Michael Dietze[1]
Elisabeth Dietze[2]

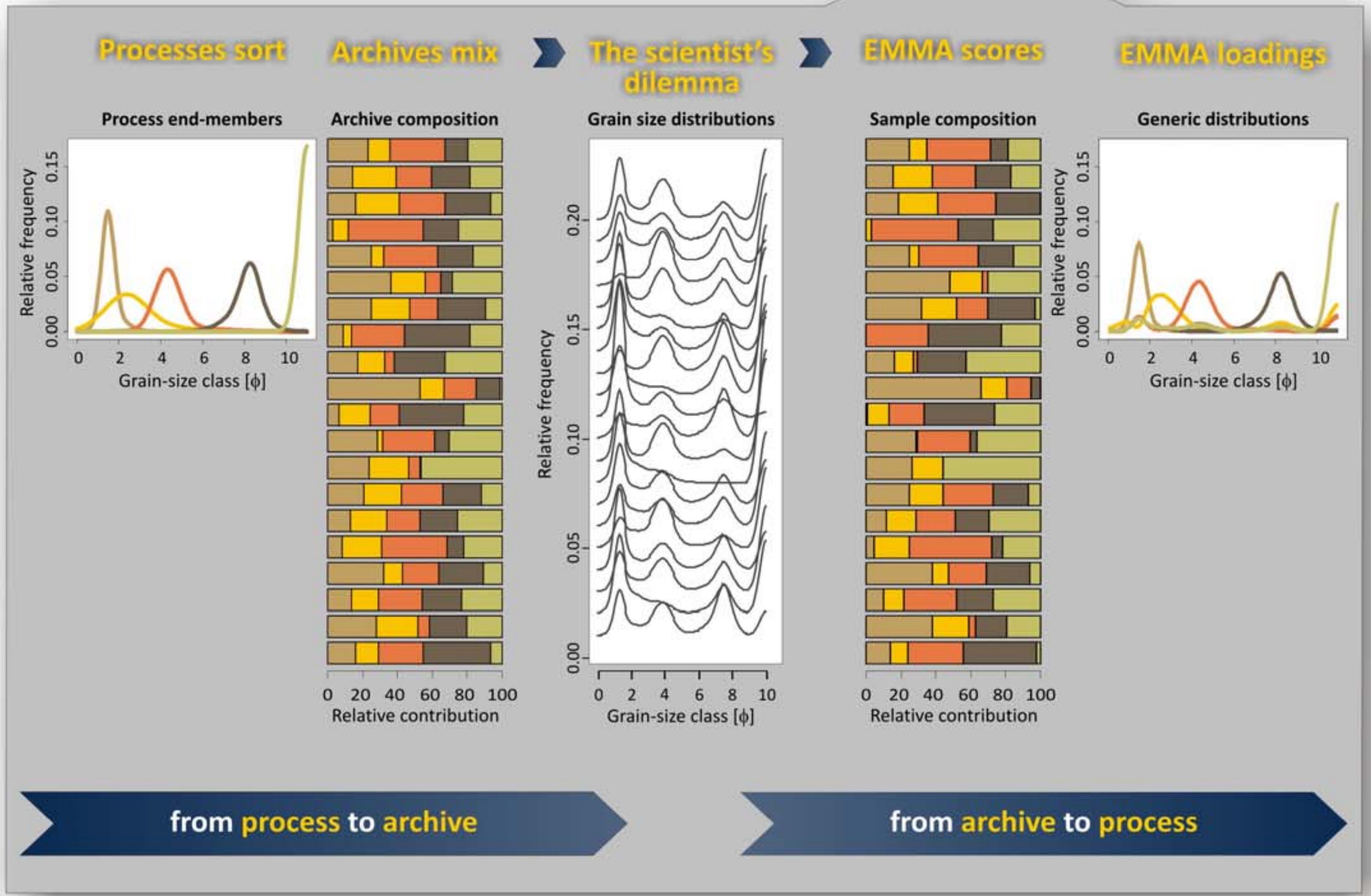
The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

Overview

Grain size analysis → Method of moments → Principle component analysis → Factor analysis → End-member modelling → Finite-mixture modelling



Why EMMA

- Meaningful interpretation of polymodal data
- Infer processes and their relative importance

Constraints

- Stationarity in processes
- Processes create non-overlapping spectrae
- Sufficient samples with respect to resolution

Strengths

- Process quantification
- Robustness and uncertainty estimation
- Open source, flexible code

