Bayesian multi-model projection of climate: bias assumptions and interannual variability

Christoph M. Buser · H. R. Künsch ·
D. Lüthi · M. Wild · C. Schär

Abstract Current climate change projections are based on comprehensive multi-model ensembles of global and regional climate simulations. Application of this information to impact studies requires a combined probabilistic estimate taking into account the different models and their performance under current climatic conditions. Here we present a Bayesian statistical model for the distribution of seasonal mean surface temperatures for control and scenario periods. The model combines observational data for the control period with the output of regional climate models (RCMs) driven by different global climate models (GCMs). The proposed Bayesian methodology addresses seasonal mean temperatures and considers both changes in mean temperature and interannual variability. In addition, unlike previous studies, our methodology explicitly considers model biases that are allowed to be time-dependent (i.e. change between control and scenario period). More specifically, the model considers additive and multiplicative model biases for each RCM and introduces two plausible assumptions ("constant bias" and "constant relationship") about extrapolating the biases from the control to the scenario period. The resulting identifiability problem is resolved by using informative priors for the bias changes. A sensitivity analysis illustrates the role of the informative prior. As an example, we present results for Alpine winter and summer temperatures for control (1961–1990) and scenario periods (2071–2100) under the SRES A2 greenhouse gas scenario. For winter, both bias assumptions yield a comparable mean warming of 3.5–3.6°C. For summer, the two different assumptions have a strong influence on the probabilistic prediction of mean warming, which amounts to 5.4°C and 3.4°C for the "constant bias" and "constant relation" assumptions, respectively. Analysis shows that the underlying reason for this large uncertainty is due to the overestimation of summer interannual variability in all models considered. Our results show the necessity to consider potential bias changes when projecting climate under an emission scenario. Further work is needed to determine how bias information can be exploited for this task.

1 Introduction

Climate projections and associated applications in impact studies have become an important topic of scientific and public interest during the last decades. Several research teams around the world are developing models to simulate the current climate and its future evolution under several greenhouse gas and aerosol scenarios.

On the large scale, general circulation models (GCMs) are used with coarse horizontal resolution. While they are capable of effectively reproducing large-scale effects and circulation patterns, they cannot predict small-scale effects for a selected region. Information about regional climate can be obtained by dynamic down-scaling (Giorgi 1990). To this end, regional climate models (RCMs) use the GCM output as their driving boundary conditions. It is advantageous to combine different results of several climate

C. M. Buser (✉) · H. R. Künsch
Seminar for Statistics, ETH Zurich,
Rämistrasse 101, 8092 Zurich, Switzerland
e-mail: buser@stat.math.ethz.ch

D. Lüthi · M. Wild · C. Schär
Institute for Atmospheric and Climate Science,
ETH Zurich, Zurich, Switzerland

models—both on the global and regional scale—to obtain a reliable data base. It is generally believed that multi-model ensembles are superior to single models, and that the ensemble may even outperform the best single participating model. Recent analysis indicates that much of this gain is due to the fact that single models are overconfident (Weigel et al. 2008). In relation to climate projections, combining different models exploits the strengths of diverse approaches and yields a more appropriate estimate of the uncertainties (Meehl et al. 2007). The combined GCM/RCM multi-model approach has been advanced by large international projects such as PRUDENCE (e.g. Christensen and Christensen 2007; Christensen et al. 2007b).

Once a large multi-model ensemble is available, one is left with the task of optimally combining this information into one probabilistic prediction of the anticipated changes in climate. In the case of medium-range weather forecasts and seasonal climate prediction, several methods exist (for an overview see Wilks 2006). Many of these methods address the task by assigning (equal) weights to all ensemble members and by subtracting the biases of each model, as known from past model performance. However, in a multi-model climate change ensemble, there are additional issues that should be considered. One would like to predict the whole climate distribution, in particular higher moments and quantiles, and there is the additional complication that the climate model biases can depend on the underlying climate, i.e. the biases are time- and state-dependent.

The last item appears particularly difficult. Indeed, the standard procedure in studies about climate change entails the implicit assumption that bias changes are negligible compared to changes in climate, i.e. the consideration of "climate change" defined as difference between scenario and control climate. This important assumption is rarely discussed in depth (but see e.g. Shackley et al. 1998), and a thorough test appears elusive, as the changes in climate considered are of a magnitude that have not occurred in the instrumental past. Yet the assumption of a time-independent (or climate-state-independent) bias is crucial. Even with a model that perfectly reproduces the current climate, there is no guarantee that the model will exhibit the true climate sensitivity (Stainforth et al. 2005). Also from a physical viewpoint, it appears unlikely that the biases of a climate model should be state-independent, as the climate system entails many non-linearities and threshold processes (e.g. related to atmospheric humidity, freezing/melting, sea ice, soil moisture, clouds, convection, etc). One method to address the role of these nonlinearities on the simulation of climate is to separately validate summer and winter seasons (e.g. Meehl et al. 2007) and to use the representation of the

seasonal cycle as a measure of the model's fidelity (Shukla et al. 2006).

The Bayesian framework is particularly attractive for combining several models. It decomposes the complicated relationship between the observations and the outputs of different models into simpler, hierarchical relationships that can be described in a reasonable and transparent way. (Gelman et al. 2003). Although the necessary integrations cannot be done analytically, Markov Chain Monte Carlo methods make it possible to deal with complicated distributions (Gilks et al. 1996).

Tebaldi et al. (2005) were among the first to use the Bayesian framework to analyze multi-model climate predictions. They obtain a probability density function (PDF) for the mean temperature changes in 22 global regions and four seasons by combining observations and output from several GCMs of 30 year regional climate averages. Their approach can be viewed as a weighted average of the individual GCM results, with weights similar to those used by the reliability ensemble average (REA) of Giorgi and Mearns (2002). The framework of Tebaldi et al. (2005) has been generalized in many directions. Smith et al. (2008) study several regions simultaneously. Tebaldi and Sanso (2008) introduce a multivariate generalization for analyzing decadal averages of temperature and precipitation for 1955–2100. Furrer et al. (2007) analyze the spatial variability of the climate change signal. They use a multivariate hierarchical Bayes model to separate it into a large scale signal of climate change and an isotropic process representing small-scale variability among models. Jun et al. (2008) analyze the spatial variability of the additive bias in detail for the control climate. Min and Hense (2007) calculate Bayes factors for a weighted multi-model average. These Bayes factors are obtained by comparing the simulations to a reference model in terms of likelihood. Sain et al. (2008) provide a multivariate approach that takes into account the spatial structure of the data. Bayesian methods are also used to aggregate station data on a regular grid for an RCM validation (Snyder et al. 2007). A review of multi-model climate projections and the different types of uncertainty is given by Tebaldi and Knutti (2007). They also discuss the problems of model dependence, tuning and evaluation.

Our approach is a different extension of Tebaldi et al. (2005). We study RCMs instead of GCMs, but the main methodological difference is that we consider not only the long-term climate mean, but also the interannual variations, by focusing on the distribution of seasonal values of the variable of interest. A possible nonstationarity of the data is taken into account by including linear trends in the control and scenario periods. For simplicity, we assume that all models have the same underlying trend.

The reason for analyzing the distribution of seasonal values is two-fold. First for impact studies, both changes in mean and variability of the climate variables are relevant (Katz and Brown 1992; Schär et al. 2004), and our approach provides this. Second, the broader approach allows us to study additive and multiplicative biases of the different RCMs in the Bayesian framework. We discuss two different assumptions for extrapolating the biases into the scenario period, which both are plausible, but lead to quite different conclusions about the likely climate changes. We can even allow these biases to be different in the control and scenario period, but we have to assume in the prior distributions that the bias changes are small.

In this paper, our variable of interest is the seasonally and regionally averaged 2 m-temperature, but other variables could in principle be considered, e.g. the regional average of the maximum temperature within a season. However, complications will arise if the assumption of normal distributions of the variables is no longer valid. We will restrict attention to the target variable (i.e. temperature), and biases in other variables, e.g. precipitation, do not enter the analysis. This procedure has also been followed by other studies (Giorgi and Mearns 2002; Tebaldi et al. 2005), although it would be desirable to account for the overall performance of a model as in the multivariate extension of Tebaldi and Sanso (2008).

The paper is structured as follows. In Sect. 2 the data and the aggregation procedure are described. In Sect. 3 the methods and the Bayesian model setup are explained. In Sect. 4 results for the Alpine region are shown. In the final Sect. 5 we draw conclusions and discuss further extensions of our approach.

2 Data

In this paper both observational data and output from the RCMs are summarized by the term "data". One has to distinguish between current climate data that comes from observations and model projections, and the future climate data that comes from models only. Our variable of interest

is the 2 m-temperature, but the same methods apply to other variables in principle. Some of the problems that can arise for other variables are discussed in Sect. 5.

2.1 Regional climate model data

For the statistical analysis there is the output of 4 RCMs (CHRM, CLM, HIRHAM, RCAO) and 1 high-resolution GCM with a stretched spectral discretization (Arpege). All simulations are part of the PRUDENCE project (http://prudence.dmi.dk) or use the PRUDENCE methodology in their set-up. Here we restrict the attention to the most salient aspects and refer to the literature for a full documentation of the numerical experiments (Christensen et al. 2007a; Christensen and Christensen 2007).

Each model has been run as a control run for the period 1961–1990 (the present) and a scenario run for 2071–2100 (the future) using an A2 emission scenario (Nakicenovic et al. 2000). All models are driven by different lateral boundary conditions as derived from global atmospheric simulations. Boundary conditions for the control runs are taken from the GCMs HadAM3H (Jones et al. 2001; Pope et al. 2000), ECHAM4/OPYC (Roeckner et al. 1996) and ECHAM5 (Roeckner et al. 2003). In Table 1 a short summary of all regional models is given. RCM data has been provided by the PRUDENCE data archive. Although there are more runs from other climate research groups in the PRUDENCE data archive, we use a subset of models that are driven by different atmospheric GCM runs. RCMs driven by the same GCM run reproduce the year-to-year variability of the driving GCM and are thus highly correlated, although inferred climate changes may considerably depend on the selected model. In order to analyze all RCMs one would need to modify the assumptions of a Bayes model such that the correlations are taken into account.

Note that 3 of the 5 simulations include the same sea-surface temperature and sea-ice distributions (i.e. Arpege, CHRM and CLM) stemming from a coupled HadCM3 simulation (for details see Rowell 2005). The HIRHAM simulation considered employs an independent HadCM3/

Table 1 PRUDENCE data overview: we use a subset of models that are driven by different atmospheric GCM runs

Institute	Model			
	RCM	Reference	GCM	Reference
CNRM (Toulouse, France)			Arpege	(Gibelin and Déqué 2003)
ETH (Zurich, Switzerland)	CHRM	(Vidale et al. 2003)	ECHAM5	(Roeckner et al. 2003)
GKSS (Geesthacht, Germany)	CLM	(Steppeler et al. 2003)	HadAM3H/1	(Jones et al. 2001)
DMI (Copenhagen, Denmark)	HIRHAM	(Christensen et al. 1996)	HadAM3H/2	(Jones et al. 2001)
SMHI (Norrköping, Sweden)	RCAO	(Jones et al. 2004)	ECHAM4/OPYC	(Roeckner et al. 1996)

The two HadAM3H simulations are independent atmospheric realizations of the same climate state

HadAM3 ensemble member, and the RCAO another ocean model (see Räisänen et al. 2004). In addition RCAO is interactively coupled with a regional ocean model of the Baltic Sea.

