A framework for quantification of groundwater dynamics - redundancy and transferability of hydro(geo-)logical indices

Benedikt Heudorfer [1], Ezra Haaf [2], Roland Barthel [2], Kerstin Stahl [1]

Background

- Groundwater time series can take various shapes. Which processes cause which shape?
- We need to quantify the dynamics as a tool to understand the causes
- Numerous indices for quantification are available from literature. We selected 63 that are suitable for groundwater time series.
- All 63 indices can describe the full range of variation. But is there a exhaustive subset without redundancy?

Objectives

Select a **subset of indices** that:

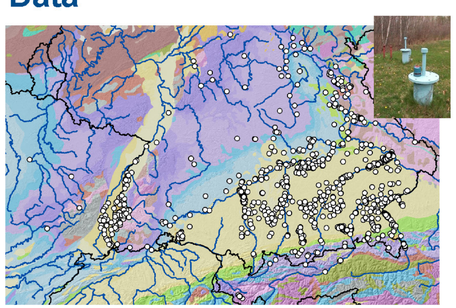
- is able to describe a **significant part of the overall variability** captured by the full set of 63 indices.
- can **describe all regime components**, i.e. there is indices for all regime components

Indices & Concepts

check poster **A49**

Groundwater Regime		
Structure	**Distribution**	**Shape**
Seasonality-Magnitude	Density	Slope
Seasonality-Timing	Modality	Scale
Inter-annual Forcing	Boundness	
Flashiness		

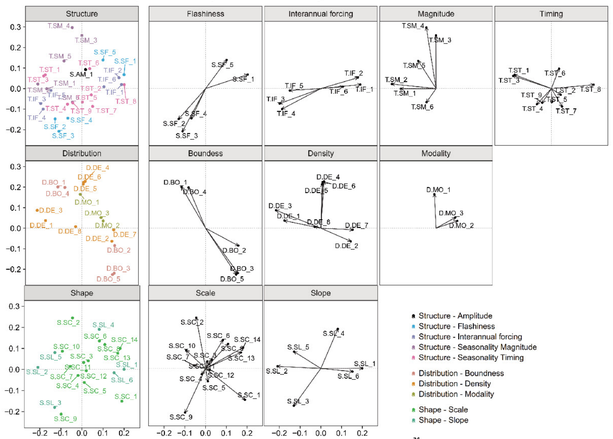
Data



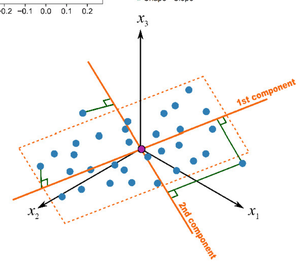
Points refer to the ~950 groundwater time series, for which the indices were calculated. Color refers to the geologc classes acording to the IGME5000 scheme.

PCA scores with redundancy

PCA = Principal Component analysis.

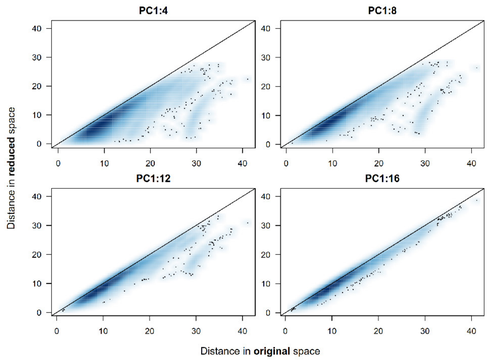


Above: Scatterplot of the loadings of the individual indices on principal component 1 and 2. Indices are grouped row-wise according to conceptual category 1, and column-wise according to category 2 (for explanation regarding conceptual categories **see twin poster #1404 (A49)**.



Visualisation of PCA.

Significant components - Shepard Diagram

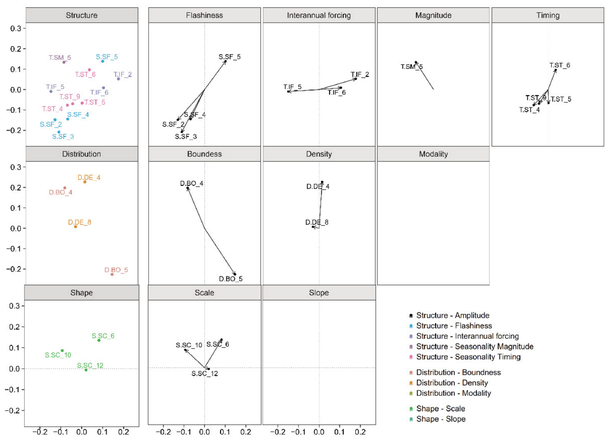


Convergence of the relation of euclidean distance of observations in original and PCA-space approaches the 1:1 line with increasing number of principal components.

- To determine the PCA-subspace Y of original space X (see below the PLS scheme), we need to identify the number of significant components of the PCA (see above). When the scatterplot approaches 1:1 line, the subspace closely resembles original space (here at 16).

Cumulative proportion of explained variance with 16 PC's: 83%.

Resulting PCA scores without redundancy



Scatterplot of the loadings of the individual indices on principal component 1 and 2 after redundancy analysis.

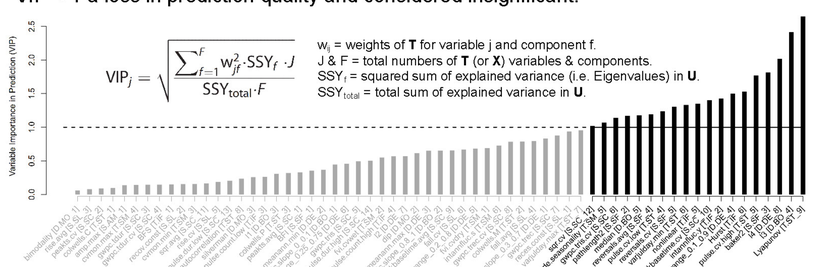
- Redundancy analysis yields 19 significant variables.
- This reduced set is able to cover most of the variability. (Note: plot shows only the first two principal component)
- However, not every category has a remaining significant index (e.g. modality & slope). These categories need to be revised.

Redundancy Analysis 2 - VIP

- The **Variable Importance in Prediction (VIP)** measures the impact of every variable in X on the subspace in Y. VIP > 1 imply a gain of prediction quality and considered significant, VIP < 1 a loss in prediction quality and considered insignificant.

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 \cdot SSY_f \cdot J}{SSY_{total} \cdot F}}$$

w_{ij} = weights of **T** for variable j and component f.
J & F = total numbers of **T** (or **X**) variables & components.
SSY_f = squared sum of explained variance (i.e. Eigenvalues) in **U**.
SSY_{total} = total sum of explained variance in **U**.

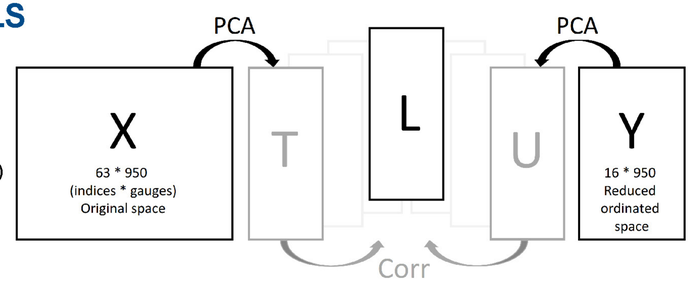


VIP-values for all 63 indices regarded in the study.

Redundancy Analysis 1 - PLS

Partial Least Squares Regression a.k.a. Projection on Latent Structures (PLS)

- The PLS method is a direct ordination method used in canonical analysis (determination of significant subset of predictors X on variables Y)
- Here it is innovatively used as selection method within indirect ordination by taking Y as the scores of the significant PCA-subspace spanned by X.
- Objectification of classical selection methods using matrix correlation of X and Y.



Scheme of the PLS method. The latent space L is the best agreement between X and Y. VIP can be performed on L.

Conclusion

- Definite set of significant indices that explains majority of variation
- Objective approach with a minimum amount of subjectivity & no collinearity
- The results can be used to link patterns of groundwater dynamics to governing processes

(1) Hydrology, Faculty of Environment and Natural Resources, University of Freiburg, Germany
benedikt.heudorfer@hydro.uni-freiburg.de www.hydrology.uni-freiburg.de

(2) Department of Earth Sciences, University of Gothenburg, Sweden
ezra.haaf@gu.se www.gvc.gu.se