ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

A combination of cluster analysis and Kappa statistic for the evaluation of climate model results

M. Kuecken[PIK], F.-W. Gerstengarbe[PIK] and B. Orlowsky[ETH]

Methodology

Objectives

- Comparison of spatial structures in climate model simulations
- Comparison of differences against random deviations
- Comparison not limited to regular grids
- Comparison expressed in few meaningful statistics
- Identification of regions of good/bad agreement

Tools

- Combination of hierarchical and non-hierarchical cluster analysis for identification of spatial structures
- κ-statistic to quantify agreement between two categorical maps
- Different κ-variants to distinguish between different sources of differences, e. g. overall distributions of classes and spatial distribution of classes

κ

Definitions:

$$p_{ab} = \frac{\text{number of points } (i,j) \text{ for which } o_{i,j} \in C_a \text{ and } s_{i,j} \in C_b}{\text{number of grid points}}$$

$$p_{a\cdot} = \sum_{k=1}^{n} p_{ak}, \quad p_{\cdot b} = \sum_{k=1}^{n} p_{kb}, \quad h_0 = \sum_{k=1}^{n} p_{kk}, \quad h_{\text{random}} = \sum_{k=1}^{n} p_{k\cdot} p_{\cdot k}$$

$$\kappa = \frac{h_0 - h_{\text{random}}}{1 - h_{\text{random}}}$$

Variants:

$$\kappa_{\text{histo}} = \frac{\sum_{k=1}^{n} \min(p_{k\cdot}, p_{\cdot k}) - \sum_{k=1}^{n} p_{k\cdot} p_{\cdot k}}{1 - \sum_{k=1}^{n} p_{k\cdot} p_{\cdot k}}$$

$$\kappa_{\text{location}} = \frac{\sum_{k=1}^{n} (p_{kk} - p_{k\cdot} p_{\cdot k})}{\sum_{k=1}^{n} \min(p_{k\cdot}, p_{\cdot k}) - \sum_{k=1}^{n} p_{k\cdot} p_{\cdot k}}$$

Note:

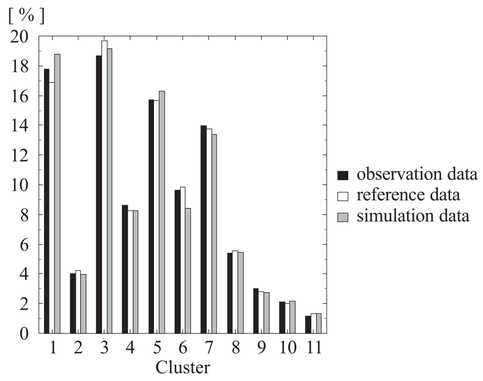
$$\kappa = \kappa_{\text{histo}} \cdot \kappa_{\text{location}}$$

Application

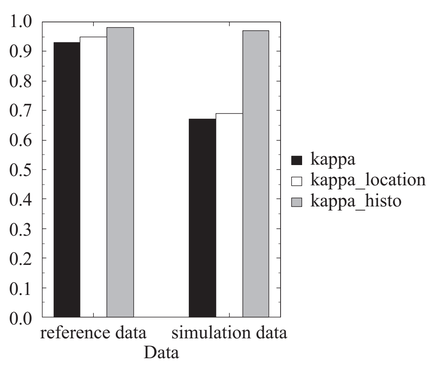
Data

- Observed monthly precipitation at 2342 stations (German Weather Service DWD) 1976 to 2000 – $o_{i,j}$
- Reference data: $r_{i,j} = o_{i,j}$+noise
- Simulated monthly precipitation [Orlowsky et al., 2008], same stations, same period: $s_{i,j}$
- Cluster analysis based on mean, variance, skewness, kurtosis, extremes, trend

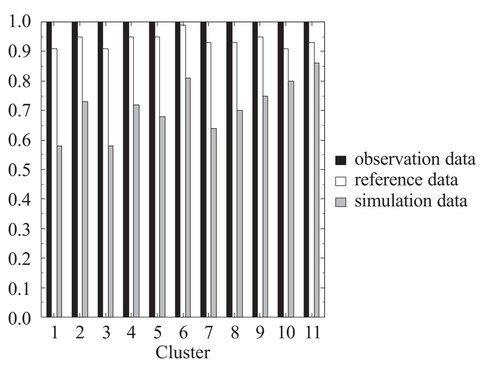
Results



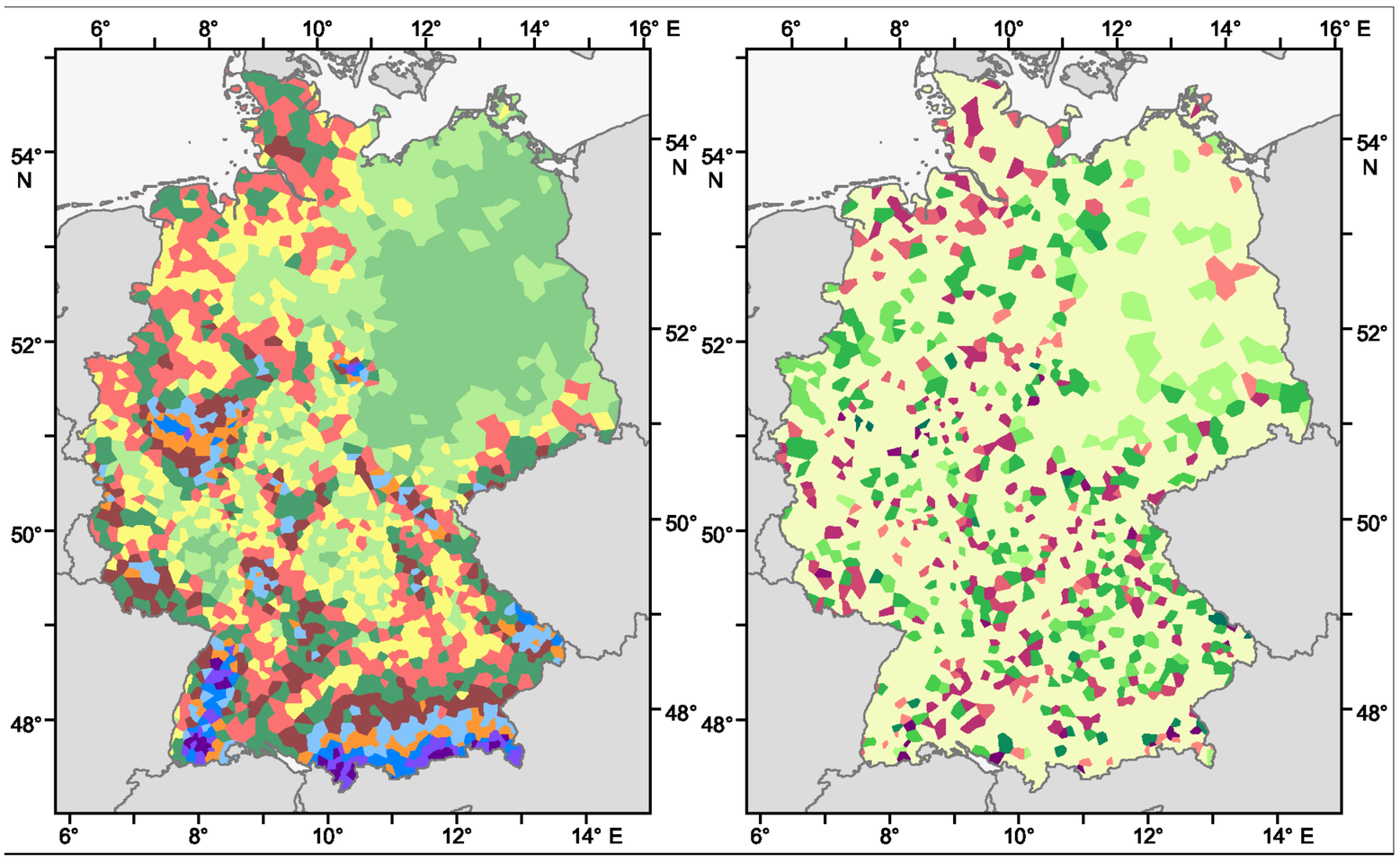
(a) Histogram of the cluster distribution



(b) κ-variants for reference and simulation data



(c) Individual $\kappa_{location}$ for each cluster



Rainfall clusters over Germany (left) and cluster changes between observation and simulation data (right).

References

Kuecken, M., Gerstengarbe, F.-W., and Orlowsky, B. (2009).
A combination of cluster analysis and Kappa statistic for the evaluation of climate model results.
Journal of the American Meteorological Society, available online.

Orlowsky, B., Gerstengarbe, F.-W., and Werner, P. (2008).
A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM.
Theoretical and Applied Climatology, 92:209–223.

Affiliations

PIK: Potsdam Institute for Climate Impact Research, Potsdam, Germany

ETH: Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland