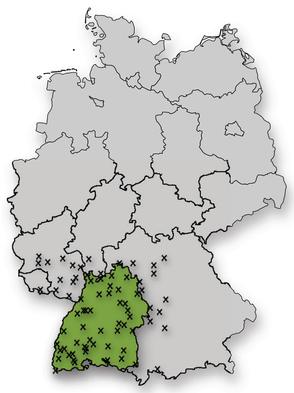
Probabilistic assessment of regional climate change by ensemble dressing

Christian Schölzel and **Andreas Hense**

Meteorological Institute at the University of Bonn
Bonn, Germany

Objectives and Outline



Objectives

- Provide short-term projections of regional climate change in the federal state of Baden-Württemberg
- Quantify the uncertainty of the projections by using ensemble simulations of regional climate models (RCM)

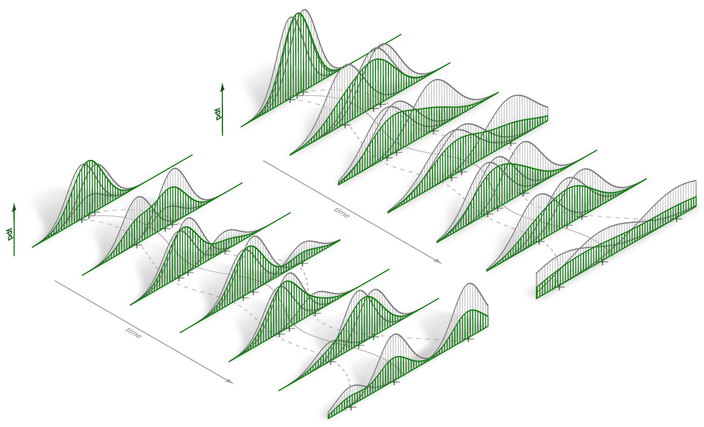
Outline

- Discuss different concepts of probabilistic ensemble post-processing in numerical weather prediction and climate change
- Introduce a multivariate extension to the classical methods of ensemble kernel dressing
- Incorporate the temporal autocovariance structure of the ensemble system

Background

Problem statement

A common way to estimate the uncertainty in climate simulation is given by ensemble simulations. The question to be answered is how ensembles can be translated into probabilistic information. The answer is not clear and strongly depends on the point of view and how the different ensemble members are generated, i.e. what they are supposed to represent.



State-of-the-art

In recent years, a variety of approaches to quantify uncertainty in numerical simulations in a probabilistic way has been attempted. They are typically referred to as *statistical post-processing of ensemble simulations* or *ensemble interpretation*. Most of these methods originate in numerical weather prediction and are therefore known under the term *ensemble MOS (Model Output Statistics)*, but can be brought forward to climate research, e.g.:

GDF: Gaussian DF interpretation
NGR: Non-homogeneous Gaussian regression
SKD: Standard kernel dressing
GED: Gaussian ensemble dressing
AKD: Affine kernel dressing
BMA: Bayesian model averaging
BHM: Bayesian hierarchical modelling

The results can be significantly different as illustrated above for Bayesian model averaging (left) and affine kernel dressing (right).

Ensemble kernel dressing

Ensemble dressing methods are closely related to the principle of kernel density estimation. For each ensemble member j a probability density function (kernel) f_K is defined around the model prediction X_j and the resulting distribution is given by an equally weighted mixture model

$$f_{ens}(y|\vec{x}) = \frac{1}{m}\sum_{j=1}^{m} f_K\left(\frac{y-(a\cdot x_j+\omega)}{\sigma_D}\right)$$

with scaling a, offset ω and the so-called *dressing variance* σ_D also referred to as *bandwidth*. The method derived in the following section is based on an equally weighted GED:

- The ensemble members (see below) are seen as indistinguishable, i.e. none of them performs better based on the prior setup.
- The output variables are seasonal averages of temperature which can be assumed to follow a Gaussian distribution.
- The statistical downscaling (see below) already incorporates observation data, so that dressing variance is estimated by the internal uncertainty in the ensemble system.

Methods

Multivariate Gaussian kernel dressing

The classical methods of kernel dressing can be extended to a multivariate approach. Analogously to univariate Gaussian ensemble dressing (GED) the probabilistic prediction is given by the multidimensional random vector $\vec{Y}$ and the ensemble members $\vec{X}_j$. Within this study the random vector components describe the temporal structure, so $\vec{Y}$ and $\vec{X}_j$ have dimension n_t, i.e. the time steps in the projection period. Hence, the conditional probability density of $\vec{Y}$ given the ensemble members follows a multivariate Gaussian mixture model

$$f_{ens}^{(mvt)}(\vec{y}|\vec{x}_1,\ldots,\vec{x}_m) = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{\sqrt{\cdots}}\exp\left(-\frac{1}{2}(\vec{y}-\vec{x}_j)^T\Sigma_D^{-1}(\vec{y}-\vec{x}_j)\right),$$

where $\sqrt{\cdots}$ denotes the typical normalisation constant for a n_t-dimensional Gaussian distribution.

