

## Data-driven derivation of geoclimatic regions from high-dimensional data for Austria

Sebastian Lehner<sup>1</sup>, Katharina Enigl<sup>1</sup>, Matthias Schlögl<sup>1</sup>

<sup>1</sup> GeoSphere Austria, Department Climate-Impact-Research

Geoclimatic regions represent climatic forcing zones, which constitute important spatial entities that serve as a basis for a broad range of analyses in earth system sciences. The plethora of geospatial variables that are relevant for obtaining consistent clusters represent a high-dimensionality, especially when working with high-resolution gridded data, which may render the derivation of such regions complex. This is worsened by typical characteristics of geoclimatic data like multicollinearity, nonlinear effects and potentially complex interactions between features. We therefore present a nonparametric machine learning workflow, consisting of dimensionality reduction and clustering for deriving geospatial clusters of similar geoclimatic characteristics. We demonstrate the applicability of the proposed procedure using a comprehensive dataset featuring climatological and geomorphometric data from Austria, aggregated to the recent climatological normal from 1992 to 2021.

The modelling workflow consists of three major sequential steps: (1) linear dimensionality reduction using Principal Component Analysis, yielding a reduced, orthogonal sub-space, (2) nonlinear dimensionality reduction applied to the reduced sub-space using Uniform Manifold Approximation and Projection, and (3) clustering the learned manifold projection via Hierarchical Density-Based Spatial Clustering of Applications with Noise. The contribution of the input features to the cluster result is then assessed by means of permutation feature importance of random forest models. These are trained by treating the clustering result as a supervised classification problem. Results show the flexibility of the defined workflow and exhibit good agreement with both quantitatively derived and synoptically informed characterizations of geoclimatic regions from other studies. However, this flexibility does entail certain challenges with respect to hyperparameter settings, which require careful exploration and tuning. The proposed workflow may serve as a blueprint for deriving consistent geospatial clusters exhibiting similar geoclimatic attributes.