The integration area of the models varies, but in all cases covers the larger part of Europe. The focus is on the Alpine region (AL: 44–48N, 5–15E) which is one of the standard regions of the PRUDENCE project (Christensen and Christensen 2007). This region lies in the center part of the integration area for all models. The spatial resolution of the data is around 0.5° (∼56 km). Model output has been interpolated on the regular CRU grid (see Sect. 2.2) so that it can easily be compared with observations from the control period.

2.2 Observational data

The observed temperature data are obtained from the Climatic Research Unit (CRU). The data is located on a regular 0.5 lon × 0.5 lat grid. It is based on station data, interpolated as a function of latitude, longitude and elevation above sea level. In New et al. (1999) there is a detailed description of the data set and the thin-plate spline that was used for interpolation. Data can be accessed via http://www.cru.uea.ac.uk. It is a widely established surface temperature data set covering the period 1901–2002. In the analysis we assume that the CRU observations represent the true climate.

2.3 Aggregation

For both seasons (winter: DJF and summer: JJA) the statistical analysis is done independently of the other season. We average the variable of interest both temporally over the 3 months of each season and spatially over all land grid points in the Alpine region. For the spatial average, a grid point has been considered as a land point if at least 50% of the corresponding area is landmass. Water grid points have been excluded from all models and the CRU data set to avoid a mixing of the sea and land temperatures.

The spatial domain considered has a size of about 20 × 8 grid points and it is one of the standard domains used for the evaluation of RCMs (see Christensen and Christensen 2007). At the spatial scale considered, both elements of the GCM/RCM model chain are important. Déqué et al. (2007) have used the PRUDENCE archive to quantify whether the regional-scale uncertainties in climate projections stem from the GCM, the RCM or from internal variability. An important conclusion reached from their analysis is that the uncertainty due to the use of different RCMs can be as large as the uncertainty due to different GCMs. More specifically, the analysis showed that uncertainties in winter conditions were primarily affected by the

GCMs (i.e. by large-scale circulations), while summer uncertainties were considerably affected by the RCMs (i.e. by parameterizations).

With this aggregation, one can ignore correlations and trends within a season and within the region. The limitation of spatial averaging is that small-scale features cannot be observed anymore since information is lost. In contrast to Tebaldi et al. (2005), we do not average over the years and retain the interannual variations of the climate which is our main interest. A potential difficulty of our approach is that trends during the periods 1961–1990 and 2071–2100 become confounded with the interannual variability. In order to avoid this, we will include linear trends in our model and integrate them out in the Bayesian framework.

3 Methods

3.1 Notation

As explained in the previous section, the data consists of $T = 30$ observations for the variable of interest in the control period (1961–1990) and of T values of the same variable generated by $M = 5$ models both for the control and scenario periods (2071–2100) under an A2 emission scenario. Having the same number of values in the control and scenario periods is not essential. We denote by $X_{0,t}$ the observations in year $1960 + t$, by $X_{i,t}$ the control output of model i in year $1960 + t$ and by $Y_{i,t}$ the scenario output of model i in year $2070 + t$ with $t = 1,…, T$ years. Although the observations $Y_{0,t}$ for the years $2070 + t$ are not available, they are included as unobserved data in the model. This will make the interpretation of model parameters more transparent. Since separate analyses are conducted for each season, it is not necessary to add an index for the season.

3.2 Bayesian formalism

As mentioned in the introduction, we are going to use a Bayesian approach to construct a probability distribution for the scenario climate given all data. In this approach one has to specify the likelihood $p(\text{Data} \mid \Theta)$, that is the conditional probability density of the data given the parameters Θ (for details see Gelman et al. 2003, Sect. 1.3), and—because all parameters in the model are considered as random variables—a joint prior distribution $p(\Theta)$ of all parameters. In our context "parameters" denote quantities of interest like long-term climate means and variances, climate changes, biases, bias changes or trends that determine the distribution of the data. Other types of parameters that are used within the RCMs are not discussed in the paper. In Sect. 3.3 the likelihood is specified for this

framework and in Sect. 3.5 the distribution of the priors will be discussed.

The foundation for Bayesian methods is the computation of the posterior density $p(\Theta \mid \text{Data})$ of the parameters given the data by Bayes formula:

$$p(\Theta|\text{Data}) \propto p(\text{Data}|\Theta) \times p(\Theta).$$

The posterior density is proportional to the product of the prior and the likelihood. The posterior predictive density of the scenario climate $p(Y_{0,t} \mid \text{Data})$ is of particular interest. This is the best estimate of the distribution of the scenario climate given all data. It is obtained by averaging the density of $Y_{0,t}$ given the parameters Θ with respect to the posterior distribution

$$p(Y_{0,t}|\text{Data}) = \int p(Y_{0,t}|\Theta)p(\Theta|\text{Data})d\Theta.$$

We will also look at the posterior predictive distributions for other variables which are defined similarly.

3.3 Distribution of data

In our framework we make three main assumptions about the conditional distribution of the data given the parameters:

Assumption 1 Conditionally on the parameters, all data are independent.

Assumption 1 implies that the likelihood has a product form. Independence means that serial correlations in the time series and possible correlations between models are ignored. The autocorrelation plots of the series do not show significant correlations, and thus the first part does not seem problematic, though in general this depends upon the region considered. In order to fulfill the second part, different RCMs driven by different GCM simulations are used. Even then, the independence assumption may nevertheless be questioned as, the GCMs and RCMs are based on the same scientific knowledge, and thus they are not completely independent (Tebaldi and Knutti 2007). It means that the PDFs do not represent all sources of uncertainty (e.g. Knutti et al. 2002).

Assumption 2 The distribution of the control climate is

$$X_{0,t} \sim \mathcal{N}(\mu + \gamma(t - T_0), \sigma^2), \tag{1}$$

$$X_{i,t} \sim \mathcal{N}(\mu + \beta_i + \gamma(t - T_0), \sigma^2 b_i^2) \tag{2}$$

with $T_0 = \frac{T+1}{2}$. Centering the time around T_0 yields that the intercept μ can be interpreted as the mean value of the climate distribution. γ is a common linear trend that is estimated from all control simulations and the CRU data set together. This trend is not of main interest, but it should be

removed to obtain stationary distributions. By introducing detrended data $X_{i,t}^{\text{det}} = X_{i,t} - \gamma(t - T_0)$, independent and identically distributed (i.i.d.) data are obtained for the control climate and the outputs of each model:

$$X_{0,t}^{\text{det}} \overset{\text{i.i.d.}}{\sim} F_0^c = \mathcal{N}(\mu, \sigma^2), \tag{3}$$

$$X_{i,t}^{\text{det}} \overset{\text{i.i.d.}}{\sim} F_i^c = \mathcal{N}(\mu + \beta_i, \sigma^2 b_i^2). \tag{4}$$

We denote distributions that describe the control climate by a superscript c. On the other hand we use a superscript s for the scenario period. The parameters μ and σ are the expectation value and standard deviation of the control climate, β_i is an additive bias of the climate mean in model i, and b_i is a multiplicative bias. In other words, we assume that model projections only imply a change in the location and spread, but not of the shape of the distribution.

Independence and identical distributions imply in particular that the detrended data are exchangeable over time, that is, their distribution is independent of permutations of the year index. In other words, a model output $X_{i,t}^{\text{det}}$ is not supposed to be close to the observation $X_{0,t}^{\text{det}}$ for the same year t, and two model outputs $X_{i,t}^{\text{det}}$ and $X_{j,t}^{\text{det}}$ for $i \neq j$ need not be close for the same t. This reflects the fact that the different data series stem from independent realizations of the (same) climate state. However, if model i is good, then the distribution F_i^c of $X_{i,t}^{\text{det}}$ should be close to the distribution F_0^c.

Assumption 3a The distribution of the scenario climate is

$$Y_{0,t} \sim \mathcal{N}(\mu + \Delta\mu + (\gamma + \Delta\gamma)(t - T_0), \sigma^2 q^2),$$

$$Y_{i,t} \sim \mathcal{N}\left(\mu + \Delta\mu + \beta_i + \Delta\beta_i + (\gamma + \Delta\gamma)(t - T_0), \sigma^2 q^2 b_i^2 q_{b_i}^2\right),$$

or equivalently

$$Y_{0,t}^{\text{det}} \overset{\text{i.i.d.}}{\sim} F_0^s = \mathcal{N}(\mu + \Delta\mu, \sigma^2 q^2), \tag{5}$$

$$Y_{i,t}^{\text{det}} \overset{\text{i.i.d.}}{\sim} F_i^s = \mathcal{N}(\mu + \Delta\mu + \beta_i + \Delta\beta_i, \sigma^2 q^2 b_i^2 q_{b_i}^2). \tag{6a}$$

This means that a mean shift $\Delta\mu$ and a multiplicative change q in the variability of the scenario climate are allowed. $\Delta\gamma$ represents a change in the trend for the scenario data. Moreover, with the parameters $\Delta\beta_i$ and q_{b_i} the additive and multiplicative biases can change between the control and scenario periods. A model may reproduce the climate well today, but an increased bias in the scenario is possible due to incorrectly parameterized or simplified physical processes. Note that the components "true change", "bias" and "bias change" are combined additively for the mean, and multiplicatively for the standard deviation.

Combining Assumptions (1) to (3a), the likelihood function is

$$
\prod_{t=1}^{T} \frac{1}{\sigma} \exp\left(-\frac{\left(X_{0,t} - \mu - \gamma(t - T_0)\right)^2}{2\sigma^2} \right) \times
$$
$$
\times \prod_{t=1}^{T} \prod_{i=1}^{M} \frac{1}{\sigma b_i} \exp\left(-\frac{\left(X_{i,t} - \mu - \beta_i - \gamma(t - T_0)\right)^2}{2\sigma^2 b_i^2} \right) \times
$$
$$
\times \prod_{t=1}^{T} \prod_{i=1}^{M} \frac{1}{\sigma b_i q q_{b_i}} \times
$$
$$
\times \exp\left(-\frac{\left(Y_{i,t} - \mu - \Delta\mu - \beta_i - \Delta\beta_i - (\gamma + \Delta\gamma)(t - t_0)\right)^2}{2\sigma^2 b_i^2 q^2 q_{b_i}^2} \right)
$$
$$
\tag{7}
$$

up to a constant which is irrelevant.

The assumption of normal distributions is reasonable due to the aggregation over a season and within the Alpine region. In addition, quantile plots of observations and model data against the theoretical normal distribution do not show strong discrepancies (see Sect. 4.2). In principle the normal assumption can be relaxed using either more general distribution families or a non-parametric approach. But even with the restriction to the normal distribution the problem is still somewhat ill-posed as we will see in the next section.

3.4 Identifiability

For the control climate, there are model values from the RCM control runs and observations from the CRU data set. Therefore it is possible to estimate both the mean value μ of the climate and the individual biases β_i for each model.

Since there are no observations $Y_{0,t}$, $\Delta\mu$ and $\Delta\beta_i$ cannot be estimated separately from the data alone, they are confounded. The model is not identifiable, that is two different parameter sets with identical sums $\Delta\mu + \Delta\beta_i$ lead to the same distribution for all data. A large value of $\Delta\mu$ could in principle be compensated by opposite model bias changes $\Delta\beta_i$ for each model. This is a general problem in statistical and dynamic down-scaling. One needs observations to calibrate and validate a model and to verify model assumptions. These observations are only available for the control climate. Therefore one has to accept the assumption that a (statistical) relationship also holds for the scenario climate, or that parameters calibrated in the control period remain valid in the scenario period. In our context we are facing the same problem by trying to separate the climate change $\Delta\mu$ and the change of the model bias $\Delta\beta_i$ of the i-th model.

There are different ways to handle the identifiability problem:

(i)　One assumes that the model bias does not change, that is $\Delta\beta_i = 0$.

(ii)　One puts restrictions on the bias change, e.g. $\sum_i \Delta\beta_i = 0$, that is the average of the model biases does not change in the scenario period.

(iii)　One introduces a soft restriction that $\sum_i \Delta\beta_i^2$ is small, that is the changes of model biases cannot be too large, where "not too large" will be defined more thoroughly later.

(iv)　One reparameterizes the model by defining new parameters $v_i := \Delta\mu + \Delta\beta_i$ which then are identifiable.

The first alternative seems to be too restrictive, especially if an RCM is calibrated and the model bias is estimated in one region with today's climate. If there is a climate shift, it is possible that the model has another bias for the new climate.

With the second alternative, a large bias change of one model forces either a large bias change of another model in the opposite direction, or many smaller compensations by the other models. In addition it does not allow the total bias to become larger (or smaller) due to a climate shift.

Although the re-parameterization in the fourth alternative solves the identifiability problem, it does not allow one to distinguish between model biases and climate change. Since the aim is a climate projection that corrects for individual model biases, this is not a real alternative to the problem.

The third solution is a regularisation of the over-parameterized problem. In a Bayesian context it can be implemented with specific choices of the priors for the affected parameters $\Delta\beta_i$. Equation 6a together with alternative (iii) will subsequently be referred to as the "constant bias" assumption and later be contrasted with an alternative "constant relation" assumption. The term "constant bias" is somehow misleading since actually bias changes are allowed, but alternative (iii) will overall tend to minimize the bias changes depending upon the prior distribution. In the next sections we will describe these assumptions and their interpretation in more detail.

The same problem as for $\Delta\mu$ and $\Delta\beta_i$ appears for q and q_{b_i}. Because these parameters represent multiplicative biases, only the products $q \cdot q_{b_i}$ are identifiable. Again this problem is solved by forcing the sum of the $\log(q_{b_i})^2$-terms to be small. This regularisation is achieved by the choice of the prior distribution of q_{b_i}.

3.5 Choice of priors

For all parameters one has to choose prior distributions. We assume that all parameters are a priori independent so that only the marginal prior distributions are needed. There are two classes of parameters: μ, $\Delta\mu$, β_i, $\Delta\beta_i$, γ and $\Delta\gamma$ are related to the mean values of the assumed normal distributions of the data. It is common to take normal priors for

these parameters since this simplifies the computations. The other class of parameters consists of σ^2, q^2, b_i^2 and $q_{b_i}^2$ which are variances or multiplicative changes of the variances. It is a common procedure (Gelman et al. 2003) to work with the precision, which is defined as the inverse of the variance, and to choose a Gamma distribution for the prior of the precision. The same procedure is used for the multiplicative change factors. Note that this reparametrization in terms of precision does not affect the results, it is used only for computational reasons. Hence we will later show the posterior for the standard deviation σ and the scale factors q, b_i and q_{b_i} which have a more direct physical interpretation.

Both the normal and the Gamma prior distributions have again parameters—called hyper-parameters—that must be specified. Table 2 presents the adopted values of the hyper-parameters. For the parameters μ, $\Delta\mu$, β_i, γ, $\Delta\gamma$, σ^{-2} and q^{-2} and b_i^{-2}, we choose them so that the priors are flat and thus carry little information. In particular, the prior variances are chosen such that only values which are far away from physical plausibility are excluded. This means the posterior distribution will be mainly determined by the likelihood, that is the data. The reason for this is that in this case little is to be gained by using expert knowledge and that we want to avoid controversies.

The situation for the parameters $\Delta\beta_i$ and $q_{b_i}^{-2}$ is different. For the reasons discussed in Sect. 3.4, we take informative priors with small variances that are concentrated around zero and one, respectively. This choice of hyper-parameters means for instance that the bias change $\Delta\beta_i$ lies between $-1.4°C$ and $1.4°C$ with a probability of 95%. Although this assumption seems somewhat restrictive, one

has to keep in mind that there are no future observation to strictly separate climate shift and bias change. Therefore one is forced to accept an assumption about a possible bias change. Our approach is reasonable. It assumes a priori that the bias change $\Delta\beta_i$ is comparable or smaller than typical biases β_i in the control period, because otherwise the scenario runs would be of little use. Since one can estimate the biases β_i from the data $X_{0,t}$ and $X_{i,t}$, there is a rational basis to choose the variance of the prior for $\Delta\beta_i$.

Only the parameters $v_i = \Delta\mu + \Delta\beta_i$ (climate shift plus additional scenario bias of model i) are identifiable. The prior assumptions above imply that $(v_1, ..., v_M)$ are a priori jointly normally distributed, where all v_i have mean zero and variance $\sigma_{\Delta\mu}^2 + \sigma_{\Delta\beta}^2$ and all pairs v_i, v_j ($i \neq j$) have a correlation $\sigma_{\Delta\mu}^2 (\sigma_{\Delta\mu}^2 + \sigma_{\Delta b}^2)^{-1}$. In other words, the correlation matrix has constant off-diagonal entries. Hence a small $\sigma_{\Delta\beta}^2$ corresponds to the a priori belief that all v_i are similar (highly correlated).

It is important to check the sensitivity of the results to the choice of the prior distributions and the hyper-parameters. This is especially important here, since the hyper-parameters are specified in order to solve the identifiability problem, and are not based on prior expert knowledge. This sensitivity analysis will be done in Sect. 4.4, and we will describe separately how the hyper-parameters have been varied.

3.6 Computation of the posterior

By Bayes formula the joint posterior density of all parameters given the data is proportional to the prior density multiplied by the likelihood of the data.

Table 2 Hyper-parameters for the prior distributions: for normal distributions hyper-parameters for the expectation (μ_0) and the variance (σ_0^2) are given

Parameter	Distribution	Hyper-parameter 1 (μ_0, shape)	Hyper-parameter 2 (σ_0^2, rate)	95% Confidence interval
μ (°C)	Normal	0 (Winter)	25	[−9.8, 9.8]
		15 (Summer)		[5.2, 24.8]
$\Delta\mu$ (°C)	Normal	0	16	[−7.8, 7.8]
β_i (°C)	Normal	0	16	[−7.8, 7.8]
$\Delta\beta_i$ (°C)	Normal	0	0.5	[−1.4, 1.4]
γ (°C year^{-1})	Normal	0	16	[−7.8, 7.8]
$\Delta\gamma$ (°C year^{-1})	Normal	0	16	[−7.8, 7.8]
σ^{-2} (°C^{-2})	Gamma	0.1	0.1	[0, 9.8]
q^{-2}	Gamma	0.1	0.1	[0, 9.8]
b_i^{-2}	Gamma	0.1	0.1	[0, 9.8]
$q_{b_i}^{-2}$	Gamma	3	3	[0.2, 2.4]

For Gamma distributions the first hyper-parameter is the shape parameter and the second hyper-parameter is the rate. The mean is shape $\times$ rate^{-1}, the variance is shape $\times$ rate^{-2}. The 95% confidence intervals show the range of values that are a priori decided to be physically plausible. The intervals are quite large for the non informative priors. For the two parameters $\Delta\beta_i$ and $q_{b_i}^{-2}$ we have chosen informative priors to solve the identifiability problem. Therefore their confidence intervals are smaller

Hence in principle, the posterior is known, but this is of little practical use. In order to deduce information about the marginal posteriors of the two main parameters of interest, $\Delta\mu$ and q, and in order to compute posterior predictive densities, high dimensional integration would be needed which is difficult. Common practice in modern statistics is to rely on Markov Chain Monte Carlo methods instead. Monte Carlo methods replace analytical calculations by empirical estimates computed with an artificially generated sample from the posterior distribution. For complicated high-dimensional distributions it is not feasible to generate an independent sample, but it is possible to generate a dependent sample with a suitable Markov Chain. This means that each member of the sample is constructed recursively from its predecessor, (see e.g. Gilks et al. 1996). For our analysis, we use the standard Gibbs sampler which updates a single component at a time, because the so-called full conditionals have a standard form. Results are based on a single Markov Chain with length 550,000 where the first 50,000 are disregarded as a burn-in period. The remaining 500,000 samples were thinned to a sample of 5,000 by taking only every hundredth point. The thinning removes the dependency within the Markov Chain so that 5,000 remaining points are an independent sample of the distribution of interest. To check the convergence of the chain, diagnostics such as autocorrelation and effective sample size were calculated. None of these diagnostic tools showed any indication that the chain has not converged. Moreover, additional simulations not shown here confirmed the results.

3.7 An alternative assumption for scenario period values

Even under the assumption that climate change and model error affect only location and scale, but not the shape of the distribution, there is at least one additional way to specify the distribution of scenario period values that can also be regarded as plausible. The "constant bias" assumption in Eq. 6a means that the difference between the expected values of the control and the scenario periods in model i is equal to $\Delta\mu + \Delta\beta_i$. Hence up to small bias changes (alternative (iii) in Sect. 3.4), all models are assumed to predict the climate scenario shift correctly.

The alternative "constant relation" assumption says that a model over- or underestimates the climate scenario shift by approximately the same factor by which it over- or underestimates the interannual variability within a season in the control period. The latter factor is equal to b_i. Allowing such an additional bias change means thus replacing Eq. 6a by

Assumption 3b   For $1 \leq i \leq M$

$$F_i^s = \mathcal{N}(\mu + b_i\Delta\mu + \beta_i + \Delta\beta_i, \sigma^2 q^2 b_i^2 q_{b_i}^2). \tag{6b}$$

The specification of the priors of the parameters with this alternative "constant relation" assumption is done as before. In particular, an informative prior is used, forcing $\Delta\beta_i$ to be near zero and q_{b_i} near one (alternative (iii) in Sect. 3.4). In this way, we will avoid the analogue basic non-identifiability problem as discussed in Sect. 3.4.

The two assumptions are shown in Fig. 1. In the left figure the "constant bias" assumption (Assumption 3a) is explained. On the x-axis the observed detrended quantiles are drawn. These are the ordered 30 observations of the yearly climatology for the period 1961–90 (red dots on the x-axis with mean μ) after subtracting the estimated trend. On the y-axis there are the quantiles for an RCM which correspond to the ordered detrended output values for the control period of the RCM. The red points in the plot show a quantile–quantile-plot of the observations against the model output for the control period. The red dashed line show the relationship between these quantiles and therefore the additive bias is the intercept and the multiplicative variability bias the slope of the line. Under the "constant bias" assumption this red line is shifted into the scenario period assuming that the bias remains constant. A small bias change $\Delta\beta_i$ and a multiplicative change of the variability q_{b_i} allow for some changes in the bias. The result is shown with the black solid line. The bias changes are restricted to be small by the informative priors on these parameters. The slightly adapted relationship between the quantiles of today's observation and the control model output is used to estimate the new climate mean $\mu + \Delta\mu$. Since there are no future observation, no points can be drawn for the quantile in the scenario period, on the x-axis.

With the "constant relation" assumption (Assumption 3b) in the right figure, one can extrapolate the observed bias relationship today (red dashed line) into the scenario period. This results in two different parts of the model bias change. The first part is a systematic part. If the slope of the line is larger than one a systematic bias increase of $(b_i - 1)\Delta\mu$ is expected. The second part of the bias change $\Delta\beta_i$ is restricted to be small by the informative priors as with the "constant bias" assumption. One has to remark that with the "constant relation" assumption the bias change can be quite large due to the systematic part since the restriction with informative prior only influences the second part of the bias change. This will result in a different estimation of the climate shift $\Delta\mu$ because a part of the signal is attributed to the bias change. This can be seen in Fig. 1 by remarking that with the same observations and model projections, $\Delta\mu$ in the right figure is smaller than in the left figure. However, because $Y_{0,t}$ is not available, it is
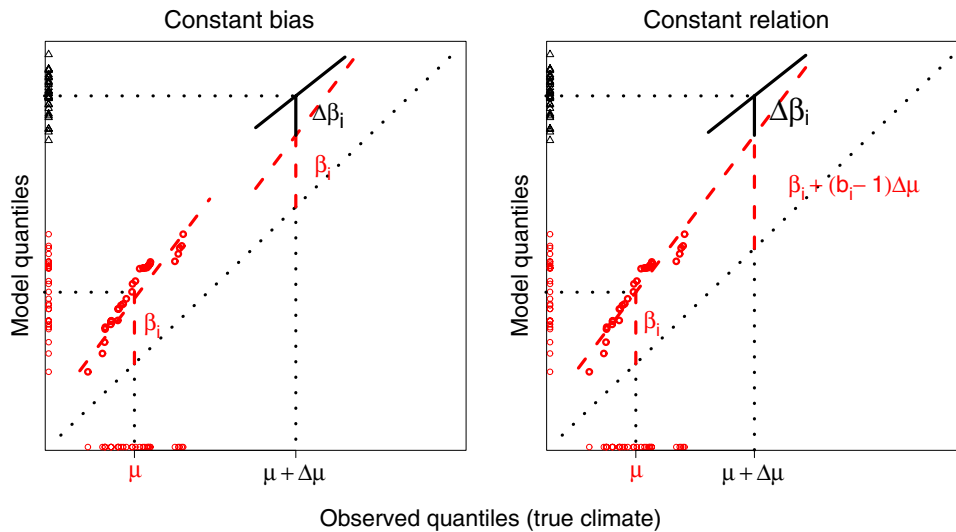
Constant bias

Constant relation

Observed quantiles (true climate)

Fig. 1 Schematic illustration of the two bias assumptions: The *red dashed lines* depict the underlying assumptions. With the "constant bias" assumption it is assumed that the additive bias of the control period (simulated minus observed) also applies to the scenario period. With the "constant relation" assumption it is assumed that the (linear) relationship between simulated and observed quantiles during the control period may be extrapolated into the scenario period. The *black solid line* depicts the resulting relation between quantiles of the scenario climate, accounting for small nonlinear changes of the biases using the Bayes approach. Thus, the *red dashed line* corresponds to the case $\Delta\beta_i = 0$ and $q_{b_i} = 1$ of the full model. The *points on the axes* depict the simulated and observed climates for the control (*red circles*) and scenario periods (*black triangles*), respectively. For the control period a quantile–quantile-plot (*red points in the plot*) is shown. The *black dotted line* is the identity $y = x$

difficult to distinguish between the two models only from the data.

Figure 1 can be justified with formulas for the quantiles of the distributions F^c_i and F^s_i. Remember that the superscript c is used for distributions and quantiles that are related to the control climate while s stands for the scenario climate. The α-quantile $z(\alpha)$ of a distribution is the value that divides the mass of the distribution into the ratio $\alpha:(1-\alpha)$. In other words, the probability that a random draw from this distribution is below $z(\alpha)$ is equal to α. The k-th smallest among T data points is an estimate of the $\alpha = k(T + 1)^{-1}$ quantile. Then by Eqs. 3 and 5

$$z^s_0(\alpha) = \mu + \Delta\mu + q(z^c_0(\alpha) - \mu) \tag{8}$$

and by Eqs. 3 and 4 for $i = 1, \ldots, M$

$$z^c_i(\alpha) = \mu + \beta_i + b_i(z^c_0(\alpha) - \mu). \tag{9}$$

Under Assumption 3a, it holds that

$$\begin{aligned} z^s_i(\alpha) &= \mu + \Delta\mu + \beta_i + \Delta\beta_i + b_i q_{b_i}(z^s_0(\alpha) - \mu - \Delta\mu) \\ &\approx \mu + \Delta\mu + \beta_i + b_i(z^s_0(\alpha) - \mu - \Delta\mu) \end{aligned} \tag{10}$$

since $\Delta\beta_i \approx 0$ and $q_{b_i} \approx 1$. In other words, the relation between the true quantiles and the quantiles of the model output has a similar structure in the control and the scenario periods (if $\Delta\beta_i = 0$ and $q_{b_i} = 1$, the structure is identical).

The 50% quantile $z(0.5)$ is the median (typical year) and in case of the normal distribution, it is equal to the mean. By using $z^c_0(0.5) = \mu$ in Eqs. 8, 9 and 10, one obtains

$$z^c_i(0.5) - z^c_0(0.5) = \beta_i = z^s_i(0.5) - z^s_0(0.5). \tag{11}$$

Hence Assumption 3a says in particular that the difference between model and observation for a typical year are similar both in the control and the scenario period, regardless of how warm a typical year is. This can be justified by saying that the physical relationships are still valid for a changed forcing and thus have about the same error for a typical year.

In contrast, under Assumption 3b, it holds that

$$\begin{aligned} z^s_i(\alpha) &= \mu + \beta_i + \Delta\beta_i + b_i(1 - q_{b_i})\Delta\mu + b_i q_{b_i}(z^s_0(\alpha) - \mu) \\ &\approx \mu + \beta_i + b_i(z^s_0(\alpha) - \mu). \end{aligned} \tag{12}$$

Note that Eqs. 12 and 9 are similar. Assumption 3b postulates therefore that one can use the same linear relation between $z^c_i(\alpha)$ and $z^c_0(\alpha)$ in the control period and between $z^s_i(\alpha)$ and $z^s_0(\alpha)$ in the scenario period. Hence if the temperature of a warm year in the control period is similar to that of a cold year in the scenario, then the difference between model and observations is about the same in both cases. This explains the name "constant relation" that describes Assumption 3b.

It is important to note that both assumptions have been made in distinct areas of climate research. Christensen et al. (2008) suggest that temperature and precipitation biases grow in a global warming scenario. As mentioned in the introduction, the "constant bias" assumption is implicit to the consideration of the "scenario minus control" signal

in climate projections, and is made throughout the IPCC report (Meehl et al. 2007). Likewise, the "constant relation" assumption is made in many statistical evaluations of seasonal forecasting, where a forecasted anomaly is considered relative to the models representation of the observed variability (e.g. Kharin and Zwiers 2003). It can thus be argued that the "constant relation" assumption is the more natural assumption for near-term climate change (e.g. the next 20 years), as we would expect the error structure of the models to be approximately conserved over shorter time periods, when the climate shifts can be considered comparatively small. Likewise, it can be argued that the "constant bias" assumption is the more natural assumption in longer-term climate change studies (e.g. 100 years), as the anticipated changes are considerably larger than the currently observed interannual variability. Further work is needed to determine how biases in the control period can be used for the estimation of biases in the scenario period.

4 Results

4.1 Climate prediction: "Constant bias" versus "constant relation" assumption

4.1.1 Summer temperature

We restrict the discussion to the Alpine region and start with the summer (JJA) season. In the upper row of Fig. 2,

the posterior distributions of $\Delta \mu$, q, γ and $\gamma + \Delta \gamma$ are given under the two assumptions "constant bias" (black solid line) and "constant relation" (red dashed line), respectively. Our method predicts an expected increase of the average temperature of 5.4°C for the "constant bias" assumption and of 3.4°C for the "constant relation" assumption. This difference is quite large and will be discussed in more detail in the next Sect. 4.2.

In contrast, the posterior for the other three parameters is similar under both assumptions. Values above and below 1 are plausible for q, hence the RCMs considered are not able to decide whether the variability of the mean summer temperatures will increase or decrease in the future, albeit there is a small tendency towards an increase in variability, (see also Fig. 3). Previous research revealed that there might be considerable increases in interannual summer variability over Central Europe (Schär et al. 2004). The aforementioned study assessed one single model chain (the CHRM driven by HadAM3H), but recent model intercomparisons indicate that this result qualitatively agrees with most RCMs (Giorgi et al. 2004; Giorgi and Bi 2005; Vidale et al. 2007; Lenderink et al. 2007) and GCMs (Seneviratne et al. 2006). The absence of a pronounced variability increase in our analysis appears mostly related to the consideration of the Alpine region, which is situated to the south of the region of maximum variability increase.

The trend is with posterior probability higher than 99% between −0.01 and 0.04°C per year for the period 1961–1990 and between 0.06 and 0.12°C per year for the period 2071–2100. In comparison, the global mean surface
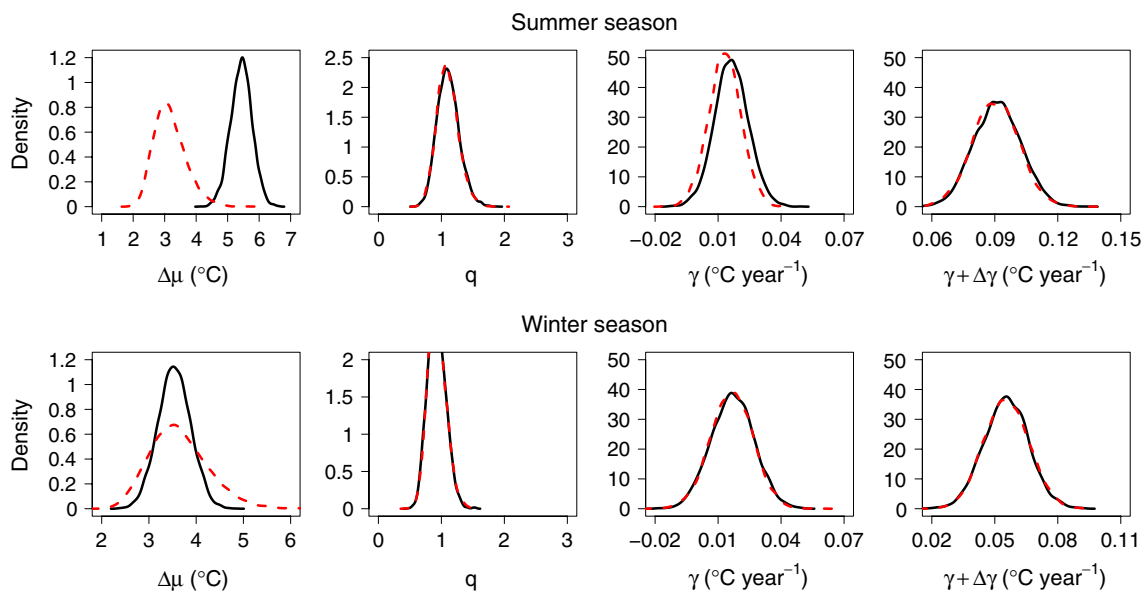

Fig. 2 Posterior densities for the climate shift $\Delta \mu$, the change of variability q, the trends γ for the control and $\gamma + \Delta \gamma$ for the scenario period in summer (*upper row*) and in winter (*lower row*). The *solid black lines* show the densities for the "constant bias" and the *red dashed lines* for the "constant relation" assumption. There is a large difference for the estimated climate change in summer between the two assumptions. Note that for q, γ and $\gamma + \Delta \gamma$, the two curves are lying upon each other
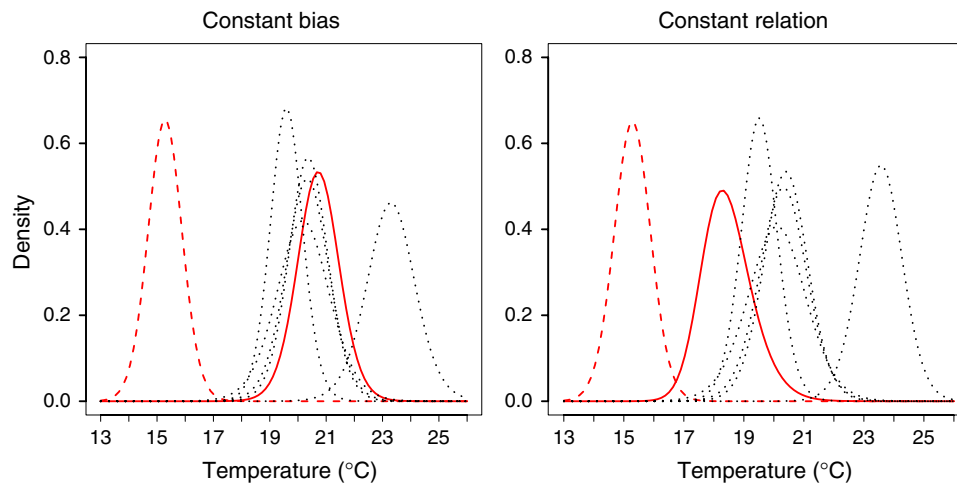
Fig. 3 Posterior predictive densities for mean summer temperature: The *dashed red line* is for the control period (observations), the *solid red line* for the scenario period (multi-model projection) and the dotted black lines for the scenario output of the individual RCMs which are corrected for the control bias. Note that the individual RCM output curves are calculated using the posterior distributions of the

Bayes model and therefore these curves can be different using the two bias assumptions. The main difference is the larger predicted climate mean change of 5.4°C for the "constant bias" assumption (*left*), compared to only 3.4°C climate shift for the "constant relation" assumption (*right*)

temperature trend of the A2 scenario for the 2071–2100 period amounts to 0.05°C per year (Meehl et al. 2007, see their Fig. 10.4). The larger trend over the Alpine region revealed above can be explained by two reasons: First, the regional warming over continental land surfaces considerably exceeds the global mean warming which is moderated by the presence of large ocean surfaces. Second, it is possible that the RCMs overestimate the trend during the scenario period, as the respective simulations use a spin-up period of merely one year and are initialized from a soil-moisture distribution that is not in complete balance with the scenario climate. However as the RCM trend exceeds the global trend by merely a factor 2, we believe that the former reason dominates.

In Fig. 3, the posterior predictive density given all data is shown with a dashed red line for the mean temperatures $X_{0,t}^{\text{det}}$ in the control period and with a red solid line for the predicted mean temperatures $Y_{0,t}^{\text{det}}$ in the scenario period. The posterior predictive densities for the output $Y_{i,t}^{\text{det}}$ of individual RCMs are given with black dotted lines. In addition to the trend, the individual biases β_i and b_i are also removed, but not the bias changes of the scenario period. The additive bias change for the scenario is $\Delta\beta_i$ under the "constant bias" assumption and $(b_i-1)\Delta\mu + \Delta\beta_i$ respectively under the "constant relation" assumption. As we will see in Sect. 4.3, the b_i's are quite large for all models in the summer season. This explains why under the "constant relation" assumption the expected value of the multi-model ensemble projection is smaller than all individual model projections of the scenario period.

Recall that in the posterior predictive density uncertainty about the parameters has been taken into account by integrating with respect to the posterior distribution of the parameters. Hence the individual model projections depend also on other models through integration over the posterior distribution of the parameters given the data. Therefore they influence each other to some extent.

For both assumptions, the range of the different models is quite large. The combined Bayesian prediction density is much narrower than an equally weighted average of the prediction densities of the 5 models. This is due to the inclusion of additive bias changes for the individual RCMs in the model. Note that the biases are not estimated by aligning the black curves as well as possible. For the control climate, they are essentially estimated by comparing the control simulations and the observed climate, and for the scenario climate the estimate depends upon the assumption. For the "constant bias" assumption, they are assumed to be similar because the prior for $\Delta\beta_i$ is concentrated around zero. For the "constant relation" assumption we assume that the biases show a linear relationship where the intercept and slope are determined by comparing the control simulations and the observed climate. The size and uncertainty of estimated biases for both assumptions will be discussed in Sect. 4.3.

4.1.2 Winter temperature

In the lower row of Fig. 2, the posterior distribution of $\Delta\mu$, q, γ and $\gamma + \Delta\gamma$ are given under the two assumptions "constant bias" and "constant relation", respectively. In contrast to the summer, the results are quite consistent under both assumptions and an expected increase of the mean temperature of around 3.5–3.6°C is observed. The

uncertainty about the climate shift $\Delta\mu$ is larger under the "constant relation" assumption.

As in summer, the posterior for the other three parameters is similar under both assumptions. Values above and below 1 are plausible for q, hence the RCMs considered are not able to decide whether the variability of the mean winter temperatures will increase or decrease in the future.

Figure 4 shows the posterior predictive densities given all data for mean winter temperature. The different lines have the same meaning as for summer. The two distributions of the control and scenario climate have a larger overlap than in summer. The individual RCMs have nearly the same variability as the combined Bayes prediction, in distinction to the spread in the summer season. This is likely due to the reduced role of soil moisture during winter.

4.2 Diagnostic check of assumptions

Although our results are reasonable and consistent with the literature, we have to verify several assumptions.

4.2.1 Normal distribution, independence

After the aggregation of daily data to a seasonal mean, and of spatial data to a regional mean, it is not surprising that the distribution of this mean is close to normal due to the central limit theorem for weak dependence. To verify the normal assumption we visually check the normal plots (ordered values against the quantiles of the normal distribution) for each model and the observations. Deviations from the normal assumption would show up as nonlinear relations. Since trends are in the model, the original data are not stationary and one should use the detrended data as introduced in Eqs. 4 and 5 for constructing the normal plot. For the summer control period, there is no obvious violation of normality in Fig. 5. One can see the very cold

winter 1962/63 as an outlier in the observations. Plots for projection in the scenario period look similar. We also checked the normal plot based on combining all data after centering and scaling the values, and there is no systematic deviation from normality. In addition to the quantile plots the assumption of normality can be checked using the Shapiro-Wilks test for normality and a goodness-of-fit test based on the linearity of the probability plot (for details see Rice 1995, chap. 9). In summer and winter, for all models and the observations there was no significant violation of the normal assumption. The smallest p-value was 0.058. Using quantile plots and goodness-of-fit tests there are no obvious violation of the normal assumption.

Furthermore, in order to examine the temporal independence between the different years, we computed the autocorrelation for each model and for the observations, assuming a stationary time series model. Even at lag one, no significant autocorrelation could be observed. There is no strong correlation between the different RCMs either. Such correlations are avoided by not including additional RCMs that are driven by the same GCM run. Since on a large scale the RCM reproduces the year-to-year process of the GCM, such correlations would be quite high. In the PRUDENCE project some RCMs are driven by the same GCM run and have a correlation between 0.8 and 0.95. We currently study a possible extension of our model that incorporates a GCM effect and thereby relaxes the restriction that all model chains consider a different GCM simulation.

4.2.2 Relation between model output and observations

In Fig. 6 are the quantile plots of the control runs and the observed temperatures (plot of ordered values of the two data sets). As before these are not the raw values, but the detrended data. Again, a linear relation is expected if our model assumptions are correct. The multiplicative

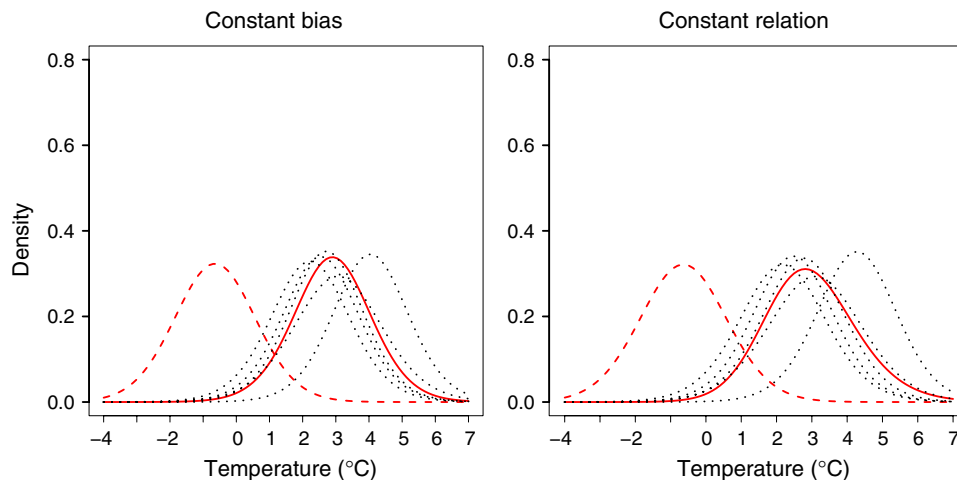**Fig. 4** Same as Fig. 3, but for mean winter temperature

Fig. 5 Normal plots of model outputs and observations for the control period (ordered, detrended mean summer temperatures against the quantiles of the normal distribution). The *red triangles* (right y-axis) are winter, the *black circles* (left y-axis) summer temperatures
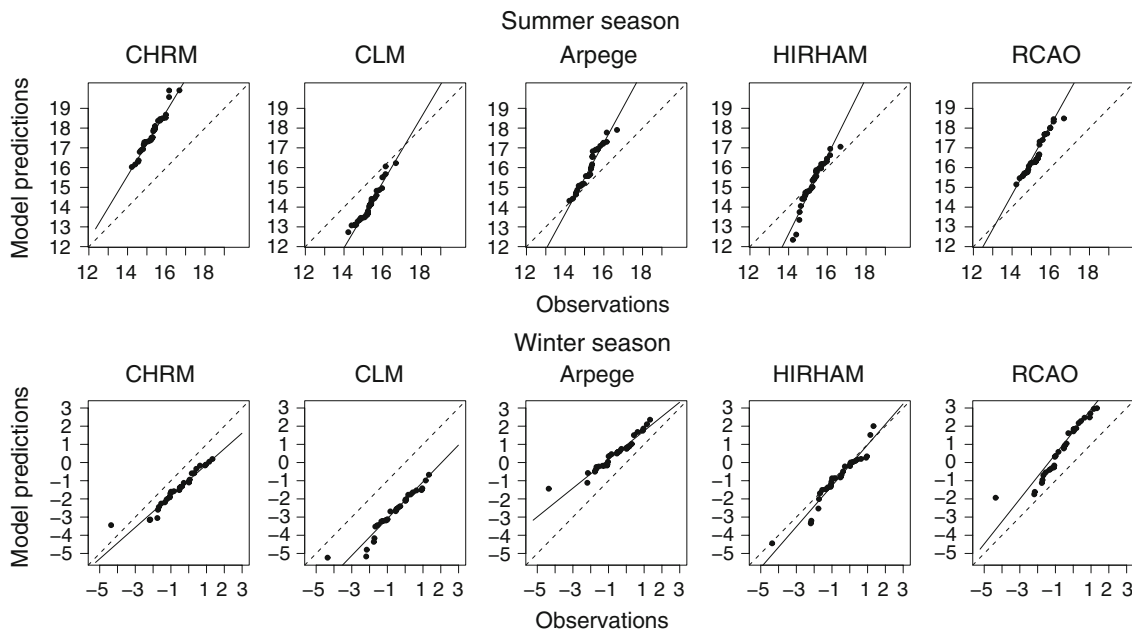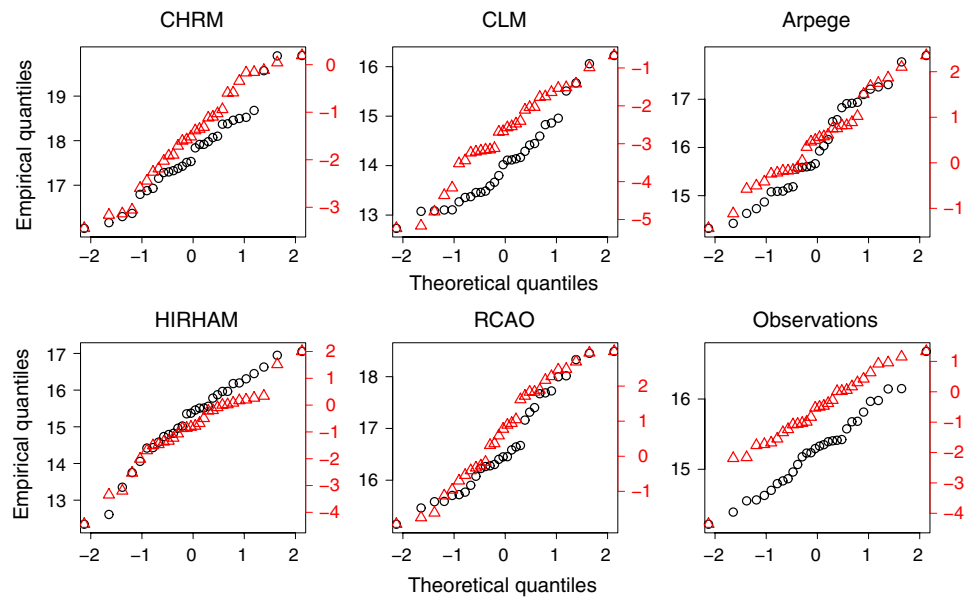
Fig. 6 Quantile plots for mean temperatures in the control period (ordered and detrended model outputs against ordered, observed and detrended temperatures)

