Dimension Estimates of Multivariate Hydro-Meteorological data and their Behaviour during Extreme „Events“

Reik Donner[1], Annette Witt[1,2], Jürgen Kurths[1]

[1]Nonlinear Dynamics Group, University of Potsdam

[2]Department of Geography, King‘s College London

(data provided by: H. Rust, M. Kallache, J. Kropp (PIK))

Motivation

Environmental time series are frequently multivariate containing spatio-temporal or complementary information

Idea: Changes in the dynamics of the underlying system may lead to temporary variations of the number of „significant“ variability patterns (dimension of the record) – Reduce information about variability of the record to a single parameter

Problem: Define dynamically meaningful estimates of nonlinear measures (dimension densities) that are applicable even for very short multivariate time series

Overview

1. KLD-based Dimension Estimates
2. Numerical Example
3. Application 1: Palaeoclimatic (multi-proxy) data
4. Application 2: Hydro-meteorological (spatio-temporal) data
5. Summary
6. Open Questions

1. KLD-based Dimension Estimates

Methods for statistical decomposition of multivariate time series:

Linear methods (need only few points in time):

- Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Karhunen-Loéve Decomposition (KLD)

Nonlinear methods (usually need many points in time – not suitable for calculation of temporary measures):

- Nonlinear Principal Component Analysis (NLPCA)
- Local Linear Embedding (LLE)
- Isometric Feature Mapping (Isomap)
- Independent Component Analysis (ICA)
- Multi-Channel Singular System Analysis (M-SSA)
- …

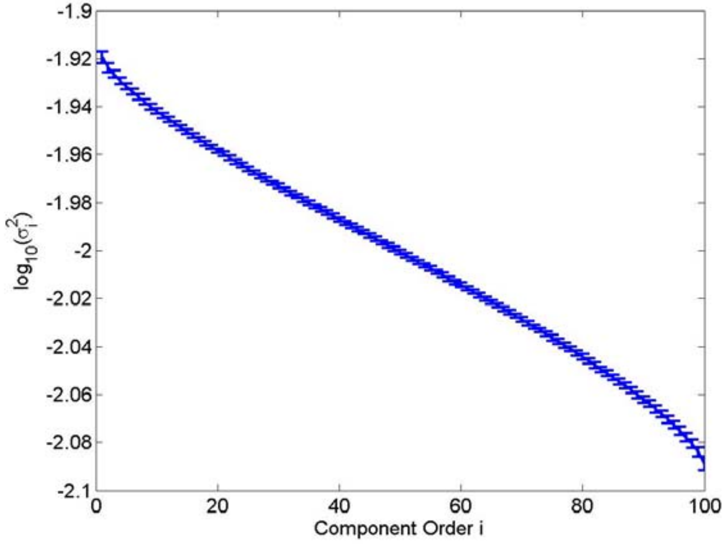
1. KLD-based Dimension Estimates

Karhunen-Loève decomposition (KLD)

Statistical decomposition of a multivariate data set A in terms of diagonalization of the covariance (scatter) matrix $S = A^T A$:

$S = U^T \Sigma U$ with $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$

(consider here only Σ)



<u>Convention:</u>

Order non-negative eigenvalues σ_i^2 of S in descending order: $\sigma_1^2 \geqslant \ldots \geqslant \sigma_N^2 \geqslant 0$ and normalize to unit sum (wherever appropriate): $\sum_{i=1}^{N} \sigma_i^2 = 1$

1. KLD-based Dimension Estimates

KLD dimension: Definition (modified from Zoldi and Greenside 1997)

KLD dimension: Minimum number of components required to capture a fraction f of the total variance of the data

$$D_{KLD}(f) = \min_{p}\left\{ \sum_{i=1}^{p} \sigma_i^2 \Big/ \sum_{i=1}^{N} \sigma_i^2 \geq f \right\}$$

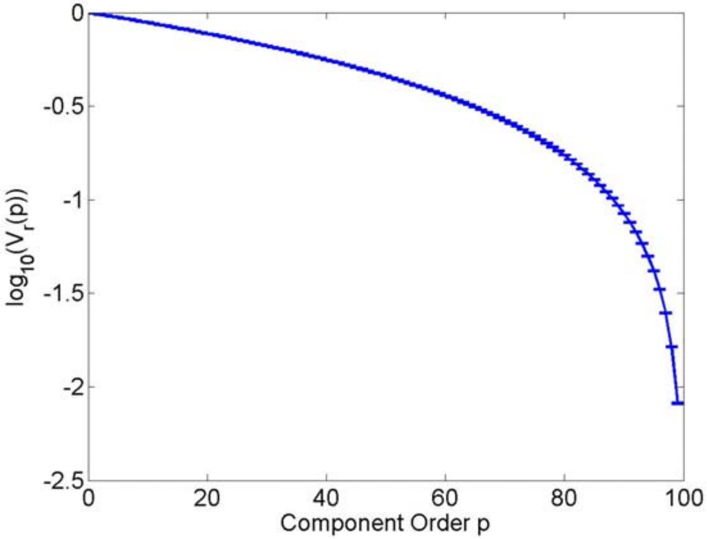
Normalized measure: KLD dimension density: $\delta_{KLD}=D_{KLD}/N$

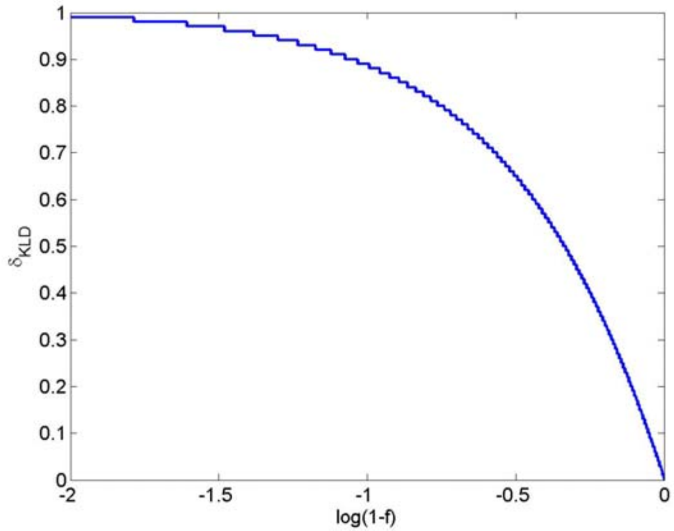
Applications: Model systems of spatio-temporal chaos, reaction-diffusion systems, electrochemical turbulence (experiment)

1. KLD-based Dimension Estimates

KLD dimension: Meaning

KLD dimension density is directly related to the remaining variances $V_r(p) = 1 - \sum_{i=1}^{p} \sigma_i^2$ rather than the component variances themselves: $V_r(p/N)$ corresponds to the inverse of the function $\delta_{KLD}(f)$





1. KLD-based Dimension Estimates

KLD dimension: Problems

Problem 1: Dynamical consideration of the measure

$\Rightarrow$ Computation of δ_{KLD} for sliding windows in time (demonstrate qualitative robustness for small amounts of data)

Problem 2: Crucial dependence on variance fraction f

$\Rightarrow$ Consider $\delta_{\mathrm{KLD}}(f)$ only as an estimate of a relative dimension (demonstrate qualitative robustness against varying f)

Problem 3: Finite number of possible values

$\Rightarrow$ Scaling of remaining variances as a new relative dimension estimate: linear variance decay (LVD) dimension density

1. KLD-based Dimension Estimates

LVD dimension: Definition

Idea: Decay of remaining variances typically dominated by exponential decay => fit a corresponding model function

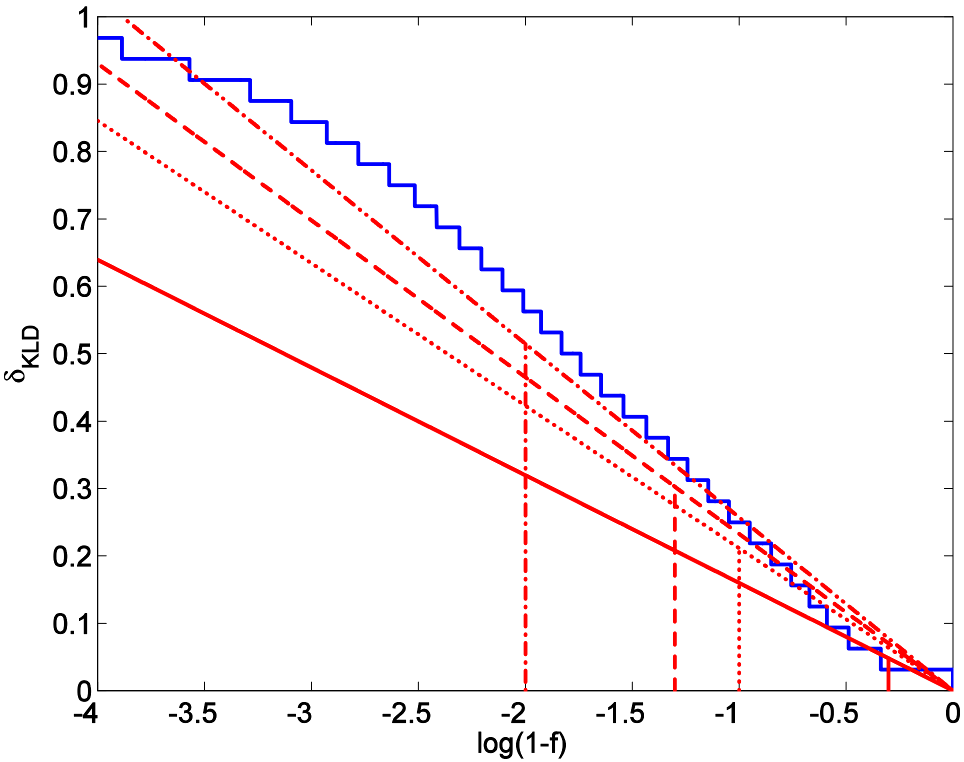
- Least-square approach to $V_r(p/N)$ (problematic for small N)
- Least-square approach to $\delta_{KLD}(f)$ (continuously defined for 0<f<1)

$$F_\alpha(f) = \int_0^f \left(\delta_{KLD}(\phi) + \alpha \log(1-\phi)\right)^2 d\phi = \int_{\log(1-f)}^0 \left(\delta_{KLD}(x) + \alpha x\right)^2 e^x dx$$

$\alpha_{\min} = \delta_{LVD}$ linear variance decay (LVD) dimension density

1. KLD-based Dimension Estimates

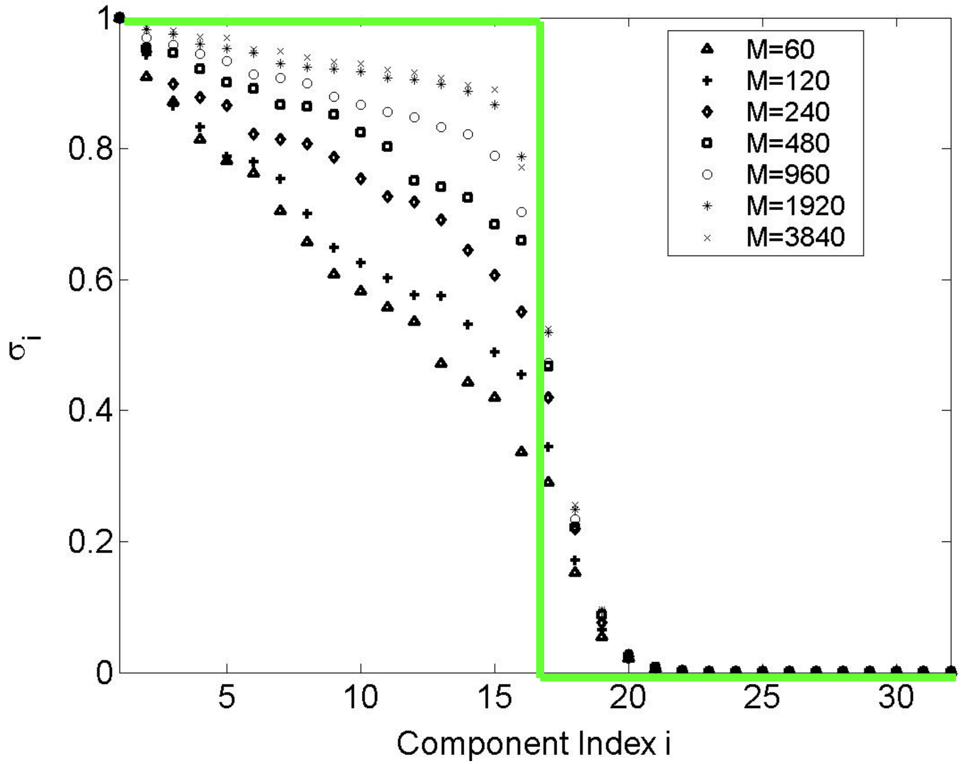
LVD dimension: Example



solid: f=0.5
dotted: f=0.9
dashed: f=0.95
dash-dotted: f=0.99

2. Numerical Example

Politi and Witt 1999: Model with prescribed dimension density d

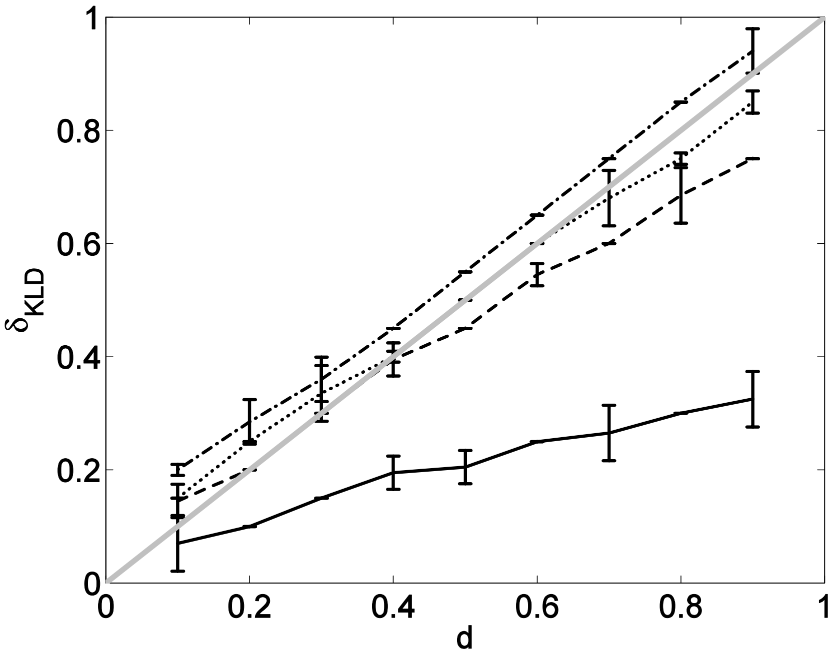


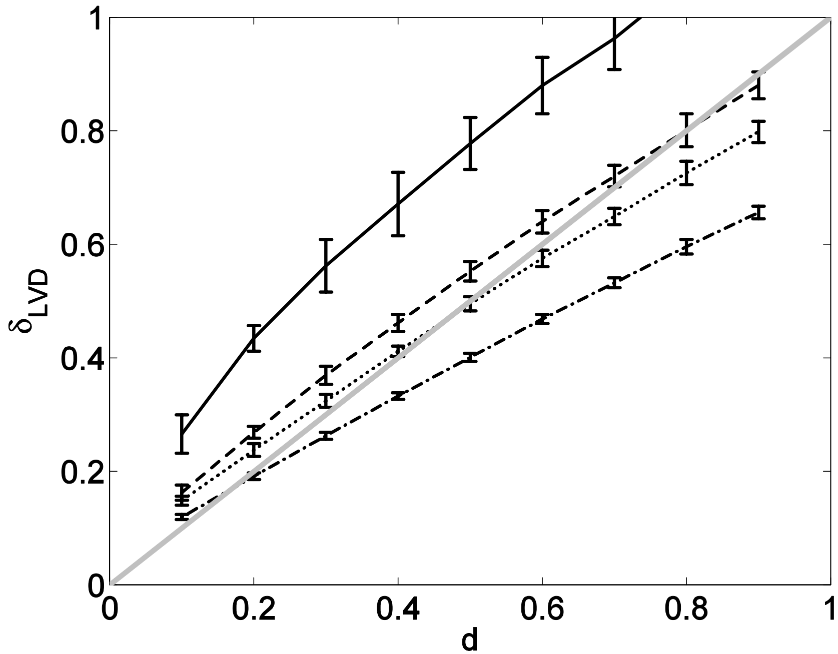
Green: asymptotic behaviour long data series $M \to \infty$

2. Numerical Example

Example: Dependence of KLD-based dimension estimates on d

(M=20, N=100, 50 realizations)





Solid: f=0.5, dotted: f=0.9, dashed: f=0.95, dash-dotted: f=0.99

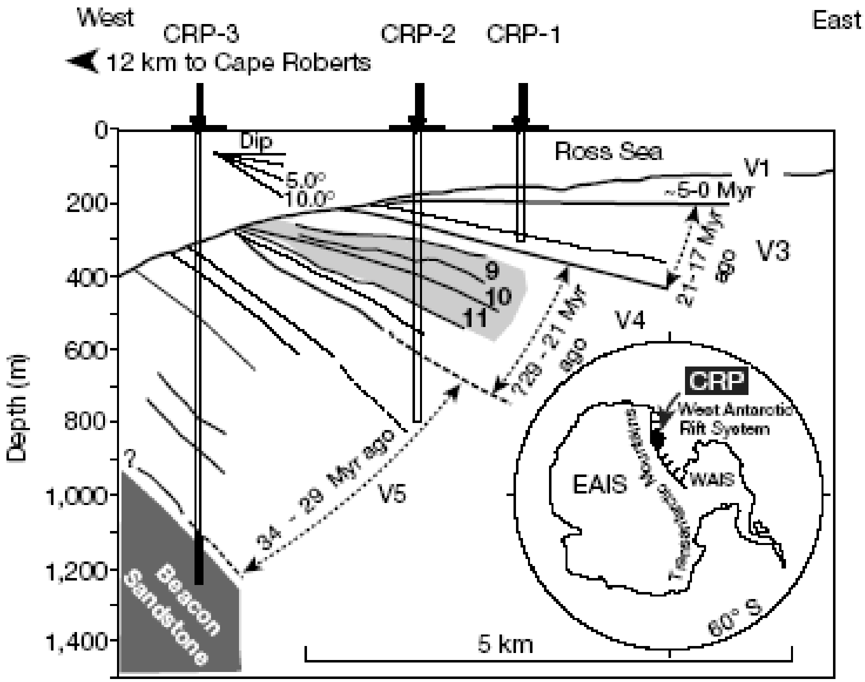
2. Numerical Example

Result for small data sets (M=20, N=100)

- Both dimension estimates are sensitive to changes in true dimension
- KLD dimension restricted to finite number of values (coarse-grained estimate with step-wise dependence on f)
- For appropriately chosen f, KLD dimension fits the true dimension over the whole range of $0<d<1$ slightly better than LVD dimension
- LVD dimension density better suited for detecting small changes in dimension of the system
- Both measures give reasonable estimates of the true dimension even for very small data sets

3. Application: Chemical Element Abundances

Description of the data

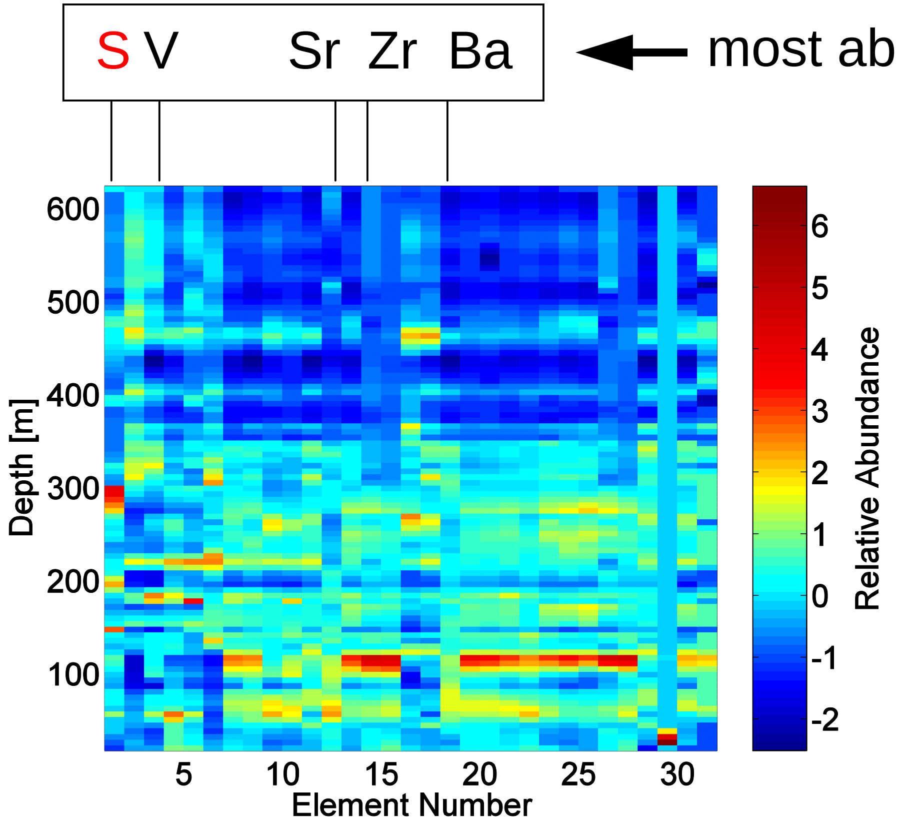


Marine sediment core CRP-2 (Cape Roberts Project, East Antarctica)

Record of 32 trace element abundances in 60 different time slices covering Oligocene/Miocene transition at about 24 Myr BP

3. Application: Chemical Element Abundances

Description of the data - continued



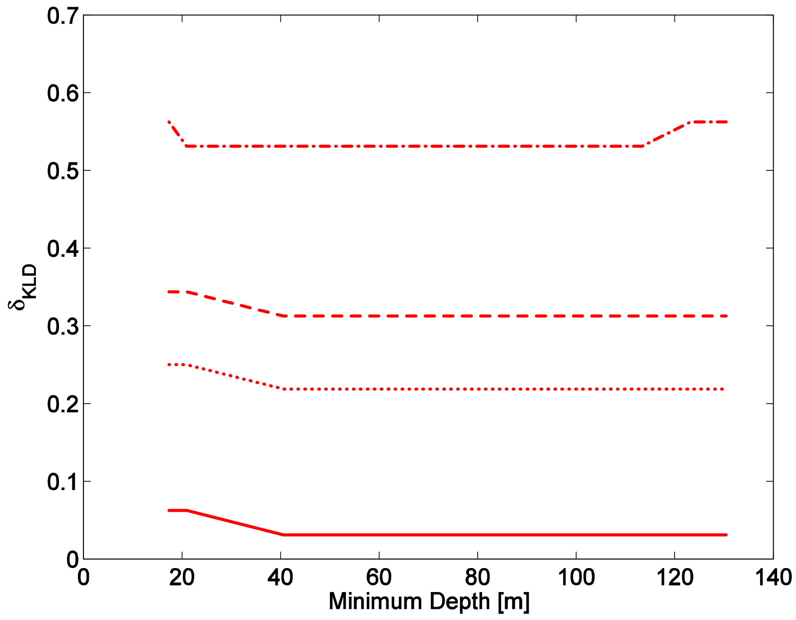
most abundant elements (>100 ppm)

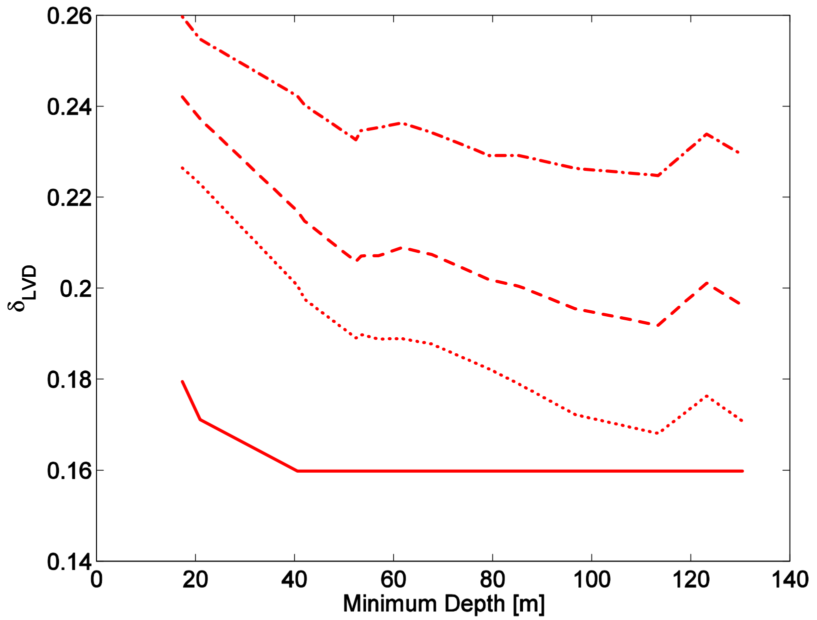
Inhomogeneous distribution of abundances of different trace elements – use data standardized to unit variance

Increase in variability in the younger part of the core

3. Application: Chemical Element Abundances

Performance of dimension estimates (M=45)



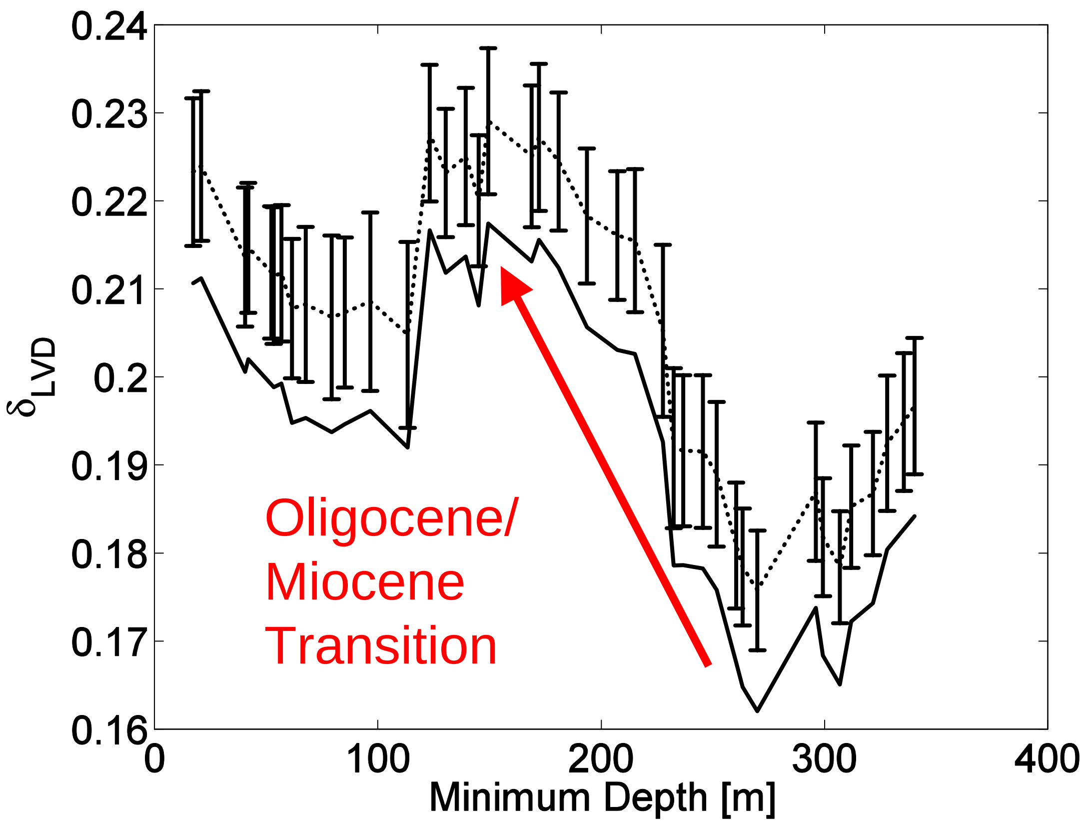


Solid: f=0.5, dotted: f=0.9, dashed: f=0.95, dash-dotted: f=0.99

LVD dimension density better suited for detecting changes !

3. Application: Chemical Element Abundances

Robustness for small windows (M=20, f=0.99)



Procedure: Substitute one randomly chosen time slice (5% of total data set) by a Gaussian random vector, repeat sufficient number of times (here: 1000)

[Figure: 95% confidence levels]

3. Application: Chemical Element Abundances

Palaeoclimatic Interpretation

Oligocene/Miocene transition (24.2 … 23.8 Myr before present):

- Further opening of Drake passage between Antarctica and South America
- Intensification of Antarctic Circumpolar Current
- Thermic Isolation of Antarctic continent
- Increase in glacial variability

⇒ Higher variability of sediment supply and change in composition (both recorded in the data)

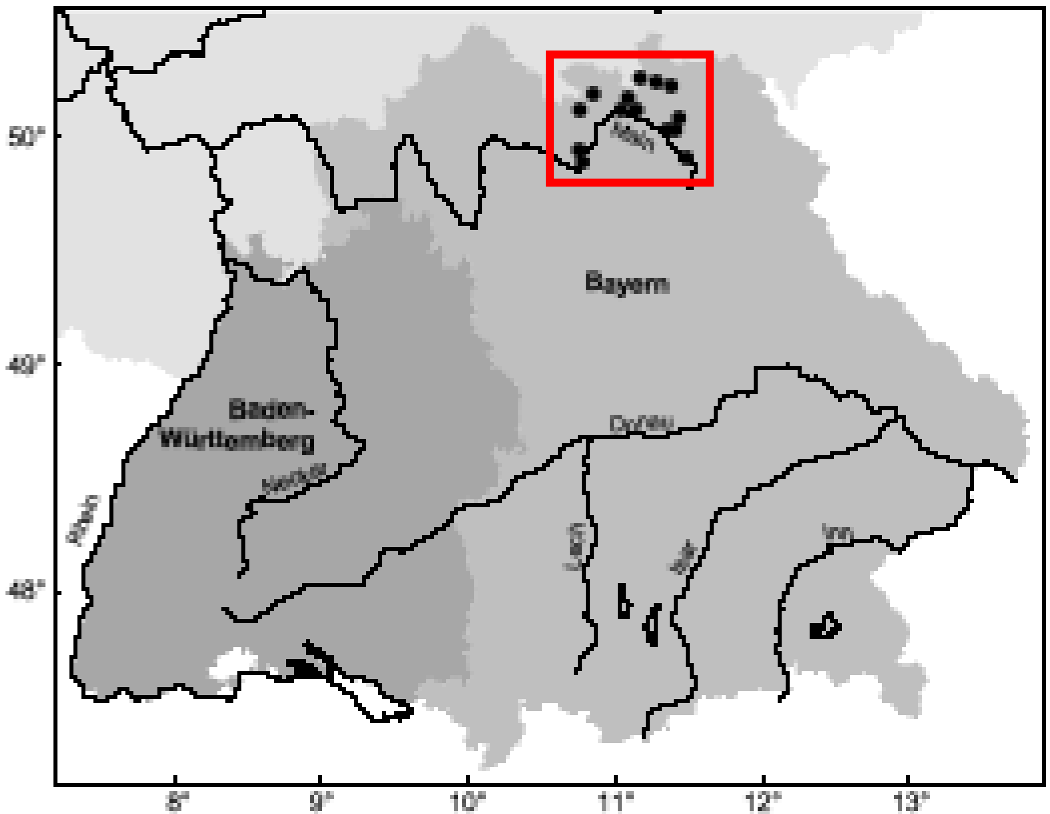
⇒ Decrease in correlations between different chemical element variations

⇒ Increase of dimension of multi-element record

(Details: R.D. and A. Witt, subm. to Int. J. Bif. Chaos)

