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 an 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

Data

PCA scores with redundancy

PCA = Principal Component analysis.

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).

Visualisation of PCA.

Significant components - Shepard Diagram

Cumulative proportion of explained variance with 16 PCs: 83%.

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 indirced 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.

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.

Conclusion
- Define 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

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.

Resulting PCA scores without redundancy

- 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.

Conclusion
- Define 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

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.

 Points refer to the ~950 groundwater time series, for which the indices were calculated. Color refers to the geologic classes according to the IGMES0000 scheme.