Dietze et al. (2012)

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

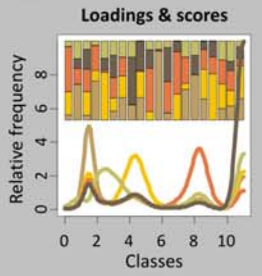
A flexible open-source toolbox for robust end-member modelling analysis

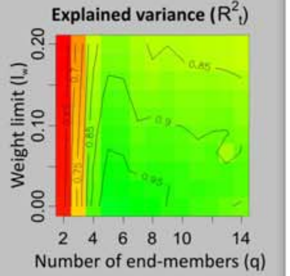
[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

Package & applications

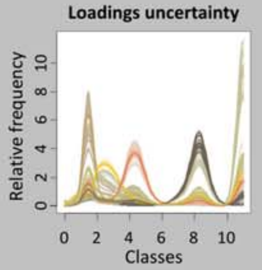
The R-package EMMAgeo

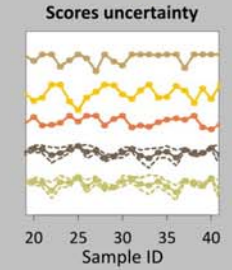
Optimisations and tests





Robustness and uncertainty





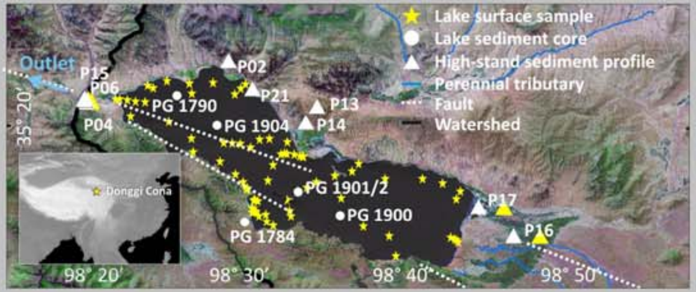
Work with the package

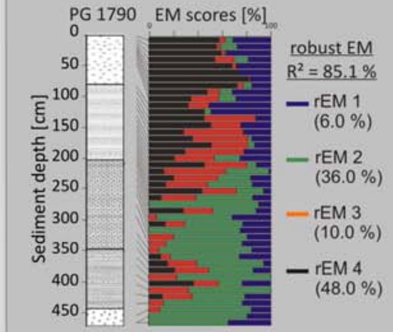
```
install.packages("EMMAgeo")
library(EMMAgeo)
```

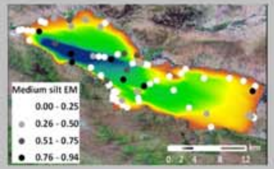
Features

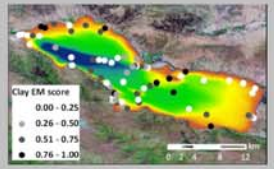
- check data consistency
- test weight limits
- test factor explanatory power
- work with user-defined loadings
- Infer model robustness
- Several measures of goodness

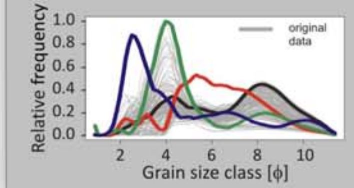
An example from the Tibetan Plateau

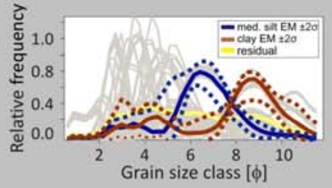












Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

The way of EMMA

Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

```
check.data(X = DC.data, q = q, lw = lw, c = 100)
[1] "Data matrix passed test... OK"
[2] "End-member vector passed test... OK"
[3] "Weight transformation limit vector
    passed test... OK"
[4] "Scaling parameter passed test... OK"
[5] "Maximum weight transformation limit value
    passed test... OK"
[6] "    Note: the following rows do not sum up
     to the specified c: 1, 42, 250."
[7] "No columns contain only zero values... OK"
```

Check the input data

- check appropriate data structure
- check for columns with only zero-values
- check for rows with NA-values

Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- residual.EM()
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

The way of EMMA

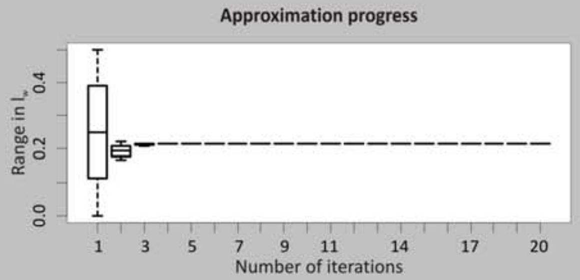
Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

```
lw.test <- seq(0, 0.5, length.out = 100)
test.lw(X = DC.data, lw = lw.test)

$step
[1] 43
$lw.max
[1] 0.2121
```

```
lw.0 <- 0 # lowest lw
lw.n <- 0.5 # highest lw
n = 10 # length of the lw-vector

for(i in 1:20) {
  lw.test <- test.lw(X = X,
                     lw = seq(from = lw.0,
                              to = lw.n,
                              length.out = n))

  lw.new = seq(lw.0, lw.n, length.out = n)[c(
    lw.test$step, lw.test$step + 1)]
  lw.0 <- lw.new[1]
  lw.n <- lw.new[2]
}
```

Test weight tranformation limits

- Weight-transformation of input data handles sample outliers, cf. Miesch (1970)
- EMMA crashes when limits set too high
- The function evaluates the last valid weight transformation limit l_w
- Use a loop to iteratively approximate l_w max



Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- residual.EM()
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

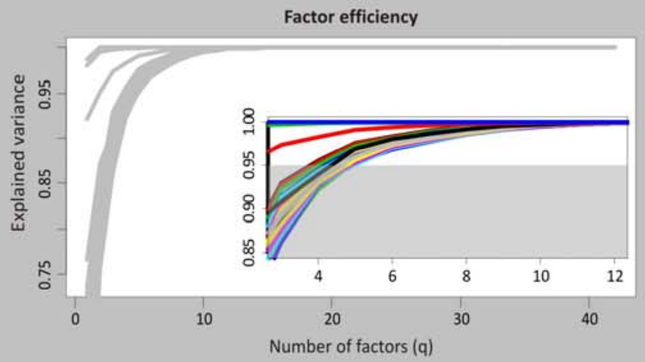
The way of EMMA

Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

```
lw = seq(0, lw.max, length.out = 20)
Lv.test <- test.Lv(X = DC.data, lw = lwplot = TRUE)
q.min <- Lv.test$q.min
q.min

[1] 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 2 1 1
```

Test explained variance & get q_{min}

- Finding the minimum number of factors/ potential end-members relies on the Kaiser criterion (i.e. $R^2 > 0.95$)
- The function returns the number of factors necessary to pass an $R^2 > 0.95$



Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- residual.EM()
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

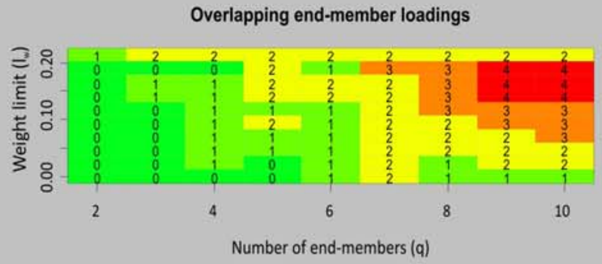
The way of EMMA

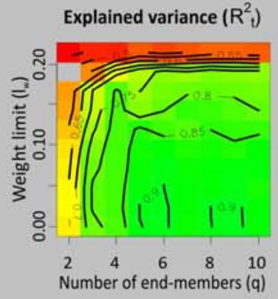
Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

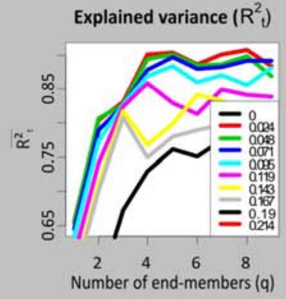
```r
q.seq <- seq(from = 2,
             to = 10)

lw.seq <- seq(from = 0,
              to = lw.max,
              length.out = 10)

params.test <- test.parameters(X = DC.data,
                               q = q.seq,
                               lw = lw.seq,
                               plot = "mRt",
                               legend = TRUE,
                               cex = 0.7)
```

Infer parameter influence & get q_{max}

- Test influence of combinations of q and l_w on diverse measures of model goodness (e.g. mRt, mRn, mRn, mEm, mEn).
- Infer number of overlapping end-member loadings
- Infer maximum number of end-members







Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- residual.EM()
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

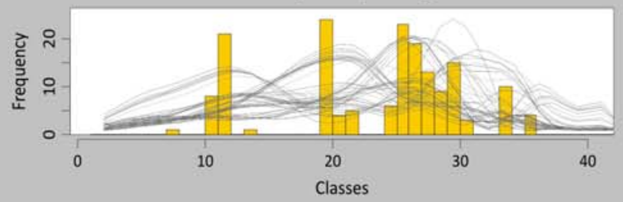
The way of EMMA

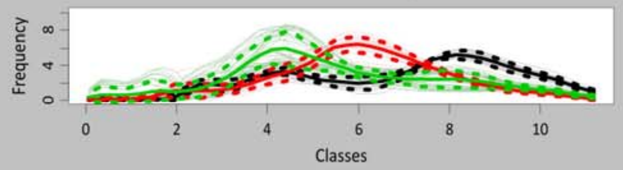
Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

```
q.min <- ifelse(q.min < 2, 2, q.min)
q.max <- ifelse(q.max < q.min, q.min, q.max)
input.matrix <- cbind(q.min, q.max, lw.seq)
input.matrix <- input.matrix[complete.cases(
                  input.matrix),]

TR <- test.robustness(X = DC.data,
                      P = input.matrix,
                      plot = TRUE,
                      ol_rej = 1,
                      mRt_rej = 0.95)

l.min <- c(10, 17, 25)
l.max <- c(15, 22, 28)
limits = cbind(l.min, l.max)

REM <- robust.EM(Vqsn = TR$Vqsn,
                 limits = limits,
                 Vqn = TR$Vqn,
                 classunits = DC.units,
                 plot = TRUE)

Vqn.mean <- REM$Vqn.mean
Vqn.sd <- REM$Vqn.sd
```

Test and extract robust end-members

- Run all parameter combinations and extract only robust ones (given by mode limits)
- Further threshold parameters: ol and $\overline{R}^2$





Package functions

mu.phi()

phi.mu()

interpolate.classes()

check.data()

test.lw()

test.Lv()

test.parameters()

EMMA()

test.robustness()

robust.EM()

residual.EM()

Mqs.uncertainty()

create.EM()

mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

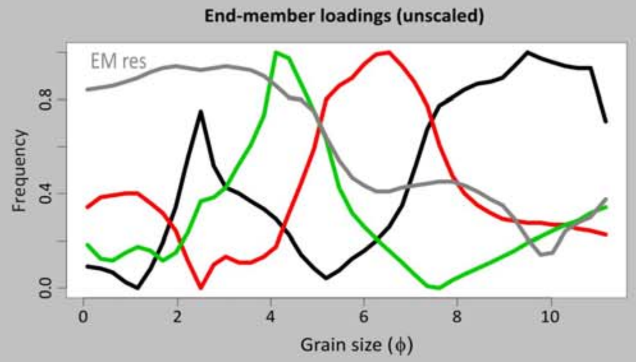
The way of EMMA

Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › **Get residual EM** › Model data › Get uncer-tainty

```
Vqn.res <- residual.EM(Vqn.mean)

plot(NA,
     main = "End-member loadings",
     xlab = expression(paste("Grain size (",
              phi, ")", sep = "")),
     ylab = "Loadings [Vol.-%]",
     xlim = range(DC.units),
     ylim = c(0, 1))

for(i in 1:nrow(Vqn.mean)) {
  lines(DC.units,
        Vqn[i,],
        col = i)
}
lines(DC.units, Vqn.res, col = "grey")
```

Retrieve the residual end-member

- Robust, unscaled end-members (Vqn) are not able to cover the entire data variance
- The residual end-member aims to describe this portion of “open” variance



Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- **residual.EM()**
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

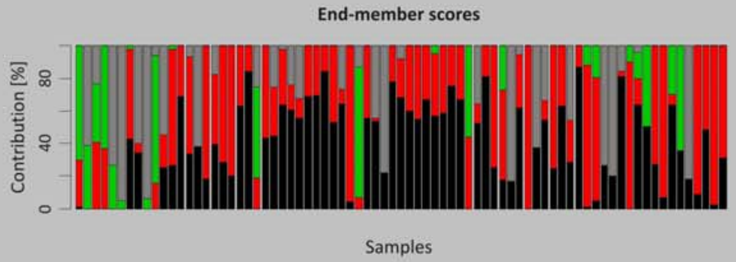
The way of EMMA

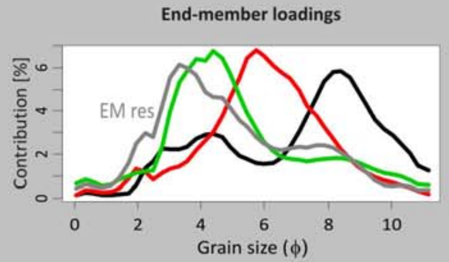
Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

```
Vqn.rob.res <- rbind(Vqn.mean, Vqn.res)

EM.rob <- EMMA(X = DC.data,
               q = nrow(Vqn.rob.res),
               lw = 0.012,
               Vqn = Vqn.rob.res,
               classunits = DC.units,
               plot = TRUE)
```

