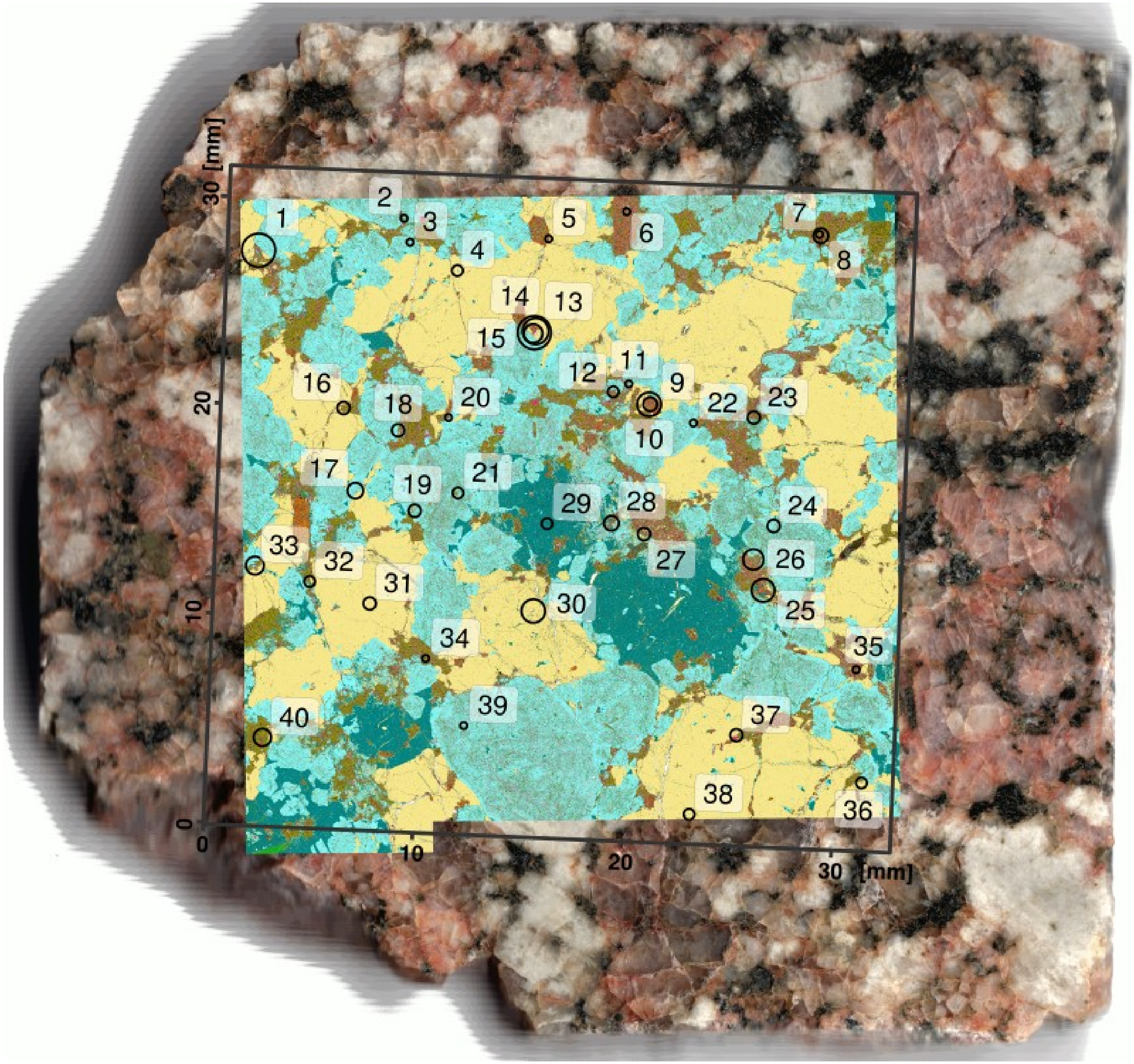



co-ordination point
QEMSCAN
?
chemical composition and petrography
epidote
staurolite
titanite
zircon
clinopyroxene
garnet
haematite
ilmenite
magnetite
zircon
LA-ICP-MS
193
40
30
20
10
0
U-Pb age distribution
0
1
2
3Ga

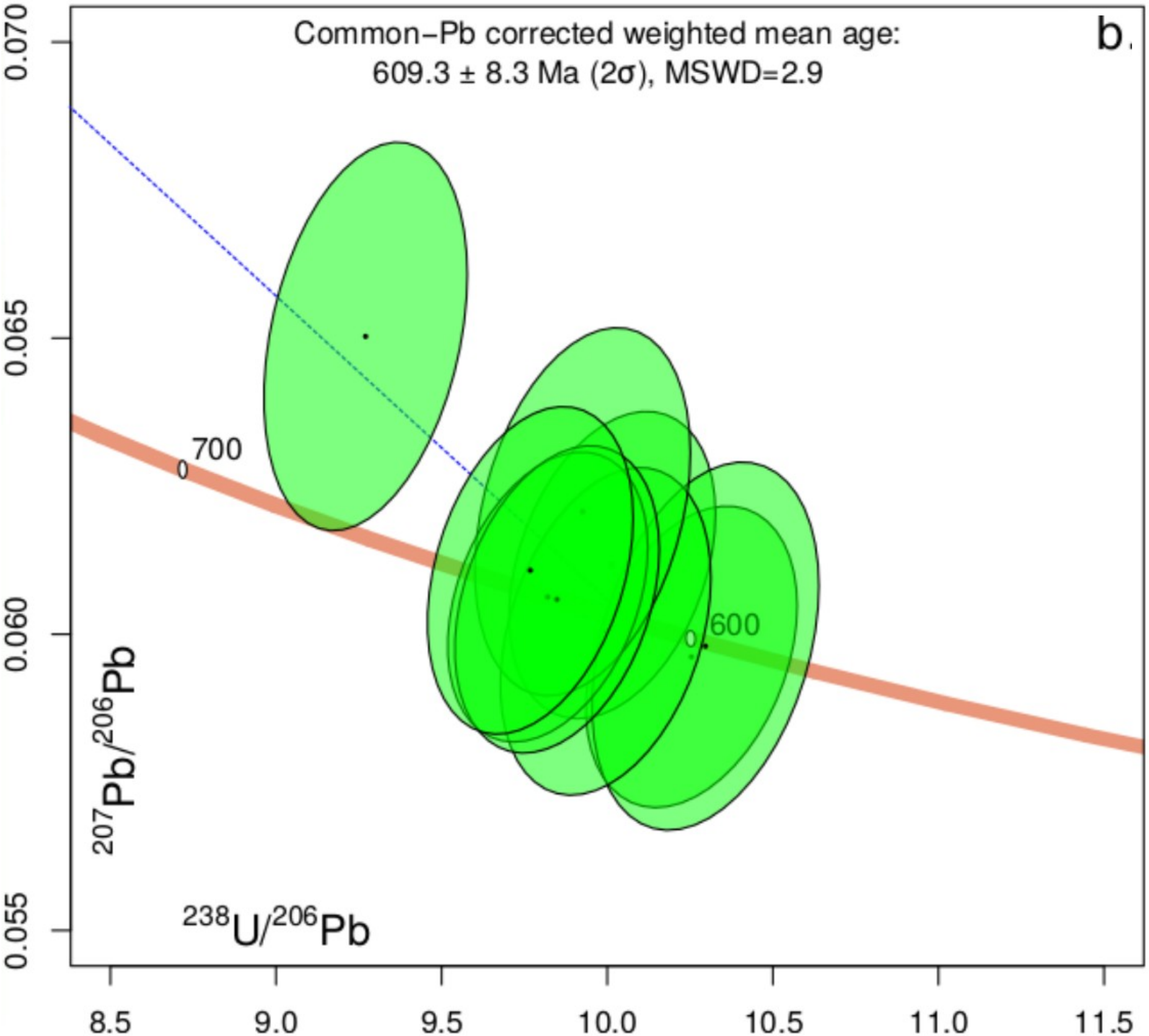


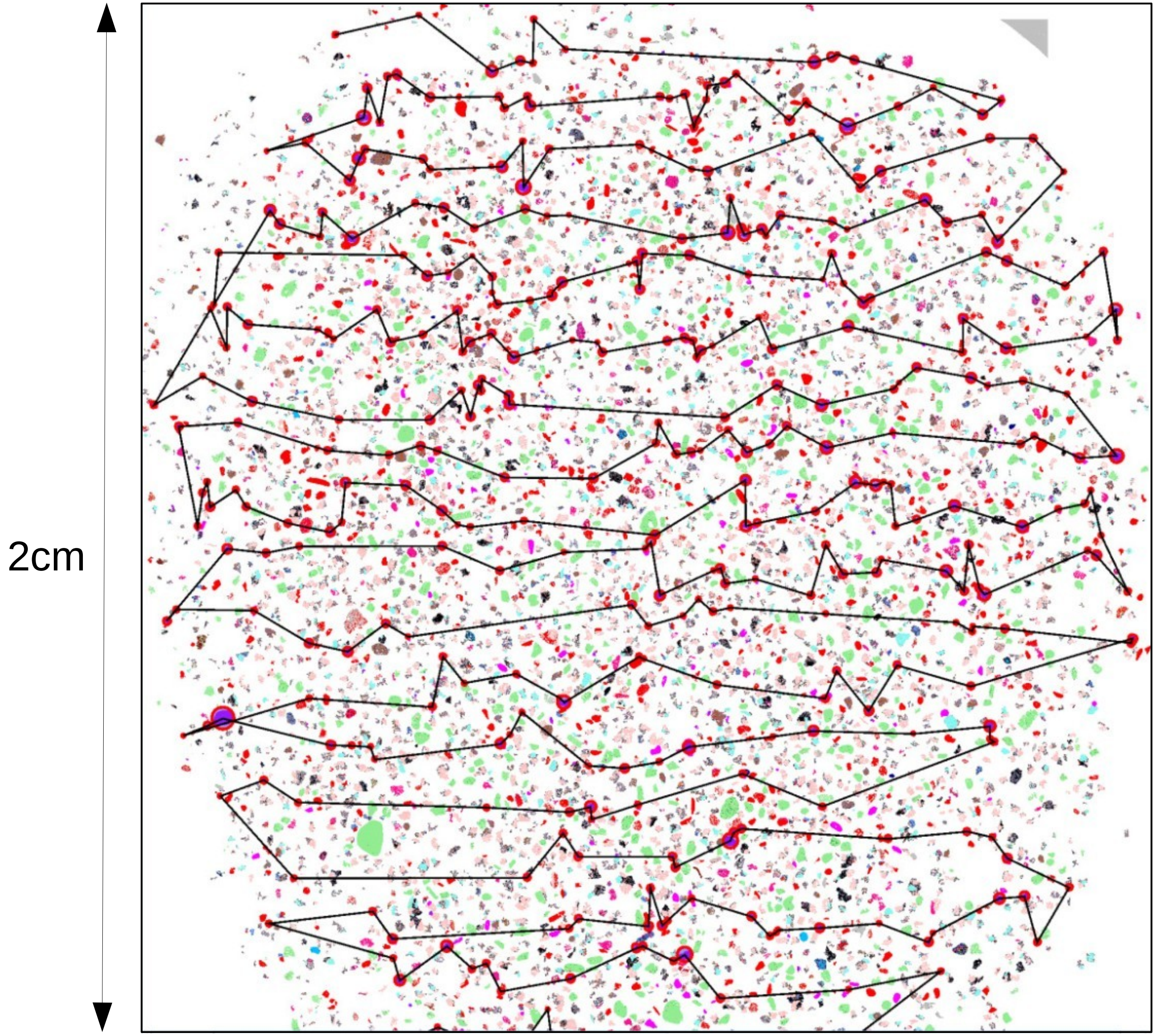
Carte géologique simplifiée du Massif armoricain

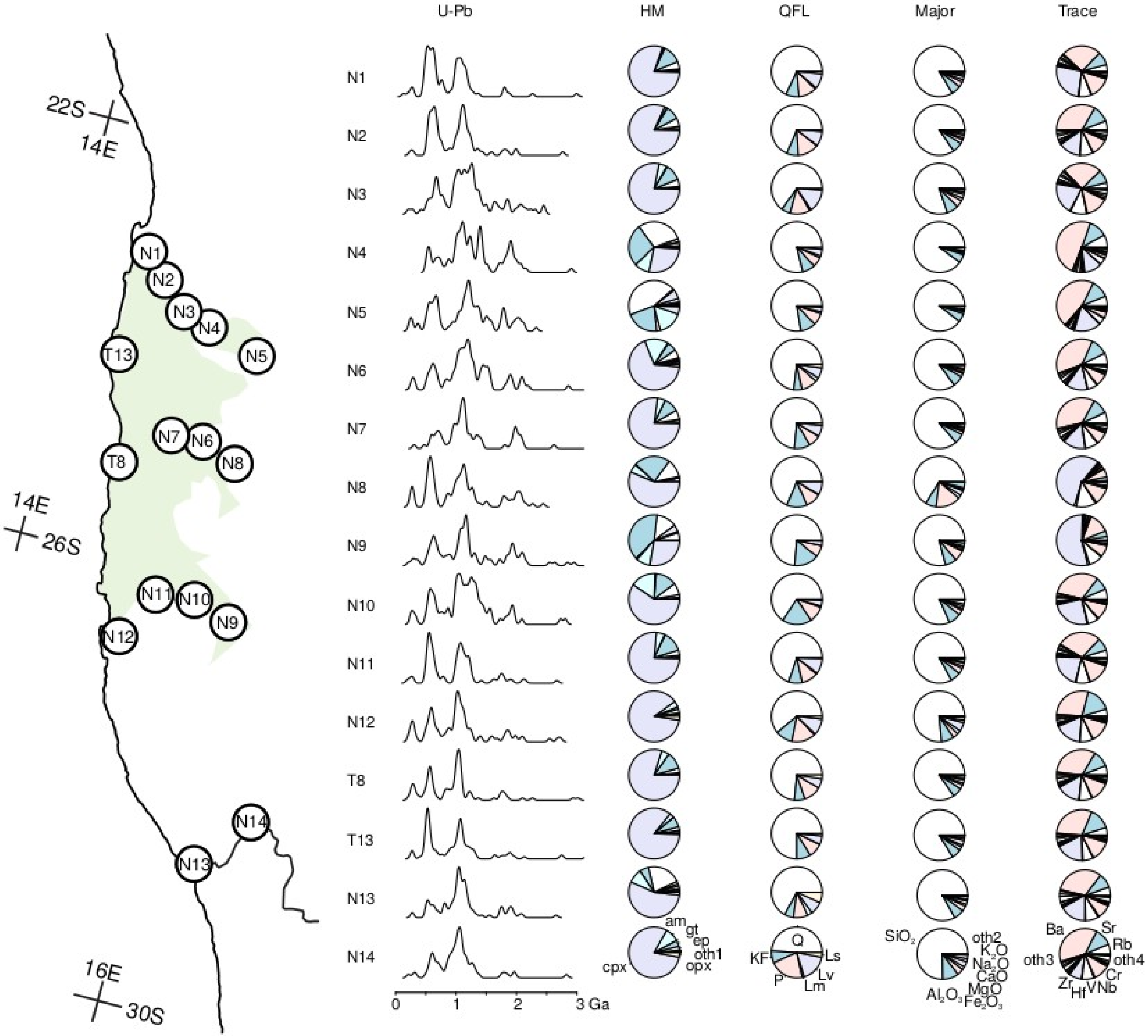
QEMSCAN map of a polished tile of l'Eree Granite