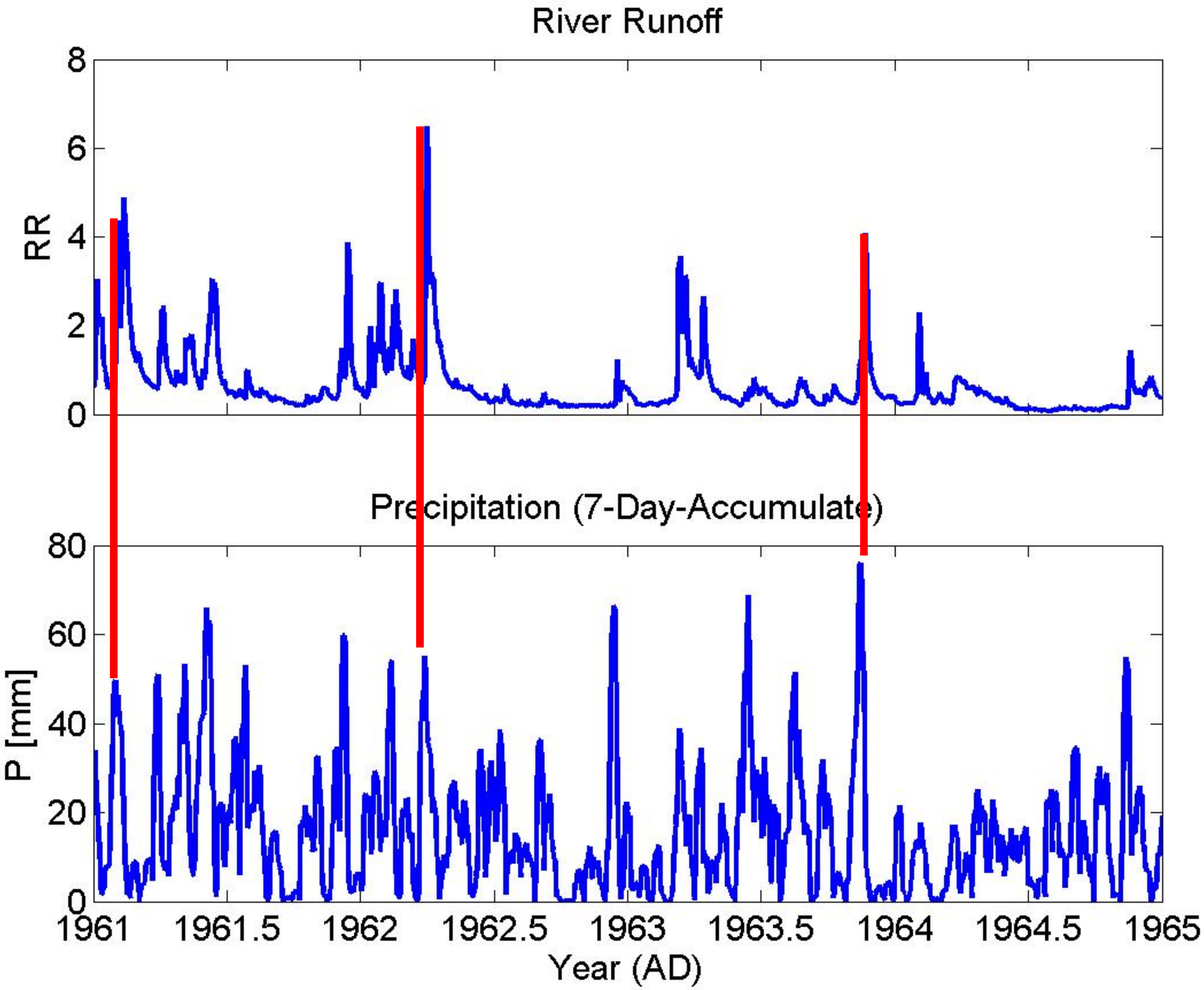
4. Application: River Runoff Data

<u>Data:</u> Upper Main river catchment area, Jan. 1961 – Dec. 1990
Daily river runoff rates (13 stations, see below)
Precipitation records (12 stations)

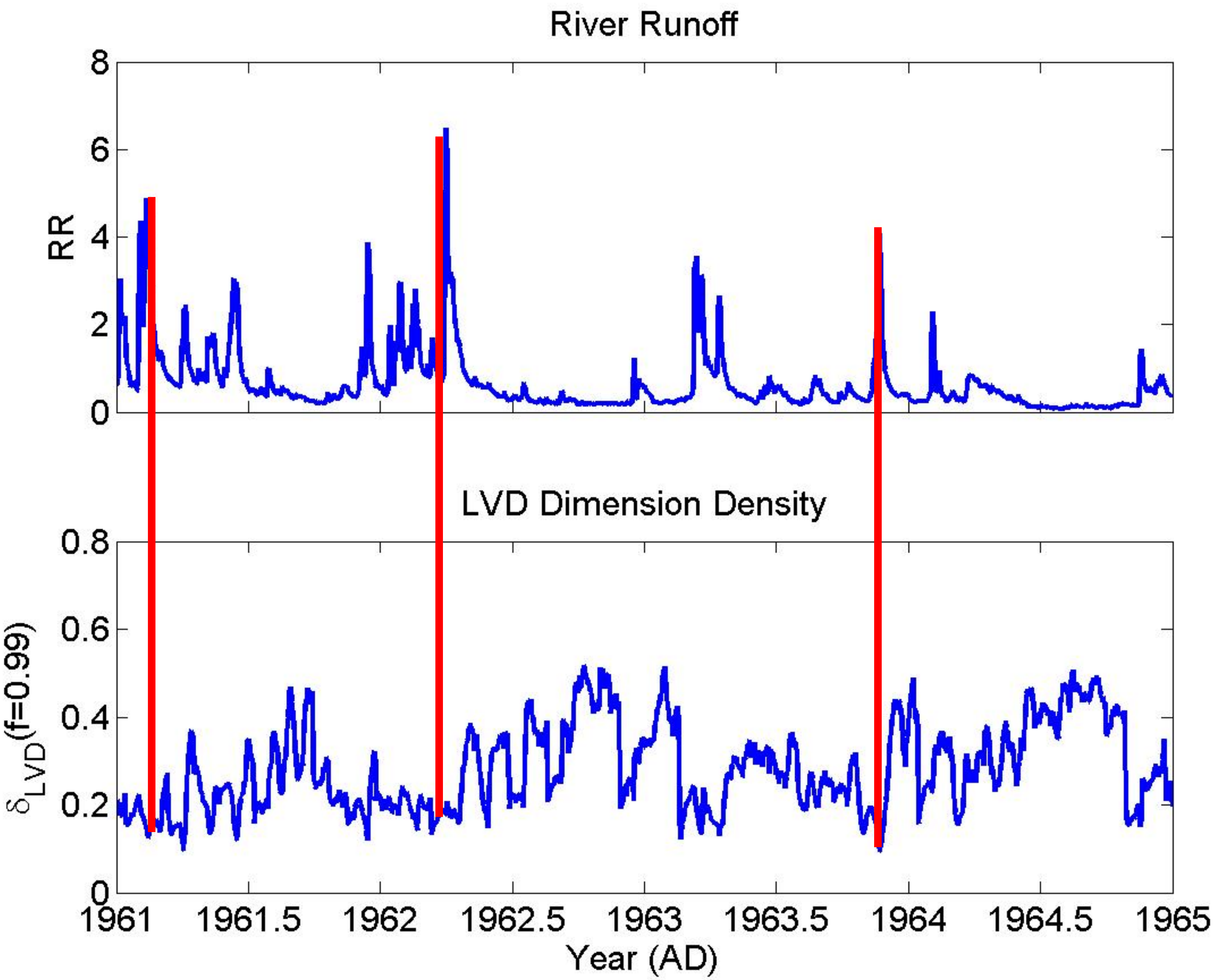


Kemmern
Leucherhof
Heinersdorf
Coburg
Schwürbitz
Horb
Unterlangenstadt
Steinberg
Wallenfels
Untersteinach
Ködnitz
Unterzettlitz
Bayreuth

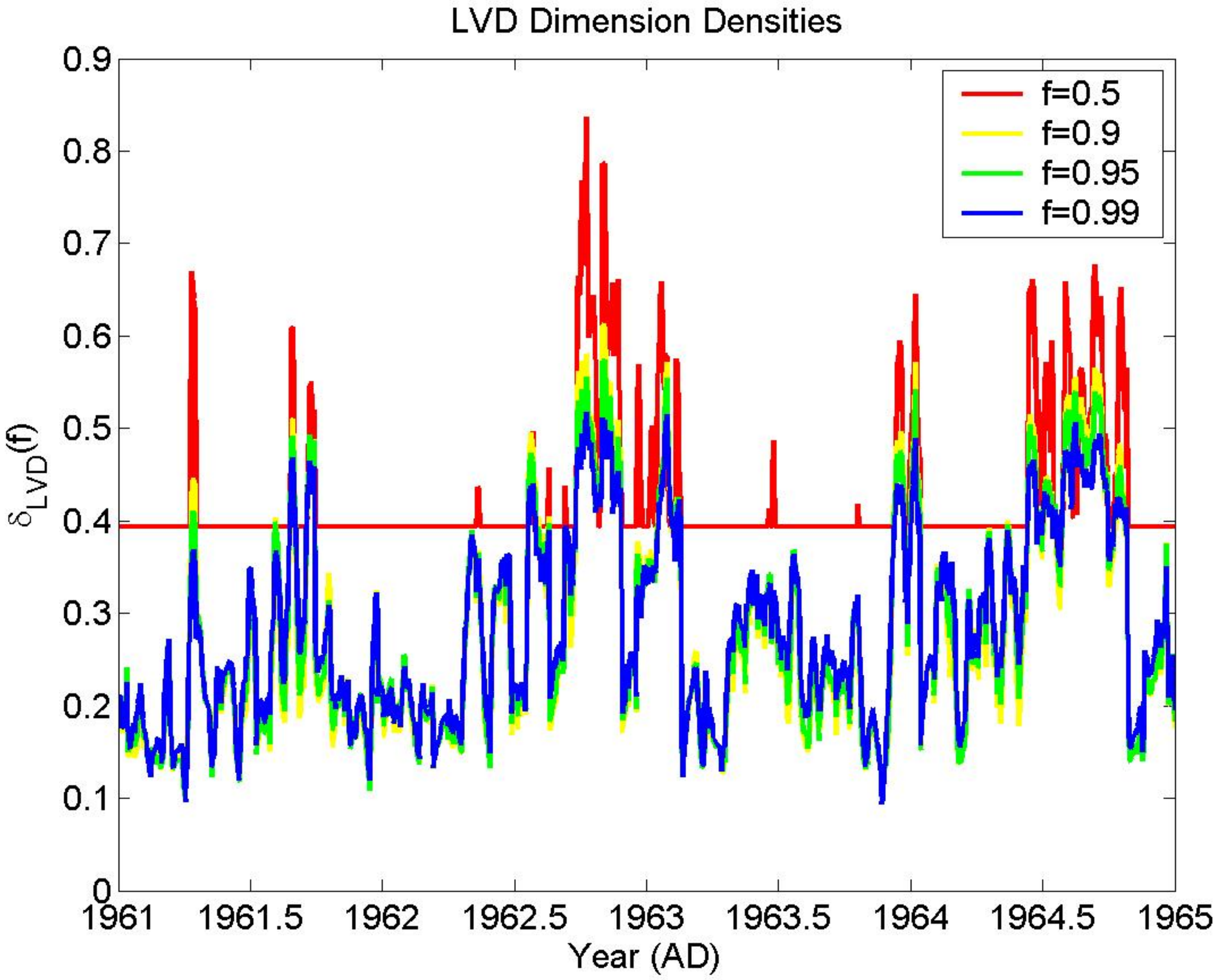
4. Application: River Runoff Data



4. Application: River Runoff Data



4. Application: River Runoff Data



4. Application: River Runoff Data

Preliminary results:

- Coincidence of extreme precipitation and river runoff
- δ_{LVD} detects extreme events by minimum values (sufficiently high cutoff fraction f required – the higher, the better)

Work in progress:

- Spatial correlation patterns
- Non-equal time correlations
- Robust dimension estimates for precipitation records (sparse correlation matrices)

5. Summary

- KLD-based estimates of fractal dimensions give robust, but (in dependence on f) relative values
- LVD dimension density (exponential scaling of remaining variances) detects qualitative changes in data structure
- Successful application to very short multivariate data sets
- Geological records: sensitivity to „extreme events“
- Hydrology: extreme river runoffs usually conincide with extreme precipitation events and are reflected by minimum dimension (i.e., maximum linear interrelationships) => work in progress

6. Open Questions

- More sophisticated models for decay of remaining variances
- Consideration of nonlinear decomposition techniques (NLPCA, ICA, Isomap) for larger data sets – linear vs. nonlinear variance decay dimensions
- Synchronization approach – application to geoscientific data
- Eigenvalues of non-equal time correlation matrices (possible relevance for hydrology ?)
- Meaning / relevance of method and results for hydrology ?

Some Advertisement…

EGU General Assembly 2006 Vienna (2-7 April 2006)

Learn more about new approaches in geoscientific data analysis: new methods and new applications of recent approaches

(Dimension reduction, Complexity, Wavelets,…)

Session NP 4.02: Advanced Methods of Data Analysis in Geosciences (convened by Annette Witt and R.D.)

And still not satisfied with extreme events ?

Session NH XX: Extreme Events – Causes and Consequences (convened by Bruce Malamud and Pascal Yiou)