Model the data with robust loadings

- EMMA() can be run with intrinsic or user-defined end-members (Vqn).
- This allows to include both, robust and residual end-member loadings.





Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- residual.EM()
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

INTRODUCTION
DETAILS
APPLICATIONS
OUTLOOK ETC.

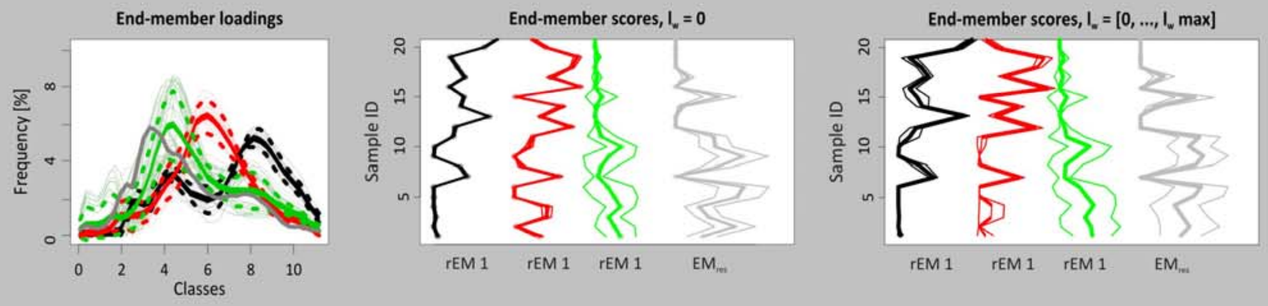
The way of EMMA

Check data › Test weight limit › Test factors › Test para-meters › Get robust EMs › Get residual EM › Model data › Get uncer-tainty

```
Vqn.rob.res.sd <- rbind(Vqn.sd,
                        rep(0, nrow(Vqn.sd)))
lw.test <- c(0, lw.max)

M <- Mqs.uncertainty(X = DC.data,
                     q = 4, lw = lw.test,
                     runs = 200,
                     Vqn = Vqn.rob.res.mean,
                     Vqn.sd = Vqn.rob.res.sd,
                     type.lw = "runif",
                     autocorrelation = 3)
```

Estimate uncertainty of Mqs and Vqsn

- Monte Carlo runs allow to estimate end-member scores uncertainty, varying number of end-members and weight transformation limits.
- The robustness test contributes uncertainty estimation for end-member loadings.
- data autocorrelation can be implemented



Package functions

- mu.phi()
- phi.mu()
- interpolate.classes()
- check.data()
- test.lw()
- test.Lv()
- test.parameters()
- EMMA()
- test.robustness()
- robust.EM()
- residual.EM()
- Mqs.uncertainty()
- create.EM()
- mix.EM()

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

INTRODUCTION
DETAILS
APPLICATIONS
OUTLOOK ETC.

The EMMA algorithm

Rescale data matrix X to constant sum c.

```r
X  <- X / apply(X, 1, sum) * c
```

Weight tranformation after Miesch (1970).

```r
qts <- function(X, lw) quantile(X, c(lw, 1-lw), type = 5)
ls <- t(apply(X, 2, qts, lw = lw))
W <- t((t(X) - ls[,1]) / (ls[,2] - ls[,1]))
```

Similarity matrix calculation (outer product).

```r
A <- t(W) %*% W
```

Eigen space extraction.

```r
EIG <- eigen(A)

V <- EIG$vectors[,order(seq(ncol(A), 1, -1))]
Vf <- V[,order(seq(ncol(A), 1, -1))]
L <- EIG$values[order(seq(ncol(A), 1, -1))]
Lv <- cumsum(sort(L / sum(L), decreasing = TRUE))
```

Varimax rotation of the eigen vector matrix Vf.

```r
Vr <- do.call(rotation, list(Vf[,1:q]))
```

Extract and sort (decreasing) factor loadings and write them to matrix Vq.
Rescale (Vqr) and normalise (Vqn) the factor loadings column-wise.

```r
Vq <- Vr$loadings[,order(seq(q, 1, -1))]
Vqr <- t(t(Vq) / apply(Vq, 2, sum)) * c
Vqr <- t(Vqr)
Vqn <- t((Vqr - apply(Vqr, 1, min)) / (
  apply(Vqr, 1, max) - apply(Vqr, 1, min)))
```

Calculate factor scores matrix (Mq) by non-negative least square fitting of Vqn and transposed row-wise weight-transformed data W.

```r
Mq <- matrix(nrow = nrow(X), ncol = q)
for (i in 1:nrow(X)) {Mq[i,] = nnls(Vqn, as.vector(t(W[i,])))$X}
```

Model the dataset (Wm) as the inner product

```r
Wm <- Mq %*% t(Vqn)
```

Rescale the factor loadings matrix Vqn to Vqsn.

```r
s <- (c - sum(ls[,1])) / apply(Vqn * unname(ls[,2] - ls[,1]), 2, sum)
Vqs <- Vqn
for(i in 1:q) {Vqs[,i] <- t(s[i] * t(Vqn[,i]) * (ls[,2] - ls[,1]) + ls[,1])}
Vqsn <- t(t(Vqs)/ apply(Vqs, 2, sum)) * c
```

Rescale factor scores (Mq) to matrix Mqs and calculate variance explanied by scores.

```r
Mqs <- t(t(Mq) / s) / apply(t(t(Mq) / s), 1, sum)
Mqs.var <- diag(var(Mqs)) / sum(diag(var(Mqs))) * 100
```

Evaluate measures of model goodness.

```r
Em <- as.vector(apply(X - Xm, 1, mean))
En <- as.vector(apply(X - Xm, 2, mean))
Rm <- diag(cor(t(X), t(Xm))^2)
Rn <- diag(cor(X, Xm)^2)
```

after Dietze et al. (2012)

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

1 TECHNISCHE UNIVERSITÄT DRESDEN
2 GFZ Helmholtz Centre Potsdam

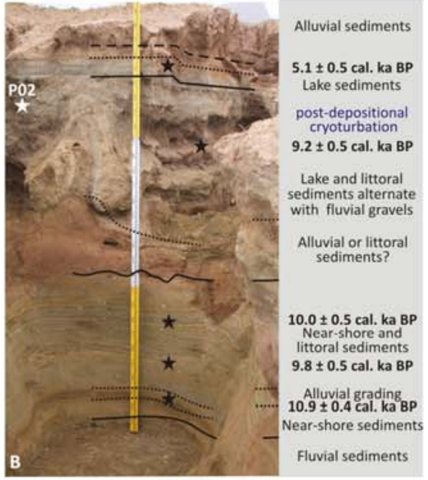
INTRODUCTION
DETAILS
APPLICATIONS
OUTLOOK ETC.

Lake high-stand deposits

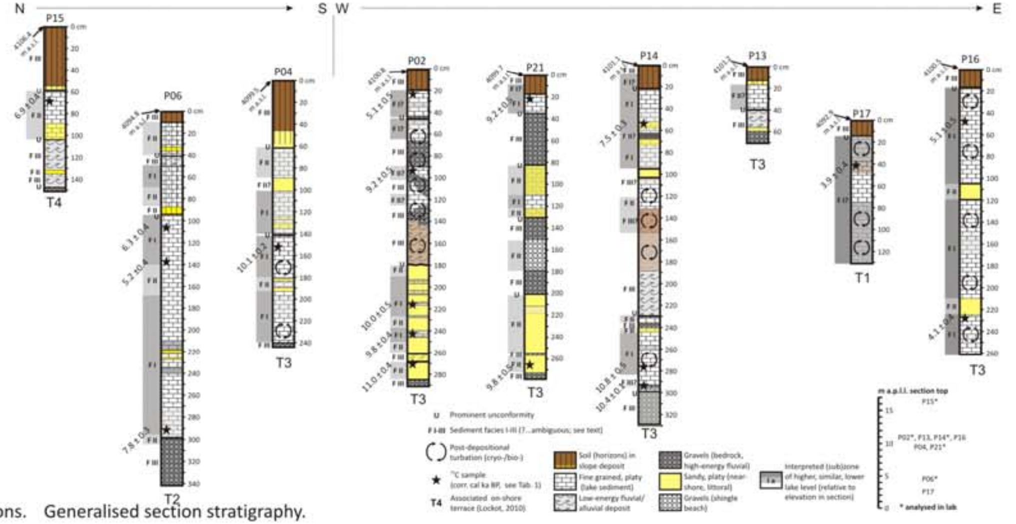
Lake high-stand sediments in onshore terraces at Lake Donggi Cona, north-eastern Tibetan Plateau allow a detailled view into the lake history.

End-member modelling of grain-size samples allowed identification and quantification of sediment delivering processes. A normalized difference between the finest and coarsest end-members was calculated as a proxy for relative lake-level change.

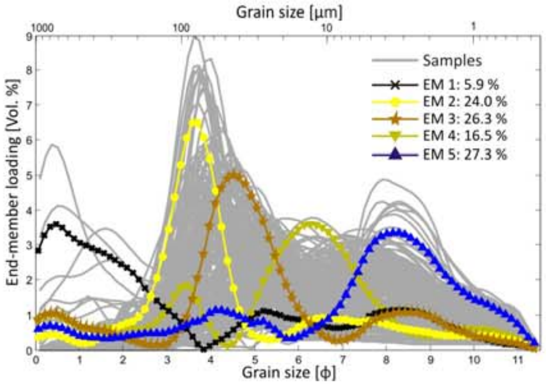
The interpreted processes are high-energy fluvial transport, low-energy unconfined flow, proximal dust, remote dust and lacustrine suspension load (EM 1 to 5, respectively).



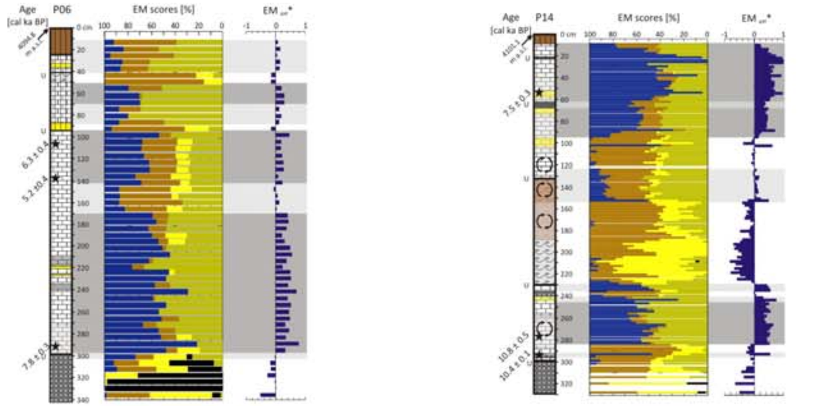
Complex stratigraphy with post-depositional alterations.



Generalised section stratigraphy.



End-member loadings of high-stand sediment grain size samples.



End-member scores and normalized difference of EM 2 and EM 5 in relation to section characteristics.

Dietze et al. (2013)

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

1 TECHNISCHE UNIVERSITÄT DRESDEN
2 GFZ Helmholtz Centre Potsdam

INTRODUCTION

DETAILS

APPLICATIONS

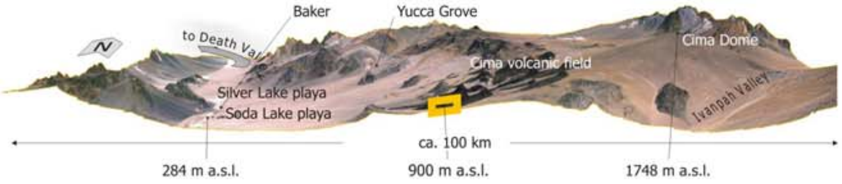
OUTLOOK ETC.

Accretionary soils

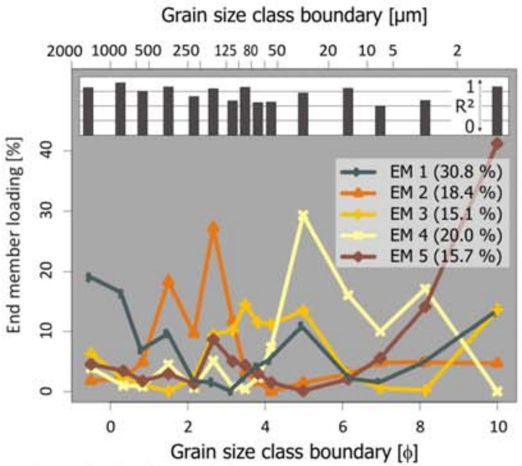
Cima volcanic field, eastern Mojave Desert, USA, hosts dissected lava flows with soil-sediment-complexes of aeolian origin, covered by stone pavements.

Relevant processes contributing to profile thickening are accumulation of dune sand, proximal dust and remote dust as well as admixture of local detritus and enrichment of pedogenic clay (EM 1 to 5, respectively).

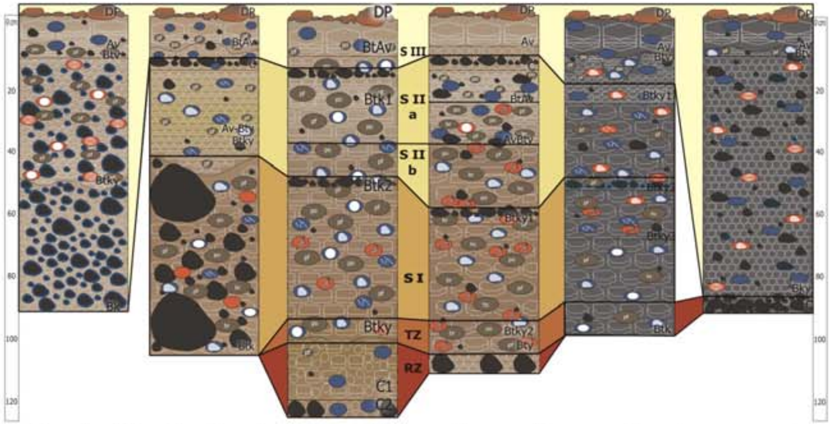
These processes are identifed by EMMAgeo, based on 104 samples with 16 classes and show a robust result even for a comparably poor sample resolution.



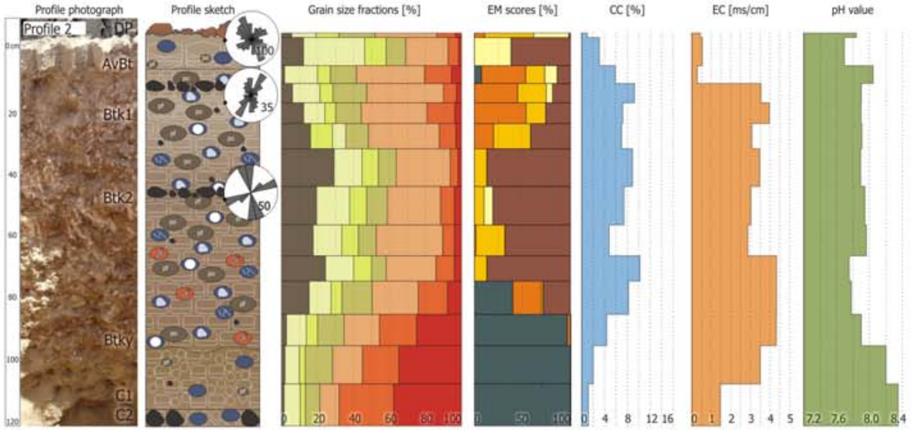
Cima volcanic field, a set of 40 cinder cones and lava flows downwind a prominent dust source.



End-member loadings and class-wise explained vairance.



Stratigraphy of six soil-sediment complexes with 3 accretionary units, covered by a stone pavement.



Photo, sketch, grain-size data, end-member scores and soil-chemical data for on accretionary section.

Dietze et al. (2011)

Michael Dietze[1]
Elisabeth Dietze[2]

The R-package EMMAgeo

A flexible open-source toolbox for robust end-member modelling analysis

[1] TECHNISCHE UNIVERSITÄT DRESDEN
[2] GFZ Helmholtz Centre Potsdam

Outlook & Ressources

How to access the R-package EMMAgeo

- http://tu-dresden.de/Members/micha.dietze/EMMAgeo

The foundations of the package

- Weltje GJ. 1997. End-member modeling of compositional data: numerical–statistical algorithms for solving the explicit mixing problem. Mathematical Geology 29: 503-549.
- Dietze E, Hartmann K, Diekmann B, IJmker J, Lehmkuhl F, Opitz S, Stauch G, Wünnemann B, Borchers A. 2012. An end-member algorithm for deciphering modern detrital processes from lake sediments of Lake Donggi Cona, NE Tibetan Plateau, China. Sedimentary Geology 243-244: 149-180.