variability bias can be seen as a change of the slope. In the summer season, obviously all models have a slope larger than one, that is they overestimate the variability of summer mean temperatures, as already noted in other publications (Vidale et al. 2007; Lenderink et al. 2007). In winter there are no systematic variability biases. Note that for the winter season the eye-catching observation in all quantile plots is the winter 1962/1963 that was extraordinary cold.

These results have different implications under the "constant bias" and the "constant relation" assumption, as

we have seen in Sect. 4.1. In the next Sect. 4.3 we will examine the biases and bias changes in more detail to explain the behaviour of the two assumptions.

4.3 Model biases

4.3.1 Summer temperature

In Fig. 7 the posterior densities of the additive biases for control and scenario summer temperatures are shown. Upper and lower row display the biases for the "constant
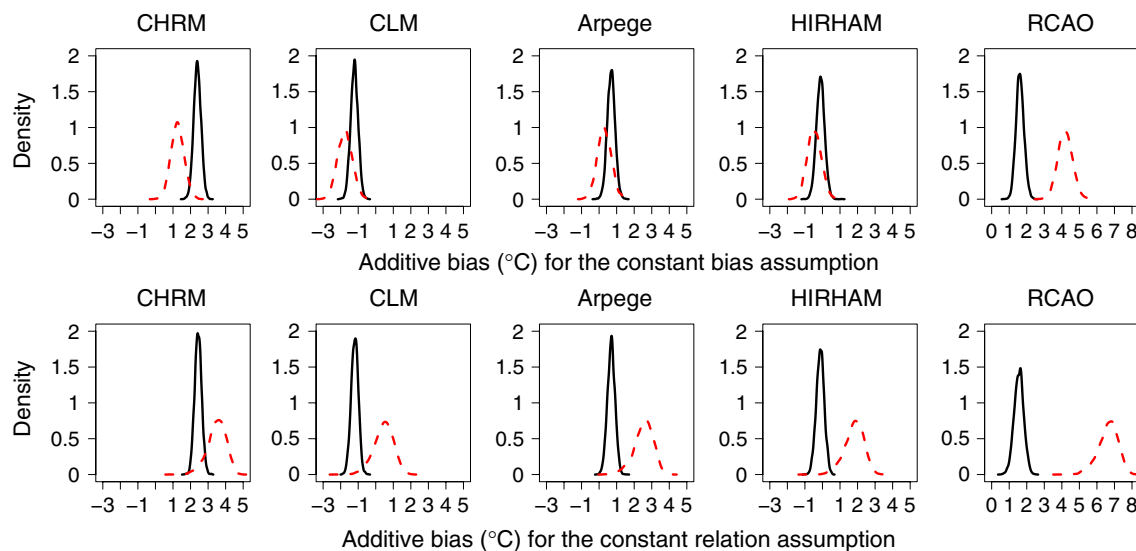
Fig. 7 Posterior densities for additive model biases in the summer season for the different RCMs. The *solid black lines* are for the control biases (β_i) and the *dashed red lines* for the scenario biases ($\beta_i + \Delta\beta_i$) in the "constant bias" assumption, and ($\beta_i + (b_i - 1)\Delta\mu + \Delta\beta_i$) in the "constant relation" assumption, respectively. Note that the scale in the RCAO panel is different from the scales of the other RCMs

bias" and "constant relation" assumption, respectively. The solid black line represents the control bias β_i and the dashed red line the scenario bias. Under the "constant bias" assumption, the scenario bias is $\beta_i + \Delta\beta_i$ whereas for the "constant relation" assumption it is $\beta_i + (b_i - 1)\Delta\mu + \Delta\beta_i$.

With the "constant bias" assumption the biases for control and scenario periods are generally similar, but the uncertainty about the biases increases in the scenario period. This was to be expected. There is no systematic increase or decrease of the biases in all models. The RCAO model has the biggest bias change. The situation changes for the "constant relation" assumption. The biases for the control period are similar to those under the "constant bias" assumption, but there is a systematic increase in the scenario biases of all models. In all models, the scenario bias is now clearly positive and the uncertainty is larger than under the "constant bias" assumption. Again it is largest for the RCAO model. Since under the "constant relation" assumption all RCMs have large positive biases, the climate shift remaining after bias correction is smaller than under the "constant bias" assumption. This results in a posterior predictive density $Y_{0,t}^{\text{det}}$ that is smaller than each RCM as observed in Fig. 3.

The difference between the scenario biases under the two assumptions is equal to $(b_i-1)\Delta\mu$. A simple point estimate of b_i is given by the slope of straight lines in Fig. 6 which are clearly greater than one. This is confirmed by Fig. 8 which shows the posterior distributions of multiplicative variability biases in summer. The solid line represents the control bias b_i and the dashed line describes

the scenario bias $b_i q_{b_i}$. Under both assumptions, the control bias b_i is larger than one for all models, and this explains why the scenario biases are substantially larger under the "constant relation" assumption. In other words, the reason for the difference between the results under the two assumptions is the overestimation of the year-to-year variability in the summer by most models (see Vidale et al. 2007; Lenderink et al. 2007). Figure 8 also shows that the scenario multiplicative bias is—under both assumptions—not much different from the control multiplicative bias. The difference is largest for the CLM model.

The ability to estimate biases of individual models both for the control and scenario period is a clear advantage of our approach. Assuming that there are no biases or that the biases remain constant over time would lead to incorrect quantifications of uncertainty.

4.3.2 Winter temperature

In Fig. 9 the posterior densities of the additive biases of the control and scenario periods are shown for the winter season. There is no systematic behaviour, neither for the control bias nor for the bias change. Some models overestimate, some underestimate the true climate shift. This holds for both the "constant bias" and the "constant relation" assumption. In Fig. 6 one can see that in the winter season there is no systematic under- or overestimation of the variability. Therefore, in the quantile plots of the models, one observes slopes which are larger and slopes that are smaller than one. In such a situation a linear extrapolation of the biases does not show a common
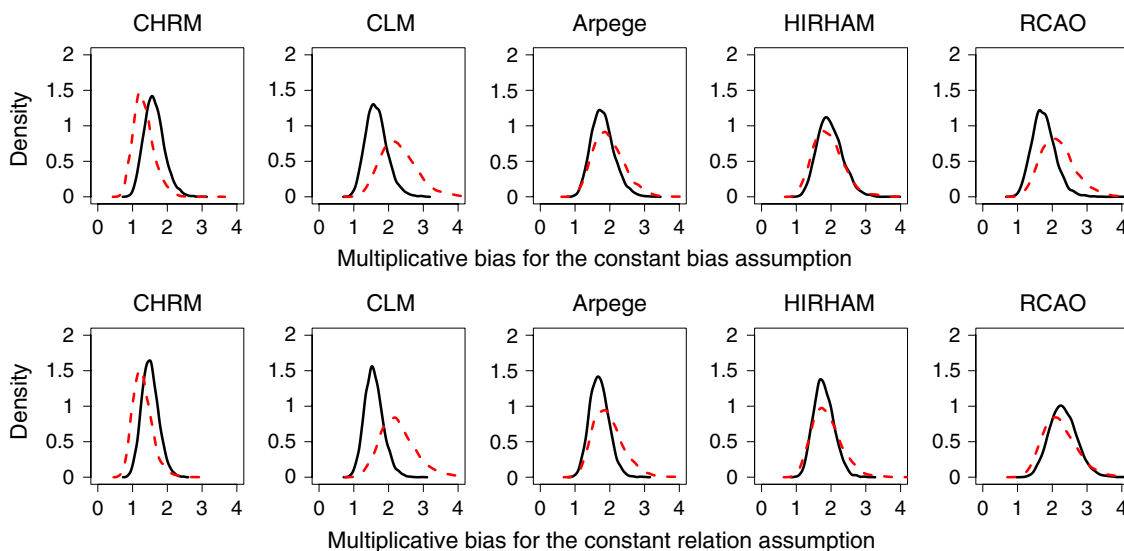
Fig. 8 Posterior densities for multiplicative summer variability biases of the different RCMs. The *solid black lines* are for the control biases (b_i) and the *dashed red line* for the scenario biases ($b_i \cdot q_{b_i}$)
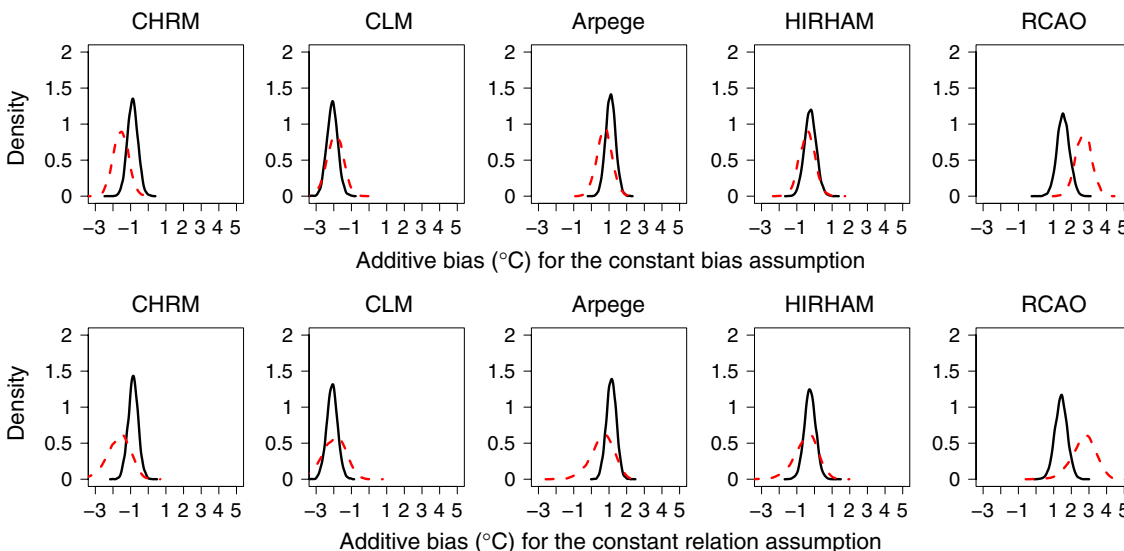

Fig. 9 Same as Fig. 7, but for winter

change for all models and the "constant bias" and the "constant relation" assumption give similar results.

In all models the uncertainty about the additive bias increases from control to scenario period. For some models the posterior mean of the bias remains unchanged and only the spread is larger. For other models the posterior mean also changes. Compared to the biases of the summer, the biases are slightly smaller in winter. Again, the RCAO model yields the largest bias and the largest bias change, but they are smaller than in the summer for the same model.

In Fig. 10 the multiplicative variability biases for the winter temperature are shown. The first obvious point is

that the uncertainty of the estimates of the multiplicative model biases in the winter season is smaller, the distributions are more concentrated around one. Overall, estimating mean winter temperature seems to be easier than summer temperature.

4.4 Sensitivity analysis

In Sect. 3.4 we described an identifiability problem of our model setup. Our solution in the Bayesian framework has been to choose informative priors for the two parameters $\Delta\beta_i$ and q_{b_i}. We used a normal distribution with expectation 0°C and variance 0.5°C^2 for $\Delta\beta_i$ and an inverse
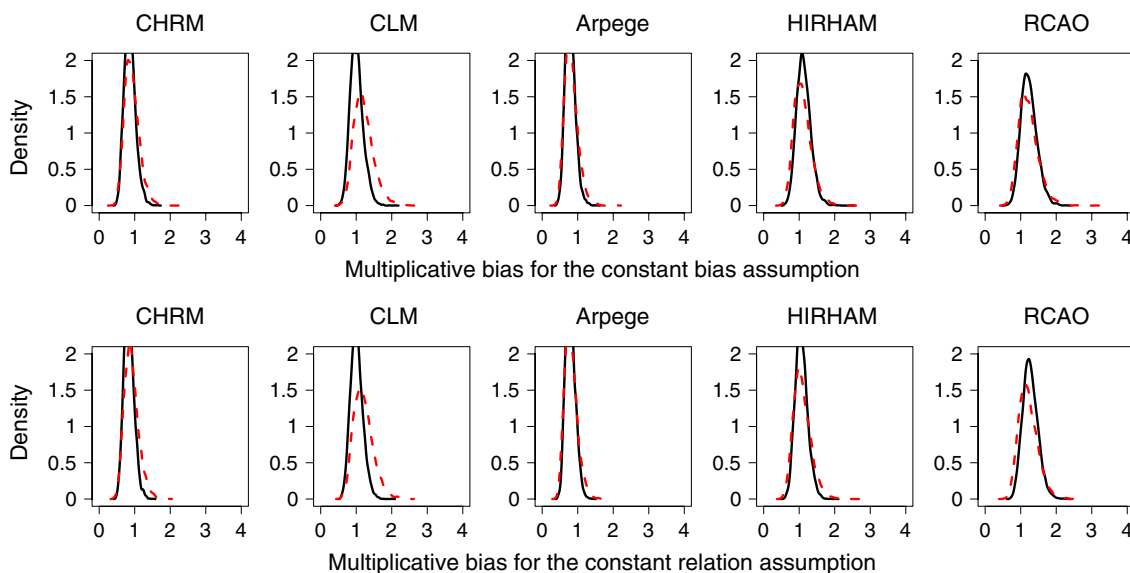
Fig. 10 Same as Fig. 8, but for winter

Gamma distribution with expectation 1 and variance 0.33 for q_{b_i} (see Table 2). Although these choices are based on some qualitative knowledge about the behaviour of the model biases, there is an additional uncertainty in this prior distribution that is difficult to quantify. We therefore vary the variances of these prior distributions over a large spectrum of possible values to examine the sensitivity of our model to the prior distributions. It would be desirable to vary several of the hyper-parameters simultaneously since there are also possible interactions between the parameters, but this is computationally expensive to do. As a compromise between varying only the hyper-parameters of one single parameter and varying all hyper-parameters together we simultaneously varied the hyper-parameters of two parameters and kept the others fixed. This has been done for all possible pairs of parameters.

There is no interaction between most parameters as long as extreme situations are avoided. There is an interaction between parameters which cannot be separated due to the identifiability problem, e.g. $\sigma^2_{\Delta\beta}$ and $\sigma^2_{\Delta\mu}$. In Fig. 11 we show results of different prior distributions for one single parameter $\Delta\beta_i$, the additive bias change. For $\Delta\beta_i$ we varied the hyper-parameter $\sigma^2_{\Delta\beta}$ of the prior distribution. Plots of the effect on the posterior for the additive bias change $\Delta\beta_i$ of the CHRM model and the corresponding climate shift $\Delta\mu$ are shown for the "constant bias" assumption, but the plots for the "constant relation" assumption look similar.

In the upper row of Fig. 11 the dashed red lines show the prior distributions and the solid black lines the a posteriori distributions of $\Delta\beta_i$. Different values of the prior variance $\sigma^2_{\Delta\beta}$ are used. For large values one can see the identifiability problem. There is a lot of uncertainty and the gain of knowledge by the observations is small. For small

values the prior and the posterior distributions are nearly identical. In such cases we assume that there is essentially no bias change and therefore the identifiability problem disappears. These different prior distributions for the bias change affect not only the posterior of the bias change, but also the posterior of the parameter $\Delta\mu$ that describes the climate shift. In the lower row of Fig. 11, it is shown how the posterior of $\Delta\mu$ (solid black line) changes by varying the prior variance $\sigma^2_{\Delta\beta}$. Note that the prior distribution of $\Delta\mu$ is fixed (dashed red line) and only the prior distribution of $\Delta\beta_i$ is changed. Furthermore if there is an uninformative prior for $\Delta\beta_i$, the correlation between $\Delta\mu$ and $\Delta\beta_i$ gets higher as expected. Nevertheless if one only considers the sum $\Delta\mu + \Delta\beta_i$ as proposed in Sect. 3.4 in alternative (iv), the identifiability problem disappears. For all $\sigma^2_{\Delta\beta}$ in Fig. 11 one obtains the same distribution for this sum (plots not shown). As indicated in Sect. 3.4 this is not a true solution to our problem since the estimation and separation of the climate shift and model bias is our main purpose.

Having a very concentrated prior distribution around 0 for the $\Delta\beta_i$'s means that there is no bias change. In that situation the a posteriori distribution for the climate shift is also very concentrated around 5°C (mean summer temperature increase). With a totally uninformative prior for $\Delta\beta_i$, the uncertainty about the climate shift increases, $\Delta\mu$ laying somewhere between 2 and 7°C. This behaviour of the climate shift has also been observed by Lopez et al. (2006, see their Fig. 3) when they are using different priors for the change of the variability for the scenario runs. Including the year-to-year variability, the uncertainty about the predicted mean summer temperatures would be even larger. But such uninformative prior for $\Delta\beta_i$ with $\sigma^2_{\Delta\beta} = 4$ or 16 is not a reasonable choice in our view. The bias $|\beta_i|$
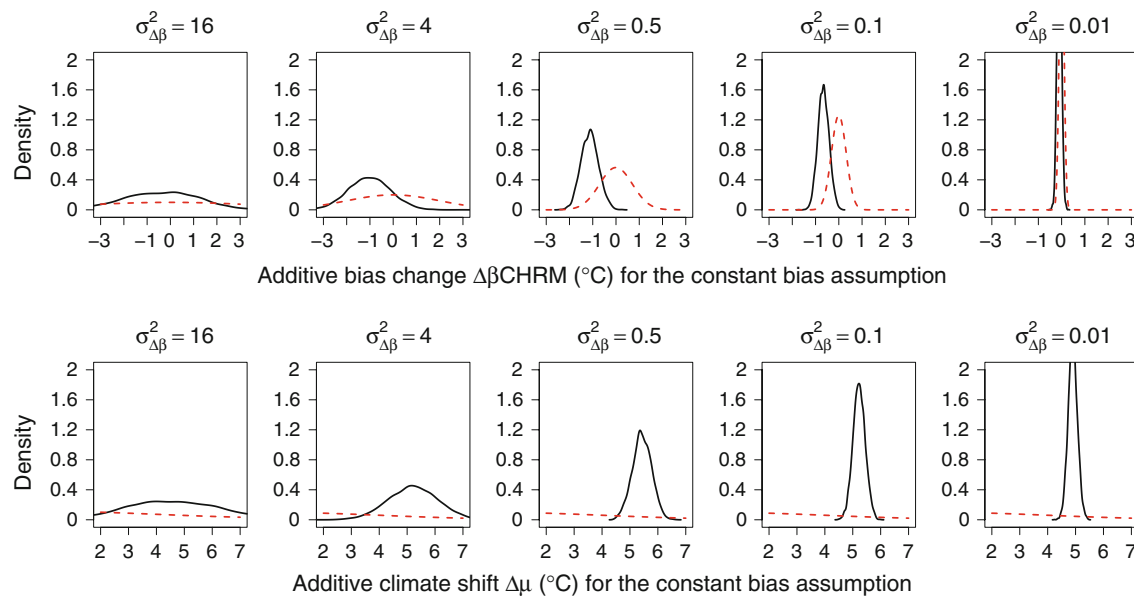
Fig. 11 Sensitivity of the posterior to the prior in case of the additive bias changes $\Delta\beta_i$ in the CHRM model (*upper row*) and for the climate shift $\Delta\mu$ (*lower row*) in summer months: The *dashed red lines* show the prior and the *solid black lines* the posterior distributions. For both rows, the variance $\sigma^2_{\Delta\beta}$ of the prior distributions for $\Delta\beta_i$ varies over a large spectrum of values. The variance for the prior distribution of $\Delta\mu$ is fixed

are with high probability less than $3°$ (see Fig. 7) for all models and therefore one would expect the bias changes $|\Delta\beta_i|$ to be smaller than $3°$ as well.

Having more uncertainty in the scenario runs implies increased uncertainty in the climate shift. The sensitivity shown here is not a disadvantage of the Bayesian approach, but it highlights a more general problem. Making the assumption of a constant bias over time leads to a too confident conclusion about the precision of a prediction. The other extreme is the assumption that there is no knowledge about the change of the bias at all. Then practically no conclusion can be drawn from the model outputs. Hence one should make a reasonable choice of the size of possible bias changes $\Delta\beta_i$. Note that only for additive and multiplicative bias changes, informative priors have been used. To validate this statement we have run a simulation in which all other priors have been taken completely uninformative (improper priors). The results have not changed.

5 Conclusions and outlook

We have developed a new Bayesian methodology for the estimation of future temperature distributions by combining the information contained in a multi-model ensemble and available observations. The new model entails two innovations: First, it has specifically been designed to provide an estimate of the full distribution of a climate variable. It thus allows the consideration of changes in variability and mean, rather than merely changes in mean.

Second additive and multiplicative biases of individual models can be taken into account, and these biases are allowed to vary with time and thus to depend upon the climate state. Although the consideration of time-dependent biases is subordinate to the main objectives of the study, it is not possible to separate the two issues, as assumptions about biases changes under an emission scenario directly influence the outcome of climate change projections.

– The new methodology is successfully applied to temperature changes as simulated by five GCM/RCM model chains, and it yields a single probabilistic estimate of climate change under an SRES A2 scenario. We can consider the predictive density of the resulting temperature changes as a kind of weighted average of shifted and scaled versions of the individual RCM predictions. The Bayesian approach incorporates a statistical way for deriving the weights, shifts and scale factors. We start with equal prior weights and the same priors for shifts and scale factors for all models. In principle, with the Bayesian approach it would also be possible to include qualitative a priori knowledge about different model behaviour in an easy way.

– The methodology does not make any a priori assumptions regarding climate change. In particular, the priors for the parameters describing the climate change signal are non-informative. A more comprehensive sensitivity analysis (not included in the paper) confirms that the choice of these priors does not influence the results.

– Our analysis does however show that there is an intrinsic identifiability problem, as the data does not allow a clear separation between bias changes and climate changes. Some additional assumptions are thus inevitable. We resolved this identifiability problem by using informative priors for the bias changes. The choice of these priors influences the results, but we believe that our choice is reasonable, and we show that the sensitivity is small as long as we avoid extreme choices. Also, the use of informative priors is well justified, as there is both established trust in climate models and justified doubts about the stationarity of model biases. In effect, our approach constrains the bias changes to be smaller in magnitude than the climate changes by about a factor 3.

– The study demonstrates that assumptions about the extrapolation of the model biases from the control into the scenario period are crucial, at least for the situation considered (Alpine summer surface temperatures). To arrive at this conclusion, we have made two different assumptions about the behaviour of the model bias, referred to "constant bias" and "constant relation" assumption. Both assumptions appear plausible and both have (implicitly or explicitly) been used in climate studies, yet the two assumptions yield different estimates of future summer mean temperatures. Indeed, with one of the two assumptions, the strong summer mean warming exhibited by most models is reduced from an ensemble mean of 5.4°C to 3.4°C, thus becoming smaller than the ensemble mean warming for the winter season. By contrast, winter temperature estimates are not affected by the bias assumptions, and this difference is explained by the difficulties (success) of the models in reproducing the observed interannual variability of the summer (winter) season. Although the current paper restricts its attention to Alpine temperatures, we note in passing that similar conclusions can be drawn if the model is applied to larger areas, e.g. Central Europe.

The aforementioned result is of general interest, as it questions an important implicit assumption of current scenario models, namely that the model bias will not significantly depend upon the climate state. This assumption is implicitly buried in the consideration of "changes in climate", which are defined as the difference between scenario and control climate.

Distinguishing in an objective way between the two aforesaid bias assumptions seems difficult. The decision cannot be made by statistical methods alone, but needs expert knowledge. Additional information about the behaviour of model biases may be gained by considering one model in different climatic regions or under different emission scenarios. Longer time series for the control runs and observations may also help to determine the behaviour of the biases and would also enable the consideration and exploitation of different variability measures (e.g. interannual versus decadal variability).

There are several extensions of our methodology beyond the current study. Since spatial and temporal aggregation is a limitation of this study, one could consider spatial averages over smaller regions (e.g. station rather than domain-averaged data), temporal averages over shorter periods (e.g. monthly rather than seasonal means), other variables (e.g. precipitation), or replace the temporal averages by a measure that considers extremes (e.g. number of days above a 90th percentile). Applying the current methodology to other models and data sets (e.g. global mean surface temperature) would also be of considerable interest. Some of these extensions would presumably require us to consider non-normal distributions. Extensions to other location-scale families of distributions (univariate distributions that are parameterized by a location parameter μ and a scale parameter σ) are straightforward, but things become more complicated when different shapes of the distribution are also involved. Other potential extensions deal with the separation of GCM and RCM uncertainties and with an individual treatment of the different RCMs trends. For the former, one would include RCMs that are based on the same GCM simulation and model the correlations with hierarchical random effects. For the latter, one would replace the common slope γ in Assumption 2 by a model-specific slope $\gamma + \delta_i$ for model i. Another question is the treatment of spatial correlations if no aggregation is done. Some of these extensions will be considered in the PhD thesis of the first author.

References

Christensen JH, Christensen OB (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. Clim Change 81:7–30

Christensen JH, Christensen OB, Lopez P, Meijgaard EV, Botzet M (1996) The HIRHAM4 regional atmospheric climate model. Technical report, DMI

Christensen JH, Carter TR, Rummukainen M, Amanatidis G (2007a) Evaluating the performance and utility of regional climate models: the PRUDENCE project. Clim Change 81:1–6

Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held I, Jones R, Kolli RK, Kwon WT, Laprise R, Rueda VM, Mearns L, Menéndez CG, Räisänen J, Rinke A, Sarr A, Whetton P (2007b) Regional climate projections. In: Climate Change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys Res Lett 35:Art. No. L20,709

Déqué M, Rowell D, Lüthi D, Giorgi F, Christensen JH, Rockel B, Jacob D, Kjellström E, de Castro M, van den Hurk B (2007) An intercomparison of regional climate models for Europe: assessing uncertainties in model projections. Clim Dyn 81:53–70

Furrer R, Sain SR, Nychka D, Meehl GA (2007) Multivariate bayesian analysis of atmosphere-ocean general circulation models. Environ Ecol Stat 14:249–266

Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian data analysis, 2nd edn. Texts in Statistical Science, Chapman & Hall/CRC

Gibelin AL, Déqué M (2003) Anthropogenic climate change over the Mediterranean region simulated by a global variable resolution model. Clim Dyn 20:327–339

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov Chain Monte Carlo in practice. Chapman and Hall, London

Giorgi F (1990) Simulation of regional climate using a limited area model nested in a general circulation model. J Clim 3:941–963

Giorgi F, Bi X (2005) Regional changes in surface climate interannual variability for the 21st century from ensembles of global model simulations. Geophys Res Lett 32:Art. No. L13,701

Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. J Clim 15:1141–1158

Giorgi F, Bi XQ, Pal JS (2004) Mean, interannual variability and trends in a regional climate change experiment over Europe. II. Climate change scenarios (2071–2100). Clim Dyn 23:839–858

Jones RG, Murphy JM, Hassell DC, Taylor R (2001) Ensemble mean changes in a simulation of the European climate of 2071–2100 using the new Hadley Centre regional modelling system HadAM3H/HadRM3H. Technical report, Hadley Centre, Exeter, UK; available at <http://prudence.dmi.dk>

Jones CG, Willen U, Ullerstig A, Hansson U (2004) The Rossby Centre Regional Atmospheric Climate Model part I: Model climatology and performance for the present climate over Europe. Ambio 33:199–210

Jun M, Knutti R, Nychka D (2008) Spatial analysis to quantify numerical model bias and dependence: how many climate models are there? J Am Stat Assoc 103:934–947

Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. Clim Change 21:289–302

Kharin VV, Zwiers FWR (2003) Improved seasonal probability forecasts. J Clim 16:1684–1701

Knutti R, Stocker TF, Joos F, Plattner GK (2002) Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature 416:719–723

Lenderink G, van Ulden A, van den Hurk B, van Meijgaard E (2007) Summertime inter-annual temperature variability in an ensemble of regional model simulations: analysis of the surface energy budget. Clim Change 81:233–247

Lopez A, Tebaldi C, New M, Stainforth D, Allen M, Kettleborough JG (2006) Two approaches to quantifying uncertainty in global temperature changes. J Clim 19:4785–4796

Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG, Weaver AJ, Zhao ZC (2007) Global climate projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Min SK, Hense A (2007) Hierarchical evaluation of ipcc ar4 coupled climate models with systematic consideration of model uncertainties. Clim Dyn 29:853–868

Nakicenovic N, Alcamo J, de Vries B, Fenhann J, Gaffin S, Gregory K (2000) Emission scenarios. A special report of working group III of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom

New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. part I: development of a 1961–90 mean monthly terrestrial climatology. J Clim 12:829–856

Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Center climate model: HadAM3. Clim Dyn 16:123–146

Räisänen J, Hansson U, Ullerstig A, Döscher R, Graham LP, Jones C, Meier HEM, Samuelsson P, Willén U (2004) European climate in the late 21st century: regional simulations with two driving global models and two forcing scenarios. Clim Dyn 22:13–31

Rice JA (1995) Mathematical statistics and data analysis, 2nd edn. Duxbury Press, Belmont

Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dmenil L, Esch M, Giorgetta M, Schlese U, Schulzweida U (1996) The atmospheric general circulation model ECHAM-4: model description and simulation of present-day climate. Technical report, MPI

Roeckner E, Bäuml G, Bonaventura L, Brokopf R, Esch M, Giorgetta M, Hagemann S, Kirchner I, Kornblueh L, Manzini E, Rhodin A, Schlese U, Schulzweida U, Tompkins A (2003) The atmospheric general circulation model ECHAM 5. Part I: model description. Technical report, MPI

Rowell DP (2005) A scenario of European climate change for the late twenty-first century: seasonal means and interannual variability. Clim Dyn 25:837–849

Sain S, Furrer R, Cressie N (2008) Combining ensembles of regional climate model output via a multivariate markov random field model. Technical report, Department of Statistics, The Ohio State University

Schär C, Vidale PL, Lüthi D, Frei C, Häberli C, Liniger M, Appenzeller C (2004) The role of increasing temperature variability for European summer heat waves. Nature 427:332–336

Seneviratne SI, Lüthi D, Litschi M, Schär C (2006) Land-atmosphere coupling and climate change in Europe. Nature 443:205–209

Shackley S, Young P, Parkinson S, Wynne B (1998) Uncertainty, complexity and concepts of good science in climate change modeling: are GCMs the best tools? Clim Change 38:159–205

Shukla J, DelSole T, Fennessy M, Kinter J, Paolino D (2006) Climate model fidelity and projections of climate change. Geophys Res Lett 33:Art. No. L07,702

Smith RL, Tebaldi C, Nychka D, Mearns LO (2009) Bayesian modeling of uncertainty in ensembles of climate models. J Am Stat Assoc 485:97–116

Snyder MA, Sanso B, Sloan LC (2007) Validation of climate model output using bayesian statistical methods. Clim Change 83:457–476

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough J, Knight S, Martin A, Murphy J, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433:403–406

Steppeler J, Doms G, Schättler U, Bitzer HW, Gassmann A, Damrath U, Gregoric G (2003) Meso-gamma scale forecasts using the nonhydrostatic model LM. Meteor Atmos Phys 82:75–96

Tebaldi C, Knutti R (2007) The use of multi-model ensemble in probabilistic climate projections. Phil Trans R Soc A 365:2053–2075

Tebaldi C, Sanso B (2008) Joint projections of temperature and precipitation change from multiple climate models: a hierarchical bayesian approach. J R Stat Soc A 171

Tebaldi C, Smith RL, Nychka D, Mearns LO (2005) Quantifying uncertainty in projection of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. J Clim 18:1524–1540

Vidale PL, Lüthi D, Frei C, Seneviratne S, Schär C (2003) Predictability and uncertainty in a regional climate model. J Geophys Res 108:Art. No. 4586

Vidale PL, Lüthi D, Wegmann R, Schär C (2007) European summer climate variability in a heterogeneous multi-model ensemble. Clim Change 81:209–232

Weigel AP, Liniger MA, Appenzeller C (2008) Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?. Quart J R Meteor Soc 134:241–260

Wilks DS (2006) Comparison of ensemble-MOS methods in the Lorenz '96 setting. Meteorol Appl 13:243–256