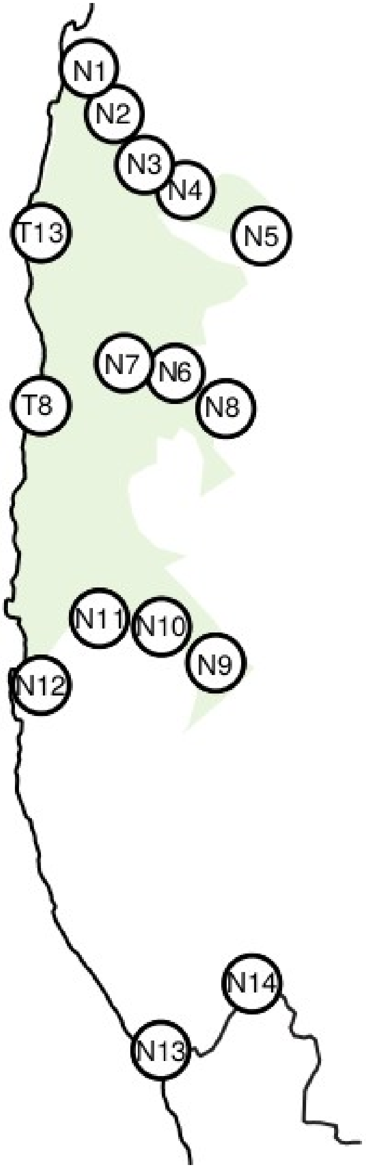
Zircon U-Pb dating confirms a Cadomian age for this rock

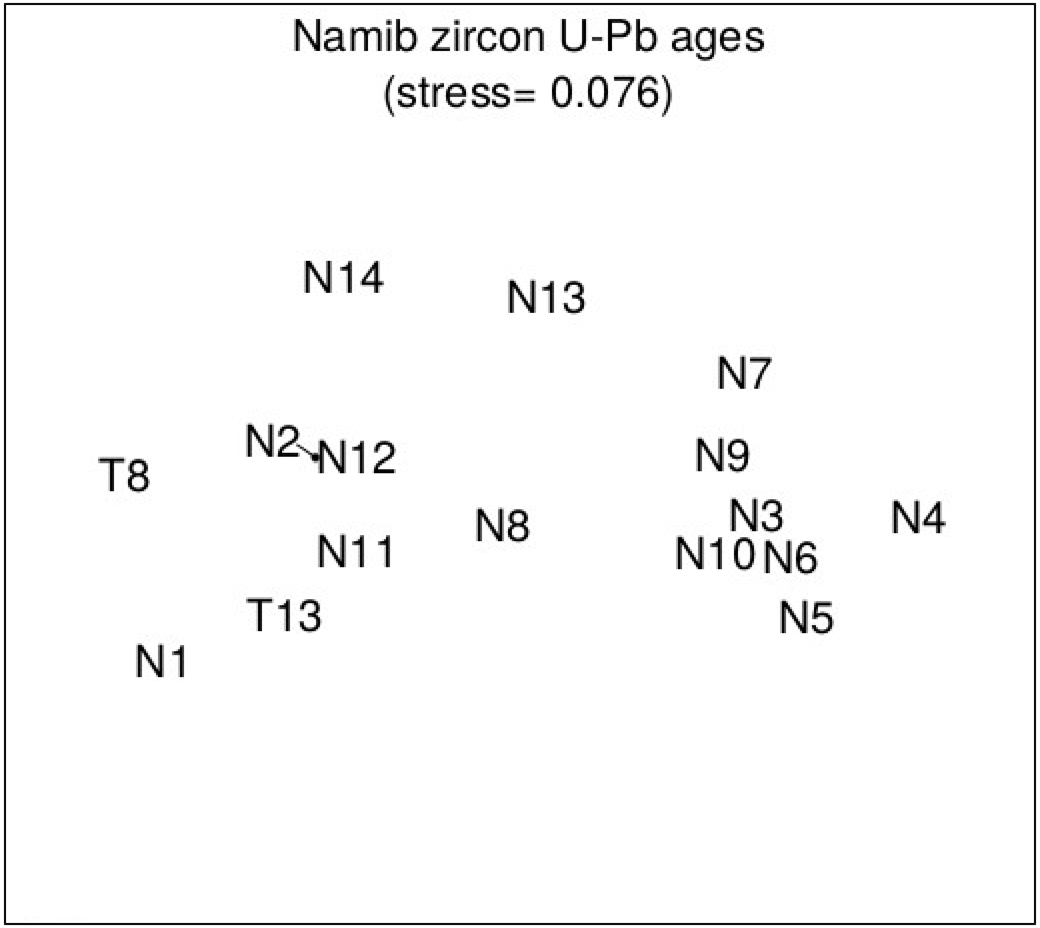



2cm

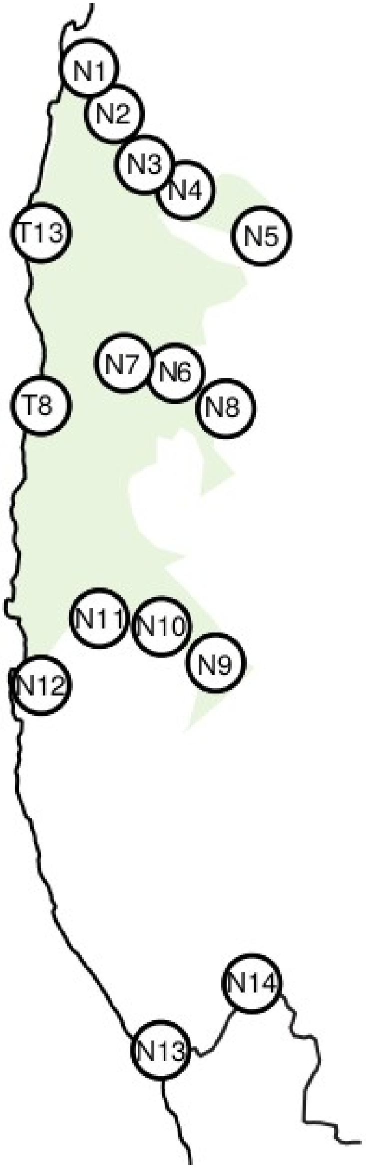


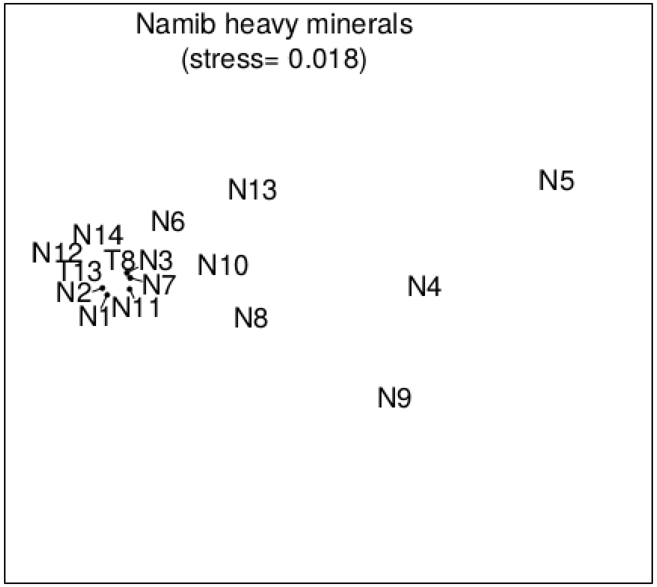
22S
14E
14E
26S
16E
30S
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
N11
N12
N13
N14
T8
T13
U-Pb
HM
QFL
Major
Trace
0
1
2
3 Ga
am
gt
ep
oth1
opx
cpx
Q
KF
Ls
Lv
Lm
P
SiO2
oth2
K2O
Na2O
CaO
MgO
Fe2O3
Al2O3
Ba
Sr
Rb
oth3
oth4
Cr
Zr
Hf
V
Nb

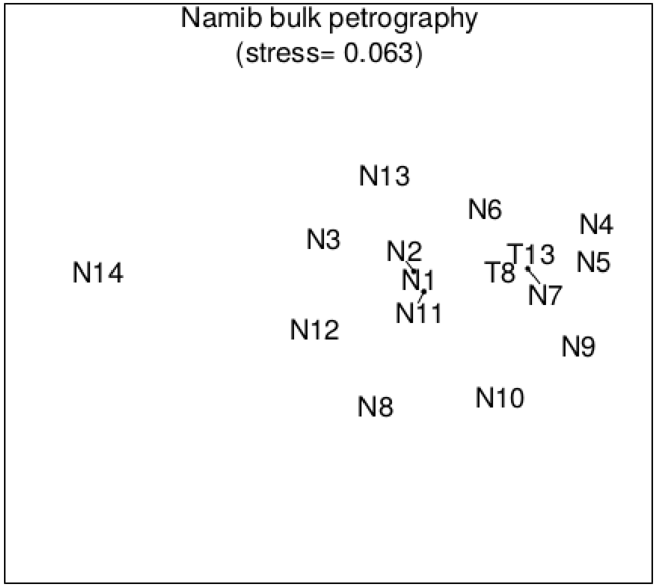

N1
N2
N3
N4
N5
T13
N7
N6
N8
T8
N11
N10
N9
N12
N14
N13

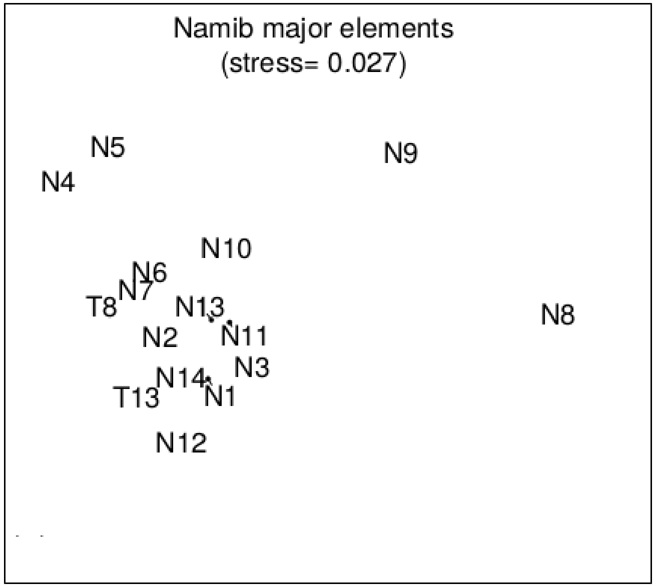

Namib zircon U-Pb ages
(stress= 0.076)
N14
N13
N7
N2
N12
T8
N9
N3
N4
N8
N11
N10
N6
T13
N5
N1

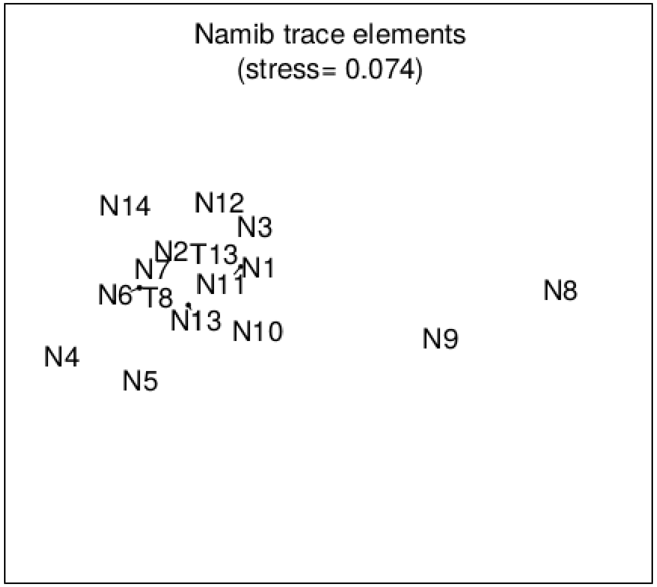
Chemical/mineralogical data yield four more MDS maps

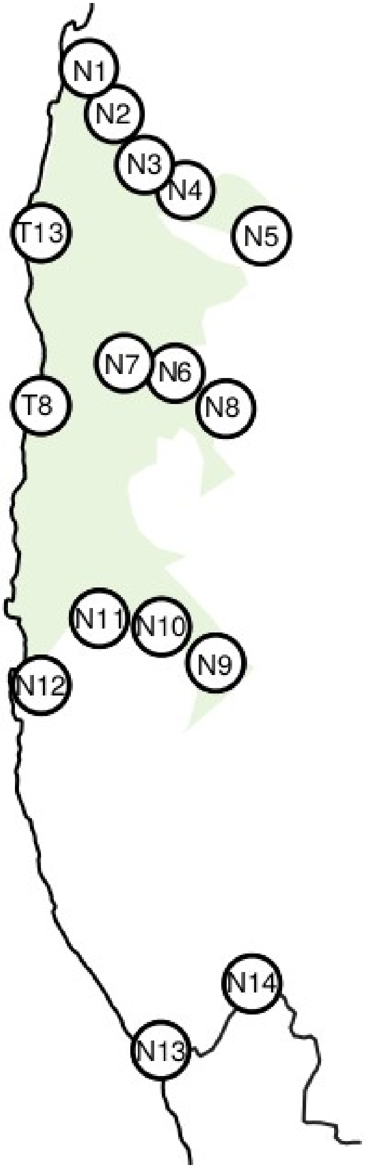


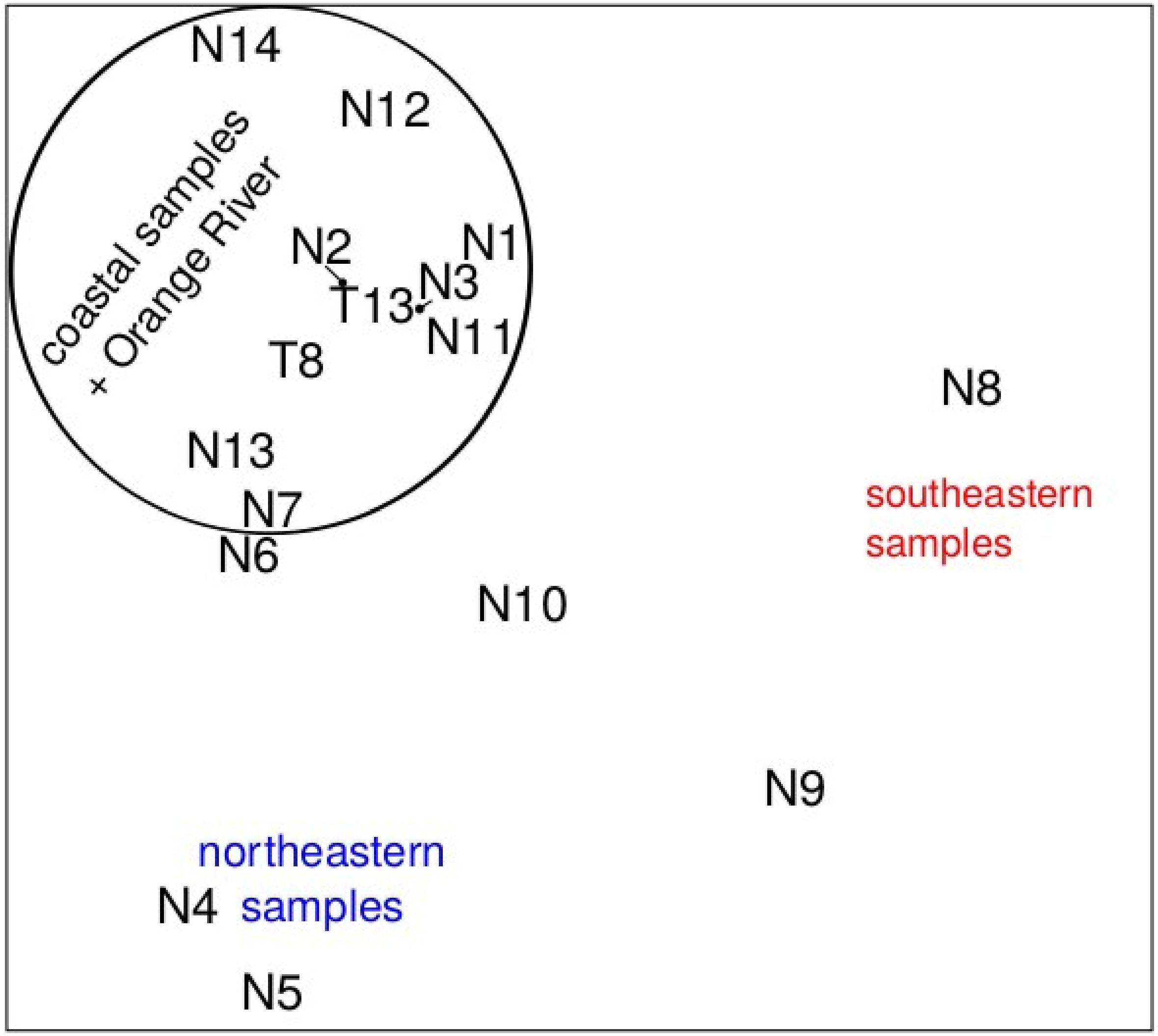




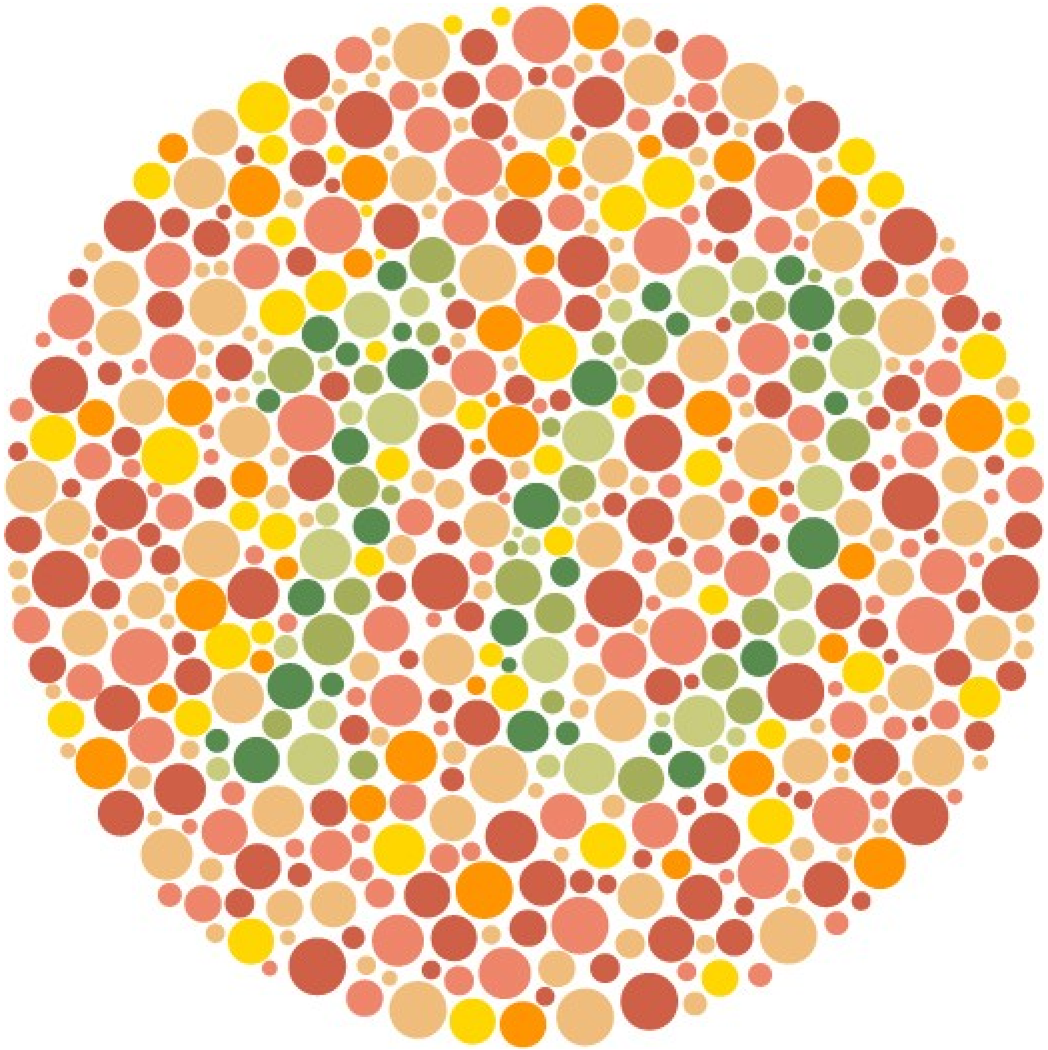





N1
N2
N3
N4
N5
T13
N7
N6
N8
T8
N11
N10
N9
N12
N14
N13

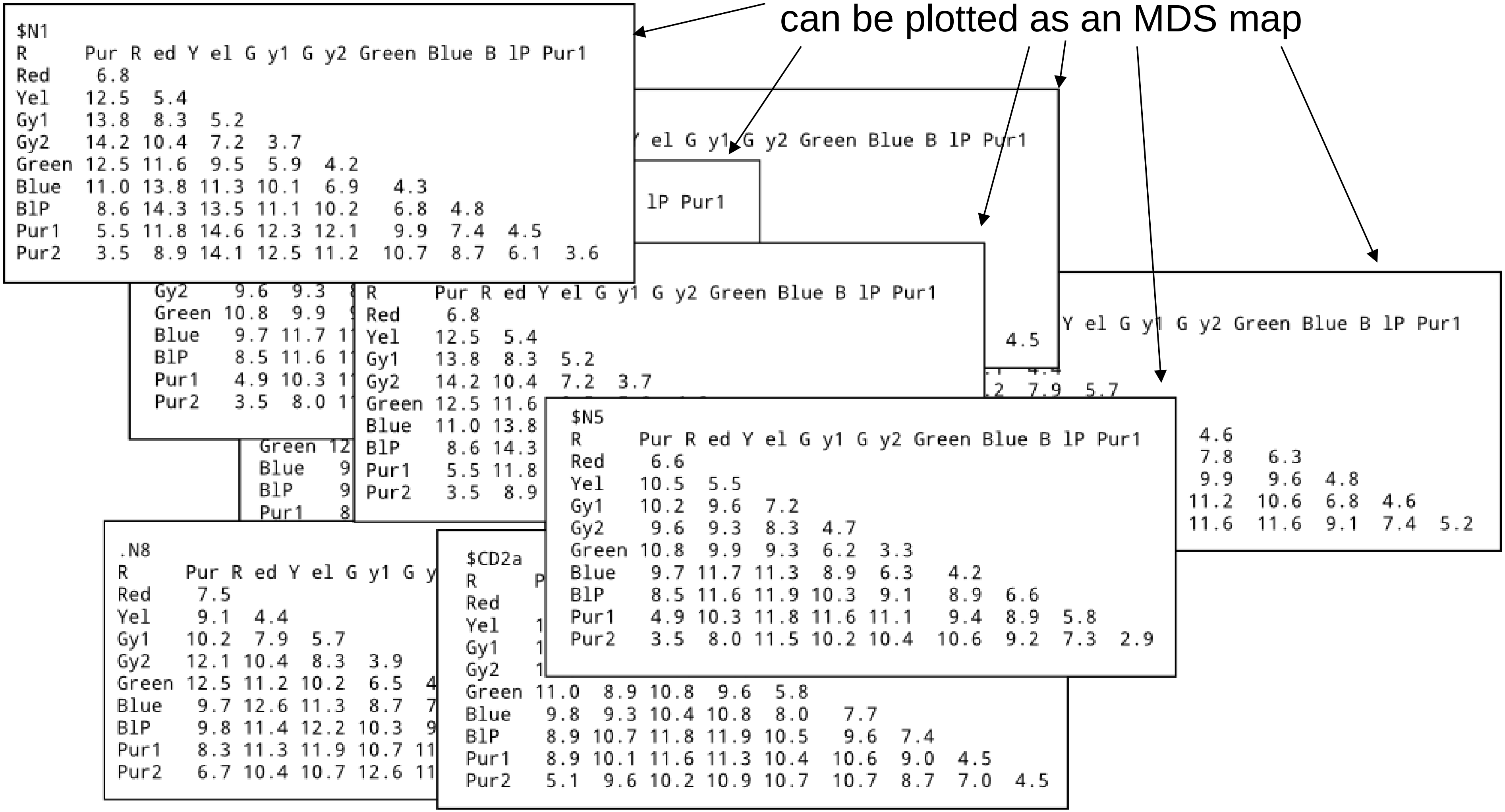

coastal samples
+ Orange River
N14
N12
N2
N1
T13
N3
N11
T8
N13
N7
N6
N10
N8
southeastern
samples
N9
northeastern
samples
N4
N5

Next example: perceived distances between colours (Helm, 1964)

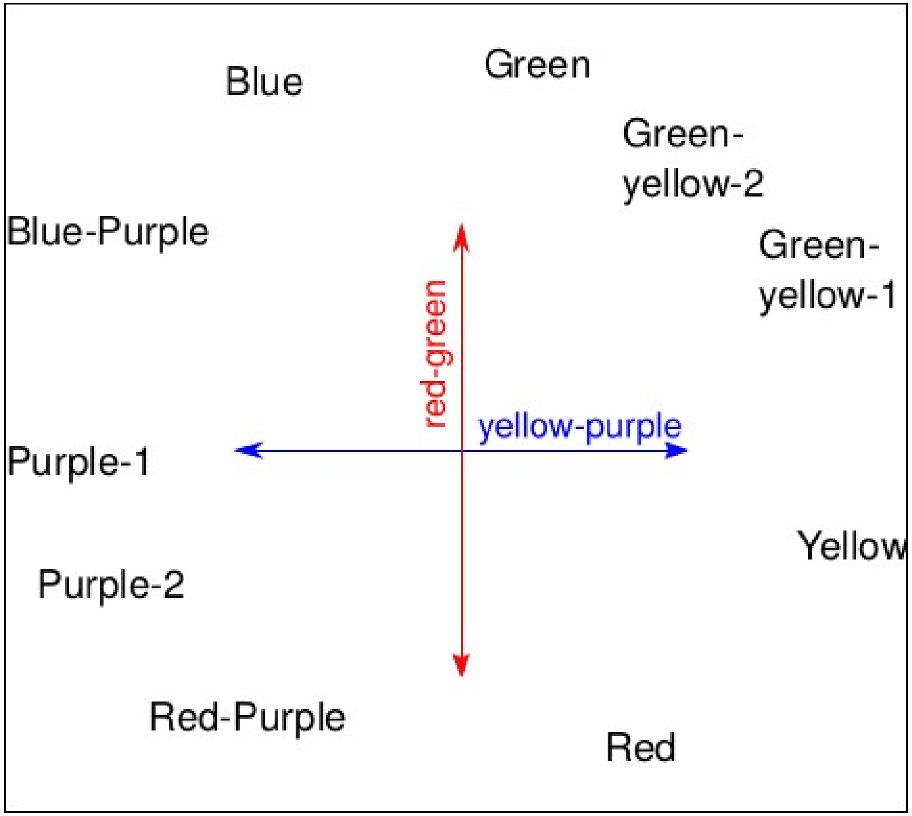


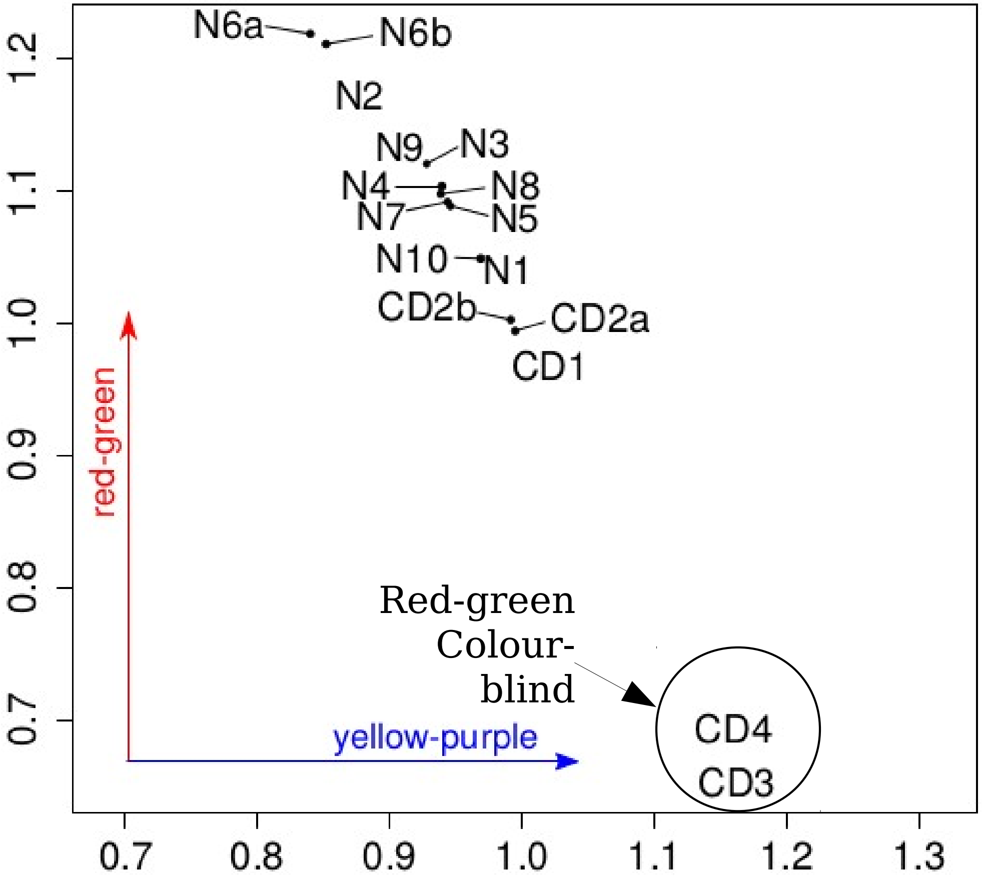
16 test subjects rank 10 colours according to their perceived similarity

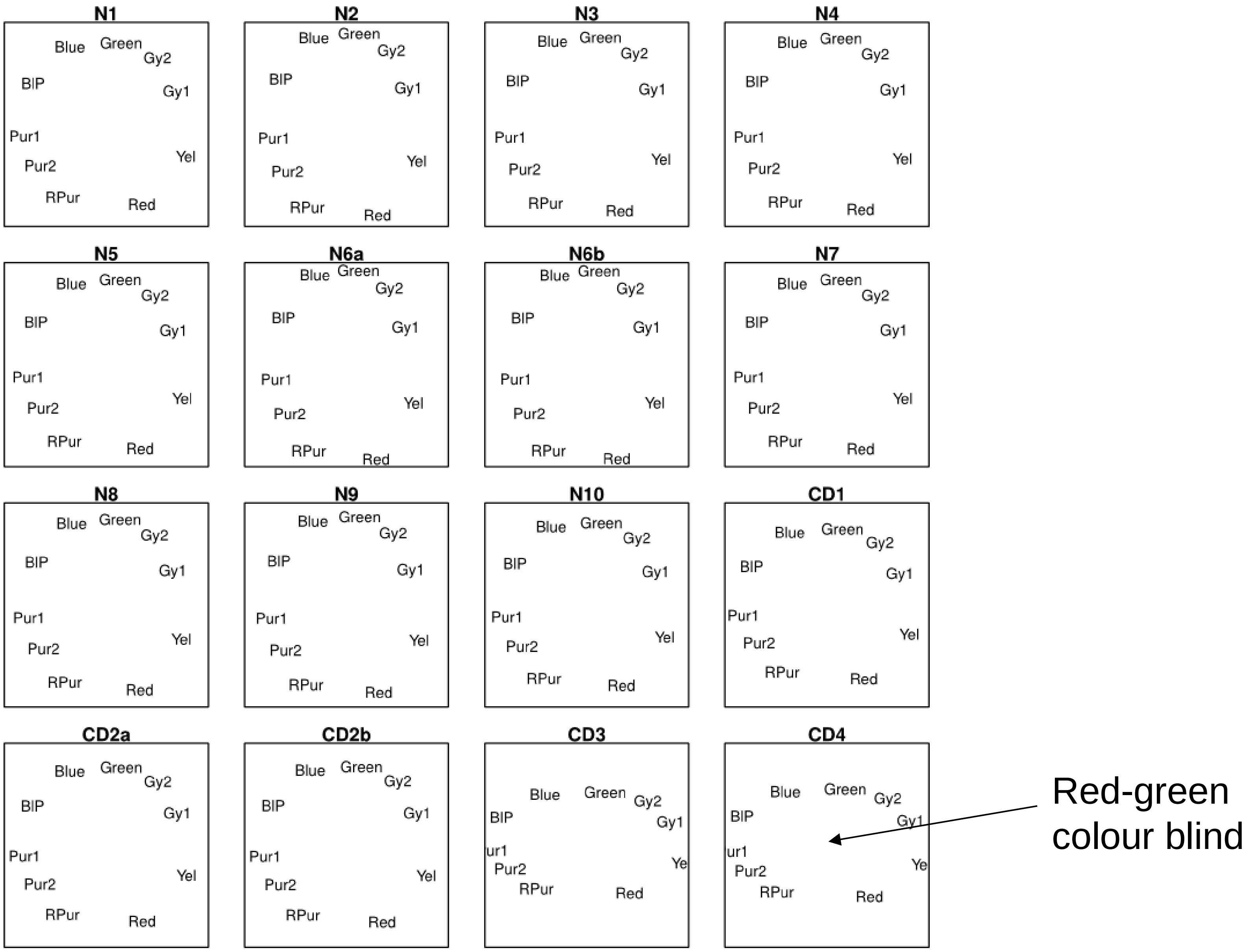
Each of these 16 9x9 distance matrices can be plotted as an MDS map



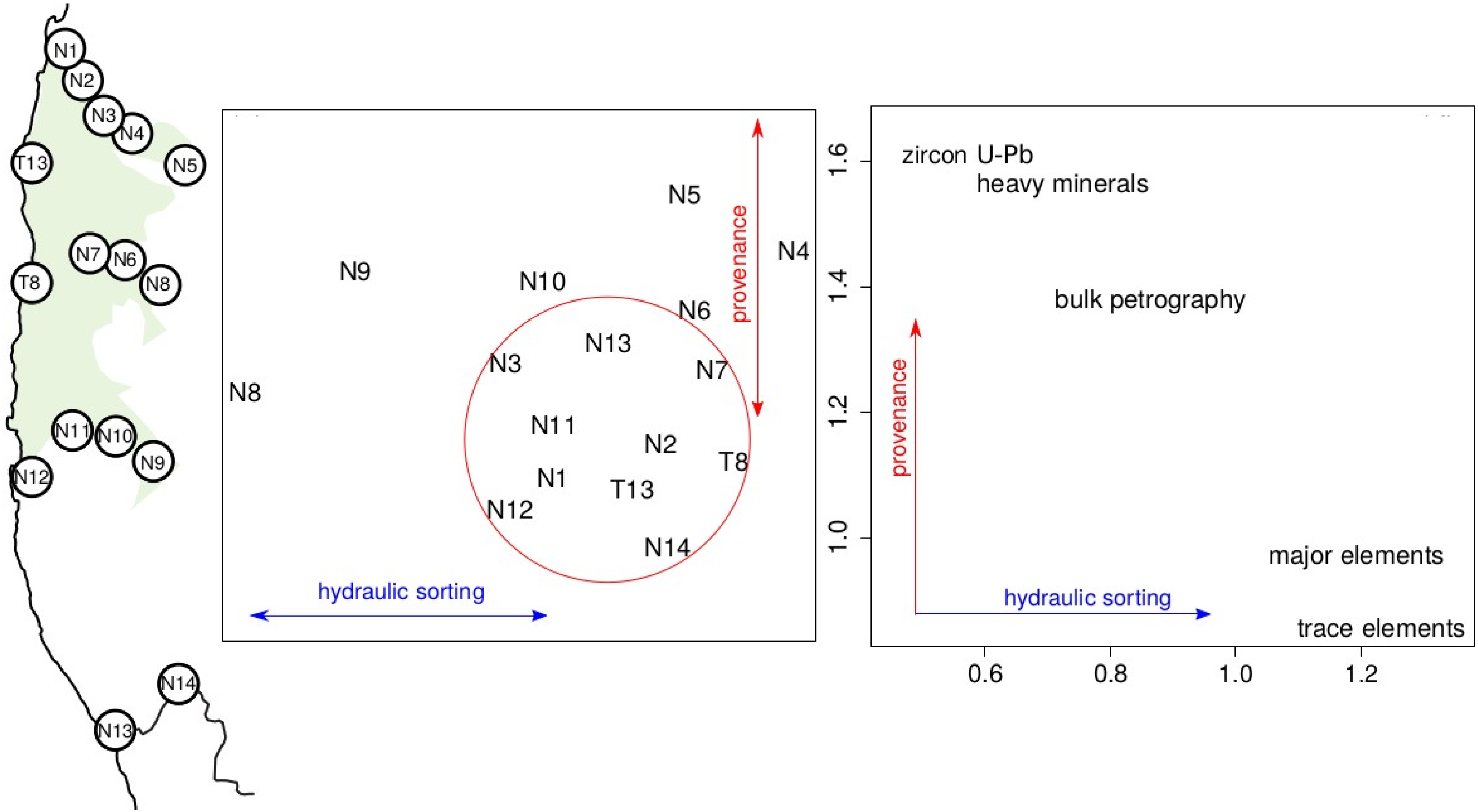
3-way MDS analysis (aka 'INDSCAL') pools all data into one analysis





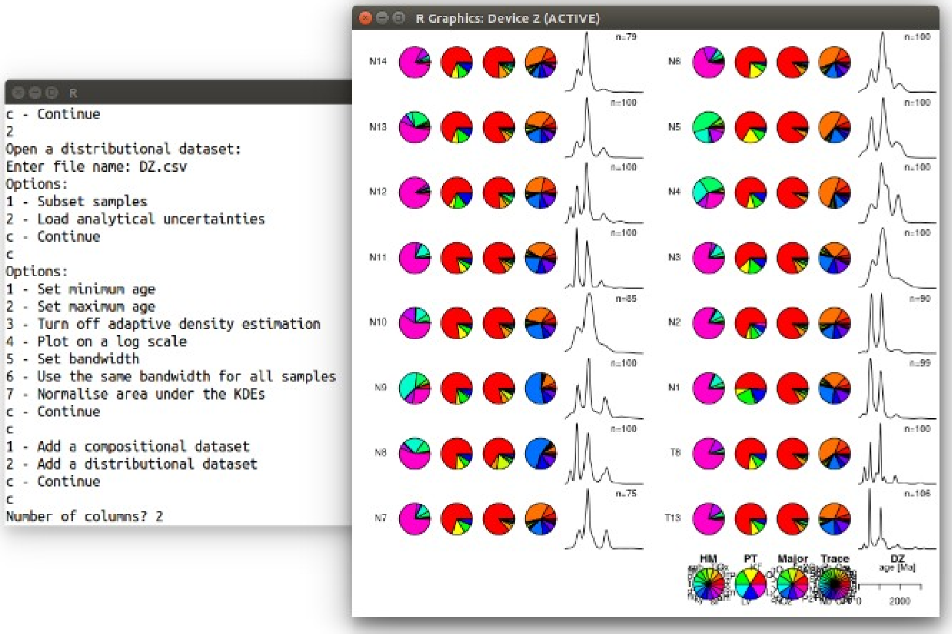

N1
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N2
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N3
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N4
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N5
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N6a
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N6b
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N7
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N8
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N9
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
N10
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
CD1
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
CD2a
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
CD2b
Blue Green Gy2 BlP Gy1 Pur1 Yel Pur2 RPur Red
CD3
Blue Green Gy2 BlP Gy1 ur1 Ye Pur2 RPur Red
CD4
Blue Green Gy2 BlP Gy1 ur1 Ye Pur2 RPur Red
Red-green colour blind

INDSCAL analysis of the Namib dataset



provenance: an R package for statistical provenance analysis

Introduction | Beginners | Advanced



provenance is a software package in the R programming environment, which aims to facilitate the visualisation and interpretation of `Big Data' in the context of sedimentary provenance analysis. After installing R from http://r-project.org, provenance can be installed by typing

```
install.packages('provenance')
```

Once installed, the package can be loaded by typing

```
library(provenance)
```

On the visualisation of detrital age distributions

Pieter Vermeesch *

Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom

Chemical Geology 312–313 (2012) 190–194

Multi-sample comparison of detrital age distributions

Pieter Vermeesch *

Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK

Chemical Geology 341 (2013) 140–146

Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE

10.1002/2017GC007109

High Throughput Petrochronology and Sedimentary Provenance Analysis by Automated Phase Mapping and LAICPMS

Key Points:

Making geological sense of 'Big Data' in sedimentary provenance analysis

Pieter Vermeesch [a,*], Eduardo Garzanti [b]

[a] *London Geochronology Centre, University College London, United Kingdom*
[b] *Laboratory for Provenance Studies, Università di Milano-Bicocca, Italy*

Chemical Geology 409 (2015) 20–27

An R package for statistical provenance analysis

Pieter Vermeesch [a,*], Alberto Resentini [b], Eduardo Garzanti [b]

chronology Centre, University College London, United Kingdom
di Petrografia del Sedimentario, Università Milano-Bicocca, Italy

Sedimentary Geology 336 (2016) 14–25

The End