The univariate dressing variance or bandwidth σ_D naturally extends to the *dressing covariance* Σ_D, which is not only intended to reflect the uncertainty in the ensemble but also its autocovariance structure. Assuming an unbiased ensemble and indistinguishable simulations the dressing covariance is proportional to the average error covariance $\Sigma_D = h_{\text{opt}}\cdot\Sigma_{\vec{\epsilon}}$. The additional factor h_{opt} accounts for the criterion which optimizes the choice of bandwidth and is based on the theory by Silverman (1986)

$$h_{\text{opt}} = \left(\frac{4}{m(n_t+2)}\right)^{\frac{1}{n_t+4}}.$$

It is an approximate measure derived by minimizing the mean integrated square error and the optimal choice of bandwidth for multivariate kernel density estimation using a Gaussian kernel.

Autocovariance structure

In the following we regard mean and trend over time by $\vec{Z} = (Z_0, Z_1)^T$. By defining a transformation $\vec{Y} \overset{P}{\longmapsto} \vec{Z}$ for mean Z_0 and trend Z_1 within the period $1,\ldots,n_t$

$$\vec{Z} = \underbrace{(PP^T)^{-1}P}_{Q}\vec{Y} \quad \text{and} \quad P = \frac{1}{n_t}\begin{pmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & 2 & 3 & \cdots & n_t\end{pmatrix}$$

we can analytically derive the density for mean and trend of $\vec{Y}$

$$f_{ens}(\vec{z}|\vec{x}_1,\ldots,\vec{x}_m) = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{\sqrt{\cdots}}\exp\left(-\frac{1}{2}(\vec{z}-Q\vec{x}_j)^T(Q^T\Sigma_D Q)^{-1}(\vec{z}-Q\vec{x}_j)\right).$$

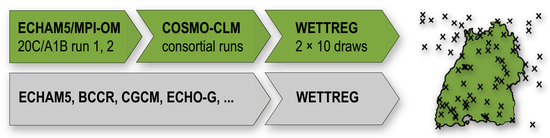
Furthermore, in the ideal case that the autocovariance is constant over time, the estimators of the error covariance $\hat{\Sigma}_{\vec{\epsilon}}$ converge to a Toeplitz structure. This property of $\hat{\Sigma}_{\vec{\epsilon}}$ is requested as condition on the structure of the estimator

$$\hat{\Sigma}_{\vec{\epsilon}} = \begin{pmatrix}\hat{\sigma}_{\vec{\epsilon}}(0) & \hat{\sigma}_{\vec{\epsilon}}(-1) & \cdots & \cdots & \hat{\sigma}_{\vec{\epsilon}}(-n_t+1)\\ \hat{\sigma}_{\vec{\epsilon}}(1) & \hat{\sigma}_{\vec{\epsilon}}(0) & & & \vdots\\ \vdots & & \ddots & & \vdots\\ \vdots & & & \hat{\sigma}_{\vec{\epsilon}}(0) & \hat{\sigma}_{\vec{\epsilon}}(-1)\\ \hat{\sigma}_{\vec{\epsilon}}(n_t-1) & \cdots & \cdots & \hat{\sigma}_{\vec{\epsilon}}(1) & \hat{\sigma}_{\vec{\epsilon}}(0)\end{pmatrix}.$$

where the autocovariance function $\hat{\sigma}_{\vec{\epsilon}}$ is estimated as follows. Let $x_j^{(t)}$ denote individual ensemble member $j = 1,\ldots,m$ for the training period $t = 1,\ldots,\tilde{n}_t$ as named below. Since the ensemble is assumed to be unbiased, each realisation can be written as true state and error. The combinations of all possible differences between the ensemble members, denoted by $\mathcal{P}$ with length $|\mathcal{P}|$, therefore provide an estimator of the error autocovariance

$$\hat{\sigma}_{\vec{\epsilon}}(\tau) = \sum_{t=|\tau|+1}^{\tilde{n}_t}\sum_{(i,j)\in\mathcal{P}}\frac{\left(x_i^{(t)}-x_j^{(t)}\right)\left(x_i^{(t-\tau)}-x_j^{(t-\tau)}\right)}{2\tilde{n}_t|\mathcal{P}|}.$$

Data



Ensemble members

- The aforementioned method is applied to the existing *consortial runs* of COSMO-CLM with a spatial resolution of approx. 18 km.
- The simulations are examined for the periods 1961–2000 and 2011–2030, so there is a comparably low dependence on the specific emission scenario (SRES A1B).
- An additional downscaling and ensemble generation by WETTREG is used. WETTREG is a combination of pattern recognition, stochastic weather generator, and regression scheme (not detailed here).

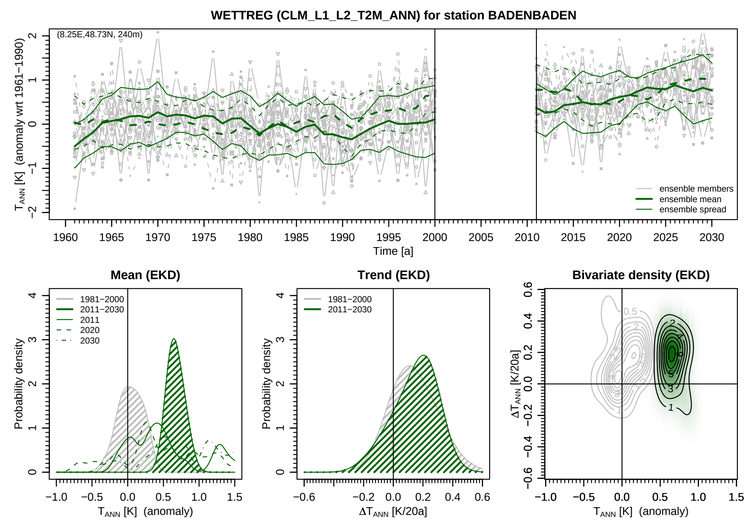
Application

In this case, we are given 10 realisations of WETTREG for each model run and we obtain 45 combinations of all differences between the ensemble members, which allows to estimate $\hat{\Sigma}_{\vec{\epsilon}}$ for each of the 73 stations in Baden-Württemberg.

Results

Point projections

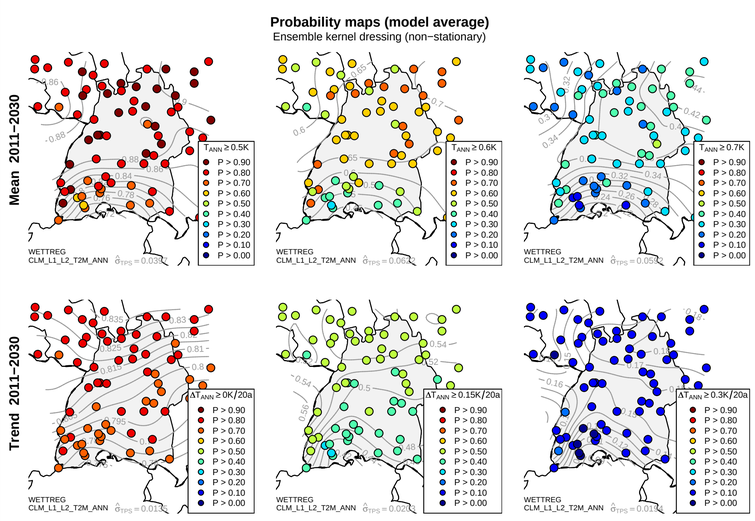
The results are given as bivariate distributions for mean and trend of annual mean temperature with respect to 1961–1990.



- The mean in 2011–2030 increases by approx. $+0.6\,K$ wrt. 1961–1990 with a comparably small uncertainty.
- The expected value of the trend in 2011–2030 is approx. $+0.15\,K/20\,a$ and shows a larger variability. There is even a 25% chance for a negative trend which indicates the uncertainty in trend determination.

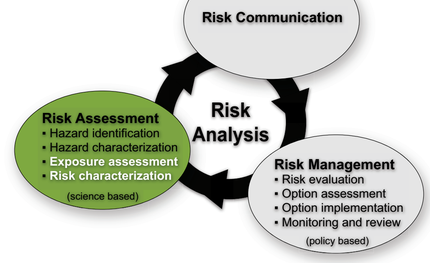
Spatial structure

A main aspect of this study is the spatial structure of temperature change in Baden-Württemberg. Since the probabilistic point of view requires to go beyond plotting pointwise expectation values, the results can be given e.g. as probability maps for the exceedances of preselected thresholds.



- Throughout the different thresholds a north-to-south pattern in the anomalies of the annual mean temperature (2011–2030) can be observed.
- The spatial structure of the trend (2011–2030) shows a similar but less distinct pattern.

Conclusions



Dissemination

- An important objective of this project (PArK) is to communicate the additional value of probabilistic climate projections to sponsors and decision makers.
- Assume a protective action that is dependent on the outcome of a RCM simulation. Probabilistic approaches in general allow to estimate the expected expense in terms of the relation between protected loss, unprotected loss, and costs of the protective action.

Outlook

- Apply the multivariate kernel dressing to precipitation amounts, i.e. incorporate non-Gaussian kernel functions, for example by using copulas.
- Apply to additional RCM simulations (without WETTREG), i.e. introduce RCM/GCM specific weights by pattern recognition.

Acknowledgments. This work is part of the PArK project which is supported by the Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW).