A Bayesian Hierarchical Model for Heterogeneous RCM–GCM Multimodel Ensembles*

CHRISTIAN KERKHOFF AND HANS R. KÜNSCH

Seminar for Statistics, ETH Zurich, Zurich, Switzerland

CHRISTOPH SCHÄR

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

(Manuscript received 29 August 2014, in final form 10 May 2015)

ABSTRACT

A Bayesian hierarchical model for heterogeneous multimodel ensembles of global and regional climate models is presented. By applying the methodology herein to regional and seasonal temperature averages from the ENSEMBLES project, probabilistic projections of future climate are derived. Intermodel correlations that are particularly strong between regional climate models and their driving global climate models are explicitly accounted for. Instead of working with time slices, a data archive is investigated in a transient setting. This enables a coherent treatment of internal variability on multidecadal time scales. Results are presented for four European regions to highlight the feasibility of the approach. In particular, the methodology is able to objectively identify patterns of variability changes, in ways that previously required subjective expert knowledge. Furthermore, this study underlines that assumptions about bias changes have an effect on the projected warming. It is also shown that validating the out-of-sample predictive performance is possible on short-term prediction horizons and that the hierarchical model herein is competitive. Additionally, the findings indicate that instead of running a large suite of regional climate models all forced by the same driver, priority should be given to a rich diversity of global climate models that force a number of regional climate models in the experimental design of future multimodel ensembles.

1. Introduction

Probabilistic projections of climate change, as for example presented in the 2013 IPCC assessment (IPCC 2013), are based on comprehensive multimodel ensembles of global climate models (GCMs; see, e.g., Taylor et al. 2012) and GCM-driven regional climate models (RCMs; see, e.g., van der Linden and Mitchell 2009; Mearns et al. 2009; Giorgi et al. 2009). In practice projections are thus usually derived from a series of heterogeneous GCM–RCM model chains. Analyzing and

interpreting such datasets is challenging from both a climate science but also statistical point of view (see, e.g., Tebaldi and Knutti 2007; Knutti et al. 2010; Stephenson et al. 2012).

Among the most frequent and simplest attempts is the calculation of the multimodel mean for a specific quantity of interest, possibly weighting each model by its ability to reproduce the observed climate as done for example by the reliability ensemble average (Giorgi and Mearns 2002).

However, assigning weights to the ensemble members is problematic for various reasons (see, e.g., Tebaldi and Knutti 2007; Knutti et al. 2010; Weigel et al. 2010; Déqué and Somot 2010; Chandler 2013). The choice of metric is not unique, such that different weighting schemes may lead to different climate scenarios (Lenderink 2010). Additionally, assigning a single numeric value to individual models is difficult when a suite of variables in different regions is considered, as performance of models varies from variable to variable and region to region. Assigning weights ultimately assigns probabilities to different models, and a formal and objective methodology based on sound probabilistic and statistical assessments is desirable.

Denotes Open Access content.

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-14-00606.s1>.

Corresponding author address: Christian Kerkhoff, Seminar for Statistics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland.
E-mail: kerkhoff@stat.math.ethz.ch

DOI: 10.1175/JCLI-D-14-00606.1

Therefore, coherent statistical treatments of multimodel ensembles have become appealing over the last decade. A Bayesian approach turns out to be especially suited because assumptions on the data generating process and expert information used in the analysis can nicely and transparently be laid out. Based on [Nychka and Tebaldi \(2003\)](#), who give a formal justification of the reliability ensemble average, [Tebaldi et al. \(2005\)](#) implemented a Bayesian analysis of a GCM multimodel ensemble. Similarly, [Greene et al. \(2006\)](#) choose model weights according to a Bayesian linear model. The approach of [Tebaldi et al. \(2005\)](#) has been generalized in many directions. Instead of analyzing different regions separately, [Smith et al. \(2009\)](#), for example, implement a multivariate analysis for various climatic regions. This approach has been further generalized by [Furrer et al. \(2007\)](#), [Jun et al. \(2008\)](#), [Kaufman and Sain \(2010\)](#), [Kang et al. \(2012\)](#), [Salazar et al. \(2011\)](#), [Sain et al. \(2011\)](#), [Furrer et al. \(2012\)](#), and [Geinitz et al. \(2015\)](#), who analyze fields of multimodel outputs via spatial processes. [Tebaldi and Sansó \(2009\)](#) and [Buser et al. \(2010b\)](#) implement a bivariate extension of [Tebaldi et al. \(2005\)](#) by modeling temperature and precipitation jointly. General considerations on the statistical treatment of multimodel ensembles with justifications for not using fully Bayesian methods are given in [Rougier et al. \(2013\)](#) and [Chandler \(2013\)](#).

In [Buser et al. \(2009\)](#) different bias assumptions for climate models are discussed. They compare two prominent assumptions called “constant bias” and “constant relation” [referred to as “bias correction” by [Ho et al. \(2012\)](#)] and show that climate projections substantially depend on the chosen assumption. The constant bias assumption underlies studies that identify climate change via the difference “scenario minus control” (see, e.g., [Maraun 2012](#)). In contrast, in the constant relation assumption, biases are climate state dependent and change with time (see also [Piani et al. 2010](#); [Boberg and Christensen 2012](#); [Christensen and Boberg 2012](#)). Within the constant relation approach, additive and multiplicative biases are considered, which are related to each other. Analyzing the same data as we do in this paper, [Kerckhoff et al. \(2014\)](#) found a preference for the constant relation assumption, although none of the assumptions seems to hold exactly. This suggests using an approach from [Buser et al. \(2010a\)](#) that provides a continuous transition between constant bias and constant relation with an additional parameter that is estimated from the data.

Besides the aforementioned bias assumptions, the literature on multimodel ensembles mainly relies on two paradigms ([Sanderson and Knutti 2012](#)). The first one is usually referred to as the “truth plus error” assumption

(see, e.g., [Tebaldi et al. 2005](#); [Buser et al. 2009](#)). It assumes that the multimodel ensemble members are centered around the true climate, with differences between the truth and the output from climate models being independent and identically distributed. Under this assumption, the error of the multimodel mean could be reduced to zero with a sufficiently large ensemble of climate models. This is not realistic as climate models are not independent, but rather share some basic structural features. The second assumption is based on the concept of exchangeability ([Annan and Hargreaves 2010](#); [Rougier et al. 2013](#)). It assumes that all members of the multimodel ensemble are equally likely and that the true climate can be regarded as an additional member of this ensemble. The difference between the truth and the multimodel mean is then comparable to the difference between a single member and the multimodel mean. More generally, [Chandler \(2013\)](#) presents an approach that assumes exchangeability conditional upon a model consensus representing a collective discrepancy between models and true climate. All these assumptions can be questioned, as there is a complex dependence structure between climate models due to their genealogy ([Masson and Knutti 2011](#)).

In this manuscript we introduce and apply a hierarchical model to seasonal and regional temperature averages from the ENSEMBLES project ([van der Linden and Mitchell 2009](#)). This project represents a large coordinated effort toward quantifying regional climate change over Europe, making use of observations and model output from 15 GCM–RCM chains; see [section 2](#). The proposed hierarchical model, described in [section 3](#), builds upon methods derived in [Buser et al. \(2009\)](#), [Buser et al. \(2010a\)](#), and [Kerckhoff et al. \(2014\)](#). Key extensions of the model include the following:

- The use of a hierarchical framework that allows for correlations between model chains. Such correlations are particularly relevant if the same RCM is driven by different GCMs ([Buser et al. 2010a](#); [A. Fischer et al. 2012](#)). Our approach follows [Kerckhoff et al. \(2014\)](#), and is similar to the one of [Heaton et al. \(2013\)](#), who introduced a Bayesian hierarchical analysis for the North American Regional Climate Change Assessment Program ([Mearns et al. 2009](#)). The hierarchical model exploits a simple linear relationship between RCMs and their driving GCMs with coefficients that depend on both the RCM and GCM components. This is a central assumption of our new model, which is well confirmed by data analysis (see [Fig. 1](#), left panel).
- The consideration of a transient setting rather than restricting attention to a control and scenario period as in many previous studies (e.g., [Buser et al. 2010a](#);

FIG. 1. Relation between GCM and RCM results exemplified by an analysis of winter (DJF) temperature means for the Mediterranean region from the ENSEMBLES dataset. (left) Scatterplot for the DMI RCM (y axis) vs its driving GCMs (x axis, color coded) for the period 1961–2099. Gray shaded areas indicate a 95% confidence interval for the regression fit. The black dashed line indicates a 1:1 relation (i.e., RCM and GCM temperatures perfectly agree). This figure underlines that RCMs strongly depend on their driving GCMs. (right) Scatterplot of the RCM increments (defined as RCM minus GCM mean temperatures) vs the GCM bias (GCM minus observations) for 1961–89. Symbols indicate different RCMs and colors the driving GCMs. The solid black line indicates a linear fit to the data cloud. Although RCMs are strongly determined by their driver, the RCM increments are anticorrelated with the GCM bias, and thus RCMs are able to compensate GCM biases. [Redrawn following Kerkhoff et al. (2014).]

A. Fischer et al. 2012; Heaton et al. 2013). This enables us to account for internal variability on multidecadal time scales. This type of variability is a major contributor to the overall uncertainty in climate projections (Hawkins and Sutton 2009; Deser et al. 2012) and may temporarily mask the effects of anthropogenic changes (i.e., global warming).

- The exploitation of a full Bayesian analysis rather than frequentist estimates as considered in Kerkhoff et al. (2014). This is essential since only a Bayesian analysis allows us to optimally separate bias changes from the climate signal, and to quantify uncertainties in the form of probability distributions. For the Bayesian analysis, a parametric approach to estimate multidecadal variability is preferable to the nonparametric approach used in Kerkhoff et al. (2014). We therefore use a basis of spline functions to represent multidecadal variability.
- The provision of a thorough statistical validation of the out-of-sample predictive performance using independent observations from 1991 to 2012. Although the predictive performance on longer time horizons would be more desirable, we think that the short-term validation provides a valuable additional check of our method.

For our model, the use of all GCM and RCM simulations is essential. Kerkhoff et al. (2014) found that RCMs are able to compensate for some of the driving GCMs' biases in mean temperature and interannual variability, even on scales that are well resolved by GCMs. This key result is evident from Fig. 1 (right panel), which shows that RCM increments or bias

“adjustments” (i.e., RCM minus GCM) are anticorrelated with the GCM biases (GCM minus OBS). As a result the overall biases of the model chains (i.e., RCM minus OBS) tend to be smaller than the GCM biases. While this result highlights the usefulness of model chains, it does not address another key question, namely whether RCMs provide information about how biases develop into the future. However, we do not make such a prior assumption in our analysis.

The reminder of this paper is organized as follows: Section 2 presents the data used, section 3 the methodology, section 4 the resulting climate change projections for four European regions, section 5 some sensitivity analysis using simpler statistical models or only parts of the available climate models, and section 6 an assessment of the predictive performance for the 20-yr period 1991–2012. The study is finally concluded in section 7.

2. Data

We analyze 2-m temperatures from the ENSEMBLES project (van der Linden and Mitchell 2009; Hewitt and Griggs 2004). In this project a suite of six GCMs with horizontal resolution of 100–300 km and 11 RCMs with 25-km resolution provide transient simulations for the period 1961–2099. The simulations use past observed greenhouse gas and aerosol concentrations, and the SRES A1B scenario (Nakicenovic and Swart 2000) after year 2000. Because of a lack of computational capacity, only 15 of all 66 possible RCM–GCM combinations were performed in the ENSEMBLES project (see Table 1).

TABLE 1. ENSEMBLES data matrix with rows indicating the institution providing an RCM and columns the driving GCM. Asterisks highlight different subsets of the dataset that are used. A single asterisk indicates the subset used in Buser et al. (2010a) and in the rcm approach. Here, only one RCM per GCM was used. This is to reduce dependence between different RCMs potentially arising through common driving GCMs. Two asterisks indicate the constrained ensemble from E. Fischer et al. (2012) that is used also in the con approach from section 5. This subensemble proved to have a realistic present-day interannual variability and soil moisture distribution and in turn projected increases in interannual summer variability (E. Fischer et al. 2012). (Expansions of acronyms can be found at <http://www.ametsoc.org/PubsAcronymList>. HadQ0, HadQ3, and HadQ16 are configurations of HadCM3. BCM is the Bergen Climate Model. HCQ0, HCQ3, and HCQ16 are configurations of HadRM3.)

RCM/GCM	HadQ0	HadQ3	HadQ16	ECHAM	ARPEGE	BCM
HCQ0	x					
HCQ3		x*				
HCQ16			x			
MPI				x**		
CNRM					x*	
Danish Meteorological Institute (DMI)				x	x	x
ETH	x*					
KNMI				x***		
International Centre for Theoretical Physics (ITCP)				x**		
Swedish Meteorological and Hydrological Institute (SMHI)		x**		x**		x***
Community Climate Change Consortium for Ireland (C4I)			x*			

Besides projections from climate models, we use observational data from version 9.0 of E-OBS gridded observations (Haylock et al. 2008). This archive spans the period 1950–2012, and the station data are already interpolated on the native grid of the ENSEMBLES RCMs such that validating the RCMs is particularly easy.

In a first postprocessing step we aggregated this data seasonally [winter is December–February (DJF), spring is March–May (MAM), summer is June–August (JJA), and autumn is September–November (SON)] and regionally using only data from land grid points for eight European climate regions [see Table 2; see also Fig. 4 in Christensen and Christensen (2007) for an illustration]. Because of this aggregation and because we analyze temperature, we think that we can neglect interpolation errors in the E-OBS data.

3. Methodology

In this section we will introduce and refine a hierarchical model that has been proposed in Kerkhoff et al. (2014). We will then estimate its parameters using a Bayesian approach; see, e.g., Gelman et al. (2013). The foundation of this approach is the computation of the posterior density by combining the prior, which specifies information from experts or previous studies, with the likelihood, which specifies information contained in the data. We first define the parameters of our statistical model, then specify the joint likelihood for the ENSEMBLES dataset and the prior, and finally describe how to explore the resulting posterior.

a. Notation

Because of the aforementioned data aggregation (see section 2) we assume independence between different

seasons and regions. We therefore analyze each season and region separately, which simplifies notation. We denote by $X_{o,t}$, $X_{g,t}$, and $X_{rg,t}$ the seasonal temperature averages for a specific season and region from the observations, GCM $g = 1, \dots, M$, and RCM–GCM combination $rg \in K$ in year t , respectively. For the observations we have data spanning the period 1961–2012. However, we only incorporate the first $T_1 = 30$ yr in the fitting procedure and keep the remaining 22 yr to validate our model. Contrarily, we use information from all climate models for the entire simulation period 1961–2099; that is, here we have $T_2 = 139$ yr of data at our disposal.

b. Hierarchical modeling of the ENSEMBLES dataset

In this subsection we present the basics of the hierarchical model, then section 3c documents the specification of time-varying parameters, section 3d discusses underlying bias assumptions, and finally section 3e explains how to estimate the parameters in a Bayesian setting.

TABLE 2. Latitudes and longitudes of the European climate regions considered in this study.

Area	Latitude	Longitude
Scandinavia (SC)	55°–70°N	5°–30°E
British Isles (BI)	50°–59°N	10°W–2°E
Middle Europe (ME)	48°–55°N	2°–16°E
Eastern Europe (EA)	44°–55°N	16°–30°E
France (FR)	44°–50°N	5°W–5°E
Alps (AL)	44°–48°N	5°–15°E
Iberian Peninsula (IP)	36°–44°N	10°W–3°E
Mediterranean region (MD)	36°–44°N	3°–25°E

Like Kerkhoff et al. (2014), we write the observations $X_{o,t}$ in year t as

$$X_{o,t} = \mu_{o,t} + \varepsilon_{o,t}, \quad (1)$$

where $\mu_{o,t}$ and $\varepsilon_{o,t}$ denote a slowly evolving observational mean, and faster-changing mean-zero yearly random fluctuations, respectively. Because of seasonal and regional aggregation, we assume that the $\varepsilon_{o,t}$ values are independent and normally distributed. Their standard deviation $\sigma_{o,t}$ characterizes the size of yearly fluctuations in observed temperature (e.g., due to changes in storm tracks). In this manuscript we use throughout the term “interannual variability” for the standard deviation of year-to-year fluctuations about the overall mean level.

Furthermore, we suppose that $\mu_{o,t}$ can be decomposed into a slowly varying climatological signal $\mu_{\text{clim},t}$ arising from external anthropogenic forcing and contributions from internal multidecadal variability $v_{o,t}$ that change at a faster pace (e.g., due to changes in the Atlantic meridional overturning circulation):

$$\mu_{o,t} = \mu_{\text{clim},t} + v_{o,t}. \quad (2)$$

The rationale behind this decomposition is as follows. While faster-changing yearly fluctuations do not affect the climate mean state, we assume that the slowly varying component of the internal variability does affect the climate on decadal time scales. The effects from multidecadal variability are thus able to mask potential trends, say from anthropogenic greenhouse gas emissions, on decadal time scales.

Similarly, the output of GCM g in year t is written as

$$X_{g,t} = \mu_{g,t} + \varepsilon_{g,t}, \quad (3)$$

where $\mu_{g,t}$ and $\varepsilon_{g,t}$ have the same interpretations as in Eq. (1). Also here, we assume that the $\varepsilon_{g,t}$ values are mean-zero, independent, and normally distributed with standard deviation $\sigma_{g,t}$. However, because climate models only imperfectly reflect reality, we allow for additive biases in the climatological component of the mean, denoted by $\beta_{g,t}$, and multiplicative biases $b_{g,t}$ in interannual variability. Additionally, we assume that each GCM has its own multidecadal variations, which have similar time scales and amplitudes as $v_{o,t}$, but are otherwise not related. Thus we decompose the GCM mean in year t into

$$\mu_{g,t} = \mu_{\text{clim},t} + \beta_{g,t} + v_{g,t} \quad (4)$$

and the GCM standard deviation in year t into

$$\sigma_{g,t} = \sigma_{o,t} b_{g,t}. \quad (5)$$

While the systematic biases $\beta_{g,t}$ and $b_{g,t}$ stem from such factors as imperfect parameterizations, missing processes, or discretization errors of the underlying equations, $\varepsilon_{g,t}$ and $v_{g,t}$ correspond to different realizations of internal variability. Because of the chaotic nature of the climate system, climate models do not reproduce the observed yearly and decadal variations.

As shown in Fig. 1 (left panel) there is a strong, linear dependence between seasonal mean temperatures from the ENSEMBLES RCMs and their driving GCMs. Therefore, Kerkhoff et al. (2014) have proposed a hierarchical statistical model that specifies the conditional distribution of the RCM data given the driving GCM. Because the intercept and slope in Fig. 1 depend on both the RCM and the driving GCM, it seems reasonable to assume that the output from RCM rg in year t given its driving GCM g can be written as

$$X_{rg,t} = \xi_{rg,t} + \phi_{rg,t} X_{g,t} + \varepsilon_{rg,t}, \quad (6)$$

where $\xi_{rg,t}$ and $\phi_{rg,t}$ denote the slowly varying intercept and slope, respectively, and the deviations $\varepsilon_{rg,t}$ from the regression line are independent and normally distributed with standard deviation $\tau_{rg,t}$.

c. Separation of the climatological signal and multidecadal variations

The components $\mu_{\text{clim},t}$, $\beta_{g,t}$, $v_{o,t}$, and $v_{g,t}$ of the mean are time-dependent functions. To distinguish them from each other, and to separate them from the yearly random fluctuations $\varepsilon_{o,t}$ and $\varepsilon_{g,t}$, we represent $\mu_{\text{clim},t}$ and $\beta_{g,t}$ as linear combinations of three basis functions $\psi_1, \dots, \psi_3$ and represent $v_{o,t}$ and $v_{g,t}$ as linear combinations of four other basis functions $\psi_4, \dots, \psi_7$. These basis functions, which reflect the different time scales of climate signal and multidecadal variations, are illustrated in Fig. 2. The first one, $\psi_1(t) = 1$, is a constant, and $\psi_2(t) = (t - T_0)/100$ with $T_0 = (1961 + 2099)/2 = 2030$ is a linear function. The construction of the other five basis functions, which uses splines, is described in the appendix. To guarantee that the different components can be separated, we choose orthogonal (uncorrelated) basis functions.

Therefore the different components of the mean functions have the form

$$\mu_{\text{clim},t} = c_{1,o} + c_{2,o} \frac{t - T_0}{100} + c_{3,o} \psi_3(t), \quad (7)$$

$$\beta_{g,t} = \beta_{1,g} + \beta_{2,g} \frac{t - T_0}{100} + \beta_{3,g} \psi_3(t), \quad (8)$$

FIG. 2. Orthogonal basis functions [$\psi_k(t)$, colors and line types] for (left) the climatological component and (right) the multidecadal variability. Dots along the x-axis indicate the position of the interior knot(s). The climatological signal is split into (solid black) a constant, (green dashed) linear, and (orange dotted-dashed) a cubic contribution, whereas the multidecadal variability comprises higher-order cubic polynomials (colors and line types) accounting for faster changes in this component.

and

$$v_{o,t} = \sum_{k=4}^7 c_{k,o} \psi_k(t), \quad v_{g,t} = \sum_{k=4}^7 c_{k,g} \psi_k(t), \quad (9)$$

with the c and β terms denoting the unknown parameters of the representation. In particular, $c_{1,o}$ is the true mean in the year 2030, $\theta_{2,o}$ is the temperature change per century, and the sign of $c_{3,o}$ decides whether $\mu_{\text{clim},t}$ is convex or concave.

By Eqs. (2), (7), and (9), the observed mean $\mu_{o,t}$ is given by

$$\mu_{o,t} = c_{1,o} + c_{2,o} \frac{t - T_0}{100} + c_{3,o} \psi_3(t) + \sum_{k=4}^7 c_{k,o} \psi_k(t), \quad (10)$$

whereas the GCM mean follows from Eqs. (4) and (7)–(9):

$$\begin{aligned} \mu_{g,t} = & c_{1,o} + \beta_{1,g} + (c_{2,o} + \beta_{2,g}) \frac{t - T_0}{100} \\ & + (c_{3,o} + \beta_{3,g}) \psi_3(t) + \sum_{k=4}^7 c_{k,g} \psi_k(t). \end{aligned} \quad (11)$$

d. Other time-dependent parameters

In our approach, the variance parameters $\sigma_{o,t}$, $b_{g,t}$, and $\tau_{\text{rg},t}$ and the regression parameters $\xi_{\text{rg},t}$ and $\phi_{\text{rg},t}$ are also allowed to vary slowly with time. In principle, they could even exhibit variations on different time scales like the mean. However, the changes in these parameters are generally smaller and trying to separate different time scales for them would be even more difficult. We therefore prefer a rather simple approach and work with linear functions. This will nevertheless capture the main changes of these parameters.

For the variance parameters (expressed on the log-scale), we assume

$$\log(\sigma_{o,t}) = \log(\sigma_o) + \frac{t - T_0}{100} \log(\delta\sigma_o),$$

$$\log(b_{g,t}) = \log(b_g) + \frac{t - T_0}{100} \log(\delta b_g). \quad (12)$$

In this way, σ_o is the interannual variability in year $T_0 = 2030$, $\delta\sigma_o$ denotes the change in interannual variability per century, b_g is the multiplicative bias in year 2030, and δb_g is the change in multiplicative bias per century.

Likewise, for the standard deviation $\tau_{\text{rg},t}$ of the deviations $\varepsilon_{\text{rg},t}$ in Eq. (6), we again use a linear function on the log-scale such that

$$\log(\tau_{\text{rg},t}) = \log(\tau_{\text{rg}}) + \frac{t - T_0}{100} \log(\delta\tau_{\text{rg}}). \quad (13)$$

Finally, for the intercept and slope parameter of the hierarchical model, we assume in accordance with Fig. 1 that

$$\xi_{\text{rg},t} = \xi_{1,\text{rg}} + \xi_{2,\text{rg}} \frac{t - T_0}{100}, \quad \phi_{\text{rg},t} = \phi_{1,\text{rg}} + \phi_{2,\text{rg}} \frac{t - T_0}{100}. \quad (14)$$

e. Bias assumptions

The main parameters of interest are $\mu_{\text{clim},t}$ and $\sigma_{o,t}$. Our ultimate goal is to estimate these parameters as precisely as possible and quantify their remaining uncertainty until the end of this century. However, without any prior assumptions on the bias change parameters $\beta_{2,g}$, $\beta_{3,g}$, and δb_g , the means and standard deviations of models and observations are not related to each other. Hence the climate models would not provide information on $\mu_{\text{clim},t}$ and $\sigma_{o,t}$, and for instance $\mu_{\text{clim},t}$ would be estimated by fitting a mean of the form Eq. (10) to the available observations. As a consequence, the uncertainty about $\mu_{\text{clim},t}$ in the later part of this century would be huge.

To gain information about $\mu_{\text{clim},t}$ from the climate models, we therefore need some assumptions about how biases change with time. Such assumptions usually restrict the size of $\beta_{2,g}$, $\beta_{3,g}$, and δb_g , either by imposing hard constraints or by specifying informative prior distributions. However, as reviewed in Kerkhoff et al. (2014), there are also plausible assumptions that postulate that the additive bias changes $\beta_{2,g}$ and $\beta_{3,g}$ are related to the multiplicative bias b_g . We therefore consider the following form of the GCM mean, which contains most bias assumptions that have been discussed in the literature as special cases:

$$\begin{aligned} \mu_{g,t} = & c_{1,o} + \beta_{1,g} + \{[1 + \kappa(b_g - 1)]c_{2,o} + \delta\beta_{2,g}\} \frac{t - T_0}{100} \\ & + \{[1 + \kappa(b_g - 1)]c_{3,o} + \delta\beta_{3,g}\} \psi_3(t) + \sum_{k=4}^7 c_{k,g} \psi_k(t). \end{aligned} \quad (15)$$

This coincides with Eq. (11) if for $k = 2, 3$ we set $\beta_{k,g} = \kappa(b_g - 1)c_{k,g} + \delta\beta_{k,g}$.

To explain the role of the parameters $\delta\beta_{2,g}$, $\delta\beta_{3,g}$, and κ , we now discuss a number of special cases. If $\delta\beta_{2,g} = \delta\beta_{3,g} = 0$ and $\kappa = 0$, we have by Eqs. (7) and (9) for $t_2 > t_1$

$$\mu_{g,t_2} - \mu_{g,t_1} = \mu_{\text{clim},t_2} - \mu_{\text{clim},t_1} + (v_{g,t_2} - v_{g,t_1}).$$

That is, the GCM climate change signal coincides with the true one, apart from multidecadal variations. This is a strict version of the constant bias assumption that is used implicitly in many studies (see Buser et al. 2009).

Alternatively, if $\delta\beta_{2,g} = \delta\beta_{3,g} = 0$ and $\kappa = 1$, we obtain for $t_2 > t_1$

$$\mu_{g,t_2} - \mu_{g,t_1} = b_g(\mu_{\text{clim},t_2} - \mu_{\text{clim},t_1}) - (v_{g,t_2} - v_{g,t_1});$$

that is, a GCM overestimating the interannual variability will also overestimate the true climate change by the same amount, again except for multidecadal variations. This is a strict formulation of the constant relation assumption (see Buser et al. 2009; Ho et al. 2012).

If $\delta\beta_{2,g} = \delta\beta_{3,g} = 0$ and κ denotes an unknown bias parameter between 0 and 1, we have a compromise between constant bias and constant relation as used in Buser et al. (2010a).

Additional flexibility is obtained if we replace the hard constraints $\delta\beta_{2,g} = \delta\beta_{3,g} = 0$ by a prior distribution that says that these parameters are most likely small compared to $c_{2,o}$ and $c_{3,o}$. If $\kappa = 0$, we obtain a weak form of the constant bias assumption, and if $\kappa = 1$ we get a weak form of the constant relation assumption. If κ is unknown, except for the restriction to the interval $[0, 1]$,

we let the data decide about which bias assumptions are likely. If we include the parameters $\delta\beta_{k,g}$ in the model, the terms $\kappa(b_g - 1)c_{k,o}$ could then be absorbed in $\delta\beta_{k,g}$. Therefore the possible values of the mean functions $\mu_{g,t}$ are the same regardless of whether the parameter κ is in the model or not. However, the posterior distribution of the mean values $\mu_{g,t}$ changes because the terms $\kappa(b_g - 1)c_{k,o}$ can reduce the size of $\delta\beta_{k,g}$.

Similarly, instead of a hard constraint $\delta b_g = 1$ for the multiplicative bias change, we will choose informative priors, which say that the δb_g is likely to be closer to one than the true change $\delta\sigma_o$. Since by Eqs. (5) and (12)

$$\sigma_{g,t} = \sigma_o (\delta\sigma_o)^{(t-T_0)/100} b_g (\delta b_g)^{(t-T_0)/100}, \quad (16)$$

such priors allow us to obtain additional information on $\delta\sigma_o$ from the estimation of $\sigma_{g,t}$.

f. The likelihood

The likelihood function quantifies how likely the parameters of our statistical model are given the data. It is the conditional density of all observations and all GCM and RCM outputs given the parameters of our hierarchical model. Assuming independence between all annual fluctuations $\epsilon_{o,t}$, $\epsilon_{g,t}$, and $\epsilon_{rg,t}$ and choosing normal distributions, the likelihood becomes

$$\begin{aligned} & \prod_{t=1}^{T_1} \frac{1}{\sigma_{o,t}} \exp \left[-\frac{(X_{o,t} - \mu_{o,t})^2}{2\sigma_{o,t}^2} \right] \\ & \times \prod_{t=1}^{T_2} \prod_{g=1}^M \frac{1}{b_{g,t} \sigma_{o,t}} \exp \left[-\frac{(X_{g,t} - \mu_{g,t})^2}{2\sigma_{o,t}^2 b_{g,t}^2} \right] \\ & \times \prod_{t=1}^{T_2} \prod_{(rg) \in K} \frac{1}{\tau_{rg,t}} \exp \left[-\frac{(X_{rg,t} - \xi_{rg,t} - \phi_{rg,t} X_{g,t})^2}{2\tau_{rg,t}^2} \right], \end{aligned} \quad (17)$$

where the meaning and the definition of the parameters are summarized in Table 3.

g. Specification of the prior

In the Bayesian approach, we need to attach prior distributions to all unknown parameters. For computational convenience, we take normal priors for all mean and regression parameters, and instead of the variance parameters we work with their inverses, called precisions, for which we use gamma priors (see, e.g., Gelman et al. 2013). Note that the reparameterization in terms of precisions is for computational reasons only, and does not affect the outcome of our analysis. For the

TABLE 3. Parameters of the Bayesian hierarchical model, their meaning, and the associated equation numbers for their definitions.

Parameter	Meaning	Equation
$\mu_{o,t}$	Observation mean	(2), (10)
$\mu_{\text{clim},t}$	Mean climatological signal	(7)
$\mu_{g,t}$	GCM mean	(4), (15)
$\beta_{g,t}$	Additive GCM bias	(8)
$v_{o,t}$	Observed multidecadal variations	(9)
$v_{g,t}$	GCM multidecadal variations	(9)
$\sigma_{o,t}$	Observed interannual variability	(12)
$\sigma_{g,t}$	GCM interannual variability	(5), (16)
$b_{g,t}$	Multiplicative GCM bias	(12)
$\xi_{\text{rg},t}$	Intercept of the regression line for an RCM vs its driving GCM	(14)
$\phi_{\text{rg},t}$	Slope of the regression line for an RCM vs its driving GCM	(14)
$\tau_{\text{rg},t}$	Interannual variability of residuals in the regression	(13)

bias parameter κ , we choose a uniform distribution on $[0, 1]$.

Prior distributions own hyperparameters that need to be specified. In case of normal priors, we have to choose means and variances and for the gamma distributions we need to specify shape and rate values. Hyperparameter values for the prior distributions associated with the hierarchical model parameters are given in Table 4.

We use weakly informative priors where possible; they carry little information and thus the posterior is mainly determined by the data. However, for those parameters associated with multidecadal variability ($c_{k,o}$, $k \geq 4$) or bias changes ($\delta\beta_{2,g}$, $\delta\beta_{3,g}$, and δb_g^{-2}) we rely on

informative priors in order to obtain better estimates of $\mu_{\text{clim},t}$ and $\sigma_{o,t}$. We have analyzed the sensitivity of our prior choices to the resulting posterior densities of $c_{2,o}$ and found similar results as in Buser et al. (2009), namely that very flat (sharp) priors lead to flat (sharp) posteriors. This sensitivity to prior choices simply reflects the fact that we cannot distinguish between a true change of the climate and a common change in the bias of all models. The priors have to reflect our understanding of the processes involved. Our approach assumes that the bias change $\delta\beta_{2,g}$ is smaller than the constant bias part $\beta_{1,g}$ and also smaller than the true changes $c_{2,o}$.

TABLE 4. Specification of the prior distribution for the parameters of the Bayesian hierarchical model. The first column indicates the parameter considered, the second column its prior distribution, the third and fourth columns are the associated hyperparameter values, and the last column indicates an interval containing 95% probability a priori for each specification, with limits rounded to one decimal.

Parameter	Distribution	Hyperparameter 1 (mean, shape)	Hyperparameter 2 (variance, rate)	95% coverage (prior interval)
$c_{1,o}$	(°C)	Normal	0 (winter) 10 (spring) 15 (summer) 10 (autumn)	100 [−19.6, 19.6] [−9.6, 29.6] [−4.6, 34.6] [−9.6, 29.6]
$c_{2,o}$	(°C)	Normal	0	25 [−9.8, 9.8]
$c_{3,o}$	(°C)	Normal	0	4 [−4, 4]
$\beta_{1,g}$	(°C)	Normal	0	25 [9.8, 9.8]
$\delta\beta_{2,g}$	(°C)	Normal	0	4 [−4, 4]
$\delta\beta_{3,g}$	(°C)	Normal	0	1 [−2, 2]
$c_{k,o}, k \geq 4$	(°C)	Normal	0	6.25 [−4.9, 4.9]
$c_{k,g}, k \geq 4$	(°C)	Normal	0	6.25 [−4.9, 4.9]
σ_o^{-2}	(°C ^{−2})	Gamma	0.1	0.1 (0, 10]
$\delta\sigma_o^{-2}$	—	Gamma	0.1	0.1 (0, 10]
b_g^{-2}	—	Gamma	0.1	0.1 (0, 10]
δb_g^{-2}	—	Gamma	3	3 [0.2, 2.5]
τ_{rg}^{-2}	—	Gamma	0.1	0.1 (0, 10]
$\delta\tau_{\text{rg}}^{-2}$	—	Gamma	3	3 [0.2, 2.5]
$\xi_{1,\text{rg}}$	(°C)	Normal	0	25 [−9.8, 9.8]
$\xi_{2,\text{rg}}$	(°C)	Normal	0	16 [−7.9, 7.9]
$\phi_{1,\text{rg}}$	—	Normal	0.8	1 [−1.2, 2.8]
$\phi_{2,\text{rg}}$	—	Normal	0	0.25 [−1, 1]

The multidecadal variations of observations and GCMs are not related, but we know from preindustrial control runs that their amplitude is below 0.4°C . We therefore choose the prior variance σ_c^2 of $c_{k,o}$ and $c_{k,g}$ for $4 \leq k \leq 7$ to match this value. By Eq. (A1) in the appendix, we obtain

$$\mathbb{E}\left(\frac{1}{T_2} \sum_{t=1}^{T_2} v_{g,t}^2\right) = \frac{4\sigma_c^2}{T_2} \approx 0.4^2 \Leftrightarrow \sigma_c \approx 2.5,$$

where $\mathbb{E}(\cdot)$ denotes the expectation operator under the prior.

Furthermore, we assume that the parameters of our hierarchical model are a priori independent. In this way, our formulation of the statistical model in section 3 amounts to using the “truth plus error” assumption for the climate signal and an exchangeability assumption for multidecadal variations. If we wanted to incorporate also the exchangeability assumption for the signal or a more general framework as in Chandler (2013), we would have to introduce correlations in the prior for $\beta_{k,g}$. However, because we allow for bias changes $\delta\beta_{k,g}$ and only have six GCMs at our disposal, we do not expect that this would have a big impact on our results.

h. Exploring the posterior

In a Bayesian analysis, conclusions about the parameters are based on the posterior distribution. According to Bayes formula, the joint posterior of all the parameters of the model given the data is proportional to the prior times the likelihood of the data. Hence, in principle the posterior is known. This, however, is of little practical help since computing the marginal posterior distributions of, for example, $\mu_{\text{clim},t}$ and $\sigma_{o,t}$, or the posterior predictive density of unknown future observations, involves high-dimensional integration, which is analytically intractable. Therefore, we rely on Markov chain Monte Carlo methods to draw samples from the posterior. These methods produce correlated samples from the posterior distribution via iterative simulation [for an introduction to Markov chain Monte Carlo methods, see, e.g., Gilks et al. (1996)]. It is necessary to discard the first part of the algorithm output (the “burn-in” period) before the simulations have converged properly to the required posterior.

We employ a Gibbs sampler with random-walk Metropolis steps since some of the so-called full conditional distributions (e.g., the one for $\delta\sigma_o^{-1}$) are neither log-concave nor from a standard family of distributions. Based on inspection of trace plots to determine convergence, we choose the first 50 000 as the burn-in period

and perform another 100 000 simulation steps. To reduce dependence and to save on storage, we only keep the result of every hundredth step.

4. Climate change projections

In this section we present the results for the main parameters of interest of our statistical model (i.e., the climatological mean $\mu_{\text{clim},t}$ and the interannual variability $\sigma_{o,t}$). We use the hierarchical model as specified in section 3 with the GCM means given by Eq. (15) and the priors as specified in Table 4. Unless otherwise stated, we mainly focus on winter (DJF) and summer (JJA) for SC, ME, AL, and MD for the sake of brevity. This longitudinal cross section through Europe should, however, be a reasonable representative for the entire continent. Results for the other regions and seasons are given in the supplementary material.

First, we consider temperature changes for 2060 and 2085 with respect to year 2000. Posterior densities of these changes are given in Fig. 3 and in Fig. S1 of the supplementary material. These results support earlier model analyses and find a pronounced positive trend in the climatological signal for Europe. However, as we go further into the future the posterior uncertainty increases. Additionally, a latitudinal and seasonal dependence is noticeable in Fig. 3: in summer (winter) strongest increases are found in southern (northern) Europe (cf. MD with SC). This is in line with the climate research literature (IPCC 2013).

To assess the results of our Bayesian analysis, we show in Fig. 3 and Fig. S1 also estimates of the same temperature changes based on a simpler model for individual GCM and RCM outputs, indicated by pluses and dashes, respectively. These estimates are obtained by taking the difference between local linear fits to 29-yr GCM and RCM time slices centered at the respective year. Because temperature changes on these scales are dominated by anthropogenic greenhouse gas forcing rather than internal climate variability [see the comment by Laprise (2014)], a smoothing window of 29 yr should be adequate to extract climatological signals without being affected by multidecadal variability. By taking differences, these estimates implicitly make the constant bias assumption. Usually, the posterior density covers the range of the estimates from the simpler model. For the Mediterranean region in winter, however, this is not the case. Here, the posterior density is shifted to somewhat lower temperature increases. The reason for this is that our Bayesian analysis attributes part of the of the temperature change signal found in the climate models to a bias that is not constant.

FIG. 3. Posterior densities for mean (top) winter and (bottom) summer temperature changes. (left)–(right) Results for Scandinavia (SC), middle Europe (ME), the Alps (AL), and the Mediterranean region (MD) with green long dashed lines indicating changes for 2060 and gold lines for 2085 with respect to 2000. Pluses at the bottom and short vertical lines at the top of each panel indicate estimates of these changes from local linear fits to 29-yr GCM and RCM time slices, respectively, centered at the respective year.

Next, we show in Fig. 4 and Fig. S2 of the supplementary material the posterior densities of the changes in interannual variability, defined as the ratio $\sigma_{o,t_1}/\sigma_{o,t_0}$, with t_1 being a future year and $t_0 = 2000$. Similarly as in the previous figure, we also show the variability changes estimated by the simpler analysis of individual GCM and RCM time series. They are computed from the residual sum of

squares from local linear fits to the same time slices as above.

In winter (as in spring and autumn; see Fig. S2), changes in interannual variability are small; the posterior density of $\sigma_{o,t_1}/\sigma_{o,t_0}$ is usually concentrated around 1 (corresponding to no change), or in the case of Scandinavia values below 1 (corresponding to a decrease in variability) are more likely. Contrarily, we find an

FIG. 4. As in Fig. 3, but for the change in interannual variability.

FIG. 5. Posterior predictive densities for mean (top) winter and (bottom) summer temperatures for (left)–(right) the different regions and years: black short dashed for 2000, green long dashed for 2060 and gold lines for 2085.

increase in interannual summer variability for most regions (i.e., future summers will likely be more variable). This also suggests that extreme summer conditions will probably occur more often in the future. Our results are consistent with the early analyses of variability changes in output from a previous model generation (Schär et al. 2004; Vidale et al. 2007) and contradict previous studies that find virtually no increase in interannual summer variability in the ENSEMBLES dataset (Buser et al. 2010a). Possible explanations of these contradictory findings will further be examined in section 5.

In contrast to the climatological mean, the posterior densities of the change in interannual variability do not always span the entire range of the estimates from a simpler model (e.g., the Alps in summer). This

indicates that our Bayesian analysis identifies models with very large interannual variability as outliers and downweights them in the estimation of $\sigma_{o,t}$. For an overview of the estimated climate model biases in the mean and interannual variability refer to Figs. S3–S6 of the supplementary material.

We have also calculated the posterior predictive densities of $X_{o,t}$ for $t = 2000, 2060$, and 2085 ; see Fig. 5 and Fig. S7 of the supplementary material. These densities represent the uncertainty about the seasonal and regional average temperature in these years, given the observations and the multimodel ensemble. They are obtained by integrating out the unknown mean $\mu_{o,t}$ and variance $\sigma_{o,t}^2$ in the density of $X_{o,t}$ with respect to the posterior:

$$p(x_{o,t} | \text{data}) = \iint \frac{1}{\sqrt{2\pi}\sigma_{o,t}} \exp \left[-\frac{(x_{o,t} - \mu_{o,t})^2}{2\sigma_{o,t}^2} \right] p(\mu_{o,t}, \sigma_{o,t} | \text{data}) d\mu_{o,t} d\sigma_{o,t}. \quad (18)$$

We approximate this integral by the average of normal densities with $\mu_{o,t}$ and $\sigma_{o,t}$ equal to the sample values from the posterior. In accordance with Figs. 3 and 4, we find a robust shift of the temperature distribution with time for all seasons and a fairly constant width of the temperature distribution for winter, whereas in summer the temperature variability increases.

5. Sensitivity study

There is no consensus on the likely trends in interannual summer variability in the ENSEMBLES data archive. While some studies find virtually no increase in variability (Buser et al. 2010a), our analysis as well as other studies that incorporate a subset of the entire

ensemble (E. Fischer et al. 2012) show a robust increase. This section aims at finding reasons for these differences and therefore we have set up and fitted a suite of five alternative statistical models, which deviate significantly from our hierarchical model, as presented in section 3 and hereafter denoted as hbm, in terms of formulation or data usage.

The first alternative is a reduction of hbm that does not contain multidecadal variability, that is, with $\mu_{o,t} = \mu_{\text{clim},t}$ and $\mu_{g,t} = \mu_{\text{clim},t} + \beta_{g,t}$ in Eqs. (2) and (4), respectively. That is, we simply set $c_{k,o} = c_{k,g} = 0$ for $k = 4, \dots, 7$ when fitting the model. We call this model “hbm w/o mdv.”

The second alternative does not make explicit use of any GCM and considers only a subset of RCMs from the ENSEMBLES project that were found to have a realistic present-day interannual variability and in turn projected a robust increase in interannual summer variability for southeastern Europe [see E. Fischer et al. (2012) and the data matrix in Table 1 herein]. In this alternative model, the distribution of the observations is unchanged, and in analogy to Eq. (3), we set

$$X_{\text{rg},t} = \mu_{\text{rg},t} + \varepsilon_{\text{rg},t}.$$

We explicitly account for multidecadal variability and allow the same bias changes as in hbm. In accordance with Eq. (15) the RCM mean is thus given by

$$\begin{aligned} \mu_{\text{rg},t} = & c_{1,o} + \beta_{1,\text{rg}} + \{[1 + \kappa(b_{\text{rg}} - 1)]c_{2,o} + \delta\beta_{2,\text{rg}}\} \frac{t - T_0}{100} \\ & + \{[1 + \kappa(b_{\text{rg}} - 1)]c_{3,o} + \delta\beta_{3,\text{rg}}\} \psi_3(t) \\ & + \sum_{k=4}^7 c_{k,\text{rg}} \psi_k(t), \end{aligned}$$

where $\beta_{1,\text{rg}}$, $\delta\beta_{2,\text{rg}}$, $\delta\beta_{3,\text{rg}}$, b_{rg} , and $c_{k,\text{rg}}$, $k \geq 4$ have obvious interpretations. The standard deviation of $\varepsilon_{\text{rg},t}$ is specified as

$$\sigma_{o,t} b_{\text{rg},t} = \sigma_o (\delta\sigma_o)^{(t-T_0)/100} b_{\text{rg}} (\delta b_{\text{rg}})^{(t-T_0)/100}.$$

We neglect potential correlations between RCMs that share the same GCM and between runs of the same RCM driven by different GCMs. We refer to this alternative model as con (for the constrained ensemble used).

The third alternative focuses on the GCM part of the hbm (i.e., only the GCMs are used whereas all RCMs are omitted). The distribution of the observations and of the GCM outputs is unchanged. As for hbm, we neglect any correlations between GCMs that might occur due to shared knowledge, code, etc., across institutions. This approach is called gcm.

The fourth alternative (hereafter rcm) is similar to con, but it is based on a different subset of RCM simulations, namely the one used in Buser et al. (2010a) (see Table 1). This subsets consists of six different RCMs driven by six different GCMs in order to avoid potential correlations between models.

As described in section 3, the hbm allows for bias changes $\delta\beta_{2,g}$ and $\delta\beta_{3,g}$, but with a prior that restricts their size. However, in climate change studies, one often assumes that these changes are much smaller than climate change and consequently can be neglected. To test if this assumption has significant effects on our climate change projections, we set $\delta\beta_{2,g}$ and $\delta\beta_{3,g}$ to zero in the approach denoted nbc (for no bias changes). We keep, however, the unknown parameter κ with values between 0 and 1 so that a continuous interpolation between constant bias and constant relation is possible.

Posterior densities of the change in interannual variability are given in Fig. 6 and Fig. S8 of the supplementary material. For comparability's sake, we included the geometric mean of the frequentist estimates from Fig. 4 and subsets of this data according to the different models used.

Differences between the approaches are usually minor in winter and especially between the hbm, the one without multidecadal variability (hbm w/o mdv), and nbc, which omits bias changes.

In summer, however, the rcm approach, which relies on the methodology derived in Buser et al. (2010a), projects no increase in interannual variability for southern Europe in contrast to our hbm (cf. red with dark blue boxplots in the bottom of Fig. 6 and Fig. S8). These differences stem from different data sources and their statistical treatment. For the rcm, a subset of RCMs (those that do not share a common driver) were used to determine the posterior distribution of the change in interannual variability. Contrarily, in our hierarchical analysis this change basically depends on the driving GCMs. Because the ENSEMBLES GCMs project a robust increase in southern Europe (see the light blue boxplots), also our hbm projects such increases.

More importantly, however, the hbm arrives at the same conclusion as the con that uses the subjectively constrained ensemble from E. Fischer et al. (2012). This constrained ensemble was chosen to ensure realistic present-day interannual variability in the current climate. In the assessment of the hbm this information is used implicitly but objectively, and it is reassuring that our hbm is able to project a similar increase in interannual summer variability.

This also indicates that the choice of models in Buser et al. (2010a) was a bit unfortunate since a set of models

FIG. 6. Box plots for the posterior distribution of the change in interannual variability for 2060 and 2085 (left and right parts of each panel, respectively) with respect to 2000 for (left)–(right) the different regions. Results in each panel are sequentially given for the Bayesian hierarchical model (hbm, red), the same model without multidecadal variability (hbm w/o mdv, brown), the constrained ensemble from E. Fischer et al. (2012b) (con, green), a Bayesian model fit to GCM data (gcm, light blue), a subset of RCMs from Buser et al. (2010a) (rcm, dark blue), and a hierarchical model omitting bias changes (nbc, pink). Results are given for (top) winter and (bottom) summer. Diamonds indicate the geometric ensemble mean from Fig. 4 for the respective models used.

with unrealistic present-day interannual variability was chosen to derive climate change projections. Biases in interannual summer variability stem from soils that are too dry to evaporate (E. Fischer et al. 2012; Bellprat et al. 2013). Likewise, the projected increases in summer variability are interpreted as resulting from the large-scale drying of the European continent in summer (Seneviratne et al. 2006; Mueller and Seneviratne 2012), so there is a physical linkage between the representation of current-day variability and projections of variability changes, and this linkage supports the formulation of our model that entails a corresponding linkage with the constant relation approach.

We have also analyzed the sensitivity of the change in $\mu_{\text{clim},t}$ (see Fig. 7 and Fig. S9 of the supplementary material). Again, differences between the approaches are small in winter. In summer, however, using the con or nbc model leads to a weaker warming compared to the other approaches (cf. green and purple to the other boxplots). This can be traced back to the choice of bias assumption (see Fig. 8 and Fig. S10 of the supplementary material). Assuming no bias changes (nbc) or choosing the constrained ensemble (con) from E. Fischer et al. (2012) leads to posterior densities of κ that are closer to 1, favoring the constant relation assumption. Contrarily, our original analysis (hbm) slightly prefers constant bias

(κ closer to 0). As already stated in Buser et al. (2009), under constant relation, parts of the climate change signal are attributed to biases such that the resulting climate change signal is smaller than under constant bias.

6. Assessment of the predictive performance

In addition to the sensitivity study, we have assessed the out-of-sample predictive performance of our six statistical models from section 5. For this assessment we use the observations from the period 1991–2012. Since these observations have not been used during model fitting, we do compare the out-of-sample predictive performance and there is no unfair advantage for more complex models.

This step is important in any statistical analysis and it is particularly easy in our Bayesian analysis, because it provides predictive distributions of the observed temperatures in each year t between 1991 and 2012. For a definition of the posterior predictive distribution see Eq. (18) and the text. Similar to Salazar et al. (2011), we use the continuous ranked probability score (CRPS) that assesses both calibration and sharpness of the predictive distribution (Matheson and Winkler 1976). The CRPS is expressed in degrees Celsius, and lower values of the CRPS indicate a better predictive performance. Because we approximate the posterior by sampling, the

FIG. 7. As in Fig. 6, but for the temperature change (°C). Here, diamonds indicate the ensemble mean from Fig. 3 for the respective models used.

predictive distribution becomes a Gaussian mixture and we can compute the CRPS for each year of the validation period according to Eqs. (5) and (6) from Grimit et al. (2006). Finally these values are then averaged over the period and reported in Table 5.

Differences between our hierarchical model (hbm) and its alternatives (hbm w/o mdv, con, gcm, rcm and nbc) are usually small (cf. column 3 to columns 4–8 of

Table 5 and Table S1 of the supplementary material). For the Alps and Mediterranean region in summer, however, larger differences between the hbm approach and the one without multidecadal variations (hbm w/o mdv; compare the third and fourth column of Table 5) exist. Here, incorporating multidecadal variations leads to an improvement in predictive power. This gain in predictive power can be attributed to the underlying

FIG. 8. As in Fig. 6, but for the bias parameter κ and without plotted diamonds.

TABLE 5. Mean CRPS values ($^{\circ}\text{C}$) over the period 1991–2012 for different Bayesian models: our standard approach presented in this manuscript (hbm), an equivalent model without the inclusion of multidecadal variability (hbm w/o mdv), a Bayesian analysis of the constrained ensemble from E. Fischer et al. (2012b) (con), a Bayesian model that uses only the ENSEMBLES GCMs (gcm), a Bayesian model for the RCMs from Buser et al. (2010a) (rcm), and a reduction of the hbm that neglects bias changes (nbc).

Region	Season	CRPS					
		hbm	hbm w/o mdv	con	gcm	rcm	nbc
SC	Winter	1.32	1.31	1.31	1.31	1.28	1.29
	Summer	0.5	0.48	0.56	0.5	0.53	0.5
ME	Winter	0.91	0.91	0.91	0.91	0.91	0.91
	Summer	0.55	0.51	0.66	0.55	0.62	0.57
AL	Winter	0.65	0.66	0.63	0.65	0.65	0.65
	Summer	0.43	0.53	0.51	0.43	0.46	0.47
MD	Winter	0.59	0.57	0.61	0.59	0.59	0.6
	Summer	0.39	0.5	0.51	0.39	0.44	0.45

functional form of the observed time series (see Fig. 9 and Fig. S11 of the supplementary material). Since the inclusion of the basis functions $\psi_4, \dots, \psi_7$ allows for more flexibility in the mean, we are able to capture the S-shaped structure of the observations which extends into the validation period.

A shortcoming of this validation exercise is that it is based on short-term projection horizons only, whereas the real climatological interest is on larger time scales. Hence it is not meant to choose or discard some

statistical models for climate projections, but it rather serves as a sanity check for the hbm. However, we think that using the CRPS in combination with long historical simulations and observational time series could help in assessing probabilistic climate change projections from different statistical models.

7. Conclusions

We have implemented a Bayesian hierarchical statistical model that enables a coherent treatment of climate model results from heterogeneous RCM–GCM multimodel ensembles. Our approach is motivated by and builds upon findings of Buser et al. (2010a) and Kerkhoff et al. (2014).

Using our Bayesian hierarchical model, we derive climate change projections from the ENSEMBLES dataset. Our analyses confirm previous studies (Buser et al. 2010a) that find a robust increase in seasonal mean temperatures for all regions by the end of the twenty-first century. For interannual summer variability, however, our results, in contrast to other studies (Buser et al. 2010a), indicate also a distinct positive trend. In this way, our findings support conclusions drawn in E. Fischer et al. (2012), but without relying on expert information. The physical interpretation links this to soil moisture dynamics, and this reinforces the importance of the representation of land–atmosphere interactions for the simulation of interannual summer variability.

FIG. 9. Time series plots of the observations (black line) for 1961–2012 for (left)–(right) the different regions in (top) winter and (bottom) summer. Superimposed are mean functions for the Bayesian hierarchical model including multidecadal variations (red long dashes) and the same without multidecadal variability (gray short dashes). The red shaded area is a 95% credible interval computed using the full model including multidecadal variations. The gray shaded area highlights the period that was left out during fitting (1991–2012).

Furthermore, our results stress the importance of bias assumptions for the outcome of climate change studies. Many assessments neglect bias changes; however, our analysis favors the most general model that includes bias changes and a compromise between the “constant bias” and the “constant relation” assumption. Moreover, the anticipated climate change signal depends significantly on the chosen bias assumption, in particular in central and southern Europe during summer. Here, neglecting bias changes leads to weaker warming compared to an approach including bias changes.

Additionally, we investigate the role of internal decadal variability on the outcome of our analysis. Using an orthogonal spline basis, we are able to separate multidecadal variability from climate change signals. In this way, our approach is an elegant, fully Bayesian advancement of the work of A. Fischer et al. (2012). Assessing the out-of-sample predictive performance by calculating the continuous ranked probability score (CRPS; see Matheson and Winkler 1976), we show that the proposed hierarchical model is competitive and that the inclusion of multidecadal variability can improve near-term projections of climate in southern Europe.

Furthermore, our analysis implies important consequences for the design of future RCM–GCM multi-model ensembles. Because the gain in predictive power when going from independent simulations (exactly one GCM drives one RCM) toward the full data matrix (where sometimes many RCMs are driven by the same GCM; see, e.g., the ECHAM GCM in Table 1) is rather marginal (cf. the third and seventh column in Table 5), priority should be given to setting up many independent RCM–GCM simulations, using a large suite of driving GCMs with different RCMs.

Acknowledgments. This study was funded by the Swiss National Science Foundation under Grant 200021-132316/1. We acknowledge the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) for providing GCM and RCM simulations and the E-OBS observational datasets, as well as the data providers in the ECA&D project (<http://www.ecad.eu>). Constructive comments from two anonymous referees are also acknowledged.

APPENDIX

Construction of Basis Functions Used in Section 3c

The construction of the basis functions $\psi_1, \dots, \psi_7$ uses splines. For an introduction to splines see, for example, Hastie et al. (2009, section 5.2). A climate science application of splines is given in Scinocca et al. (2010). A

spline with knots at $s_0 < s_1 < \dots < s_{n+1}$ is a twice continuously differentiable function on $[s_0, s_{n+1}]$ that is equal to a cubic polynomial on each interval $[s_i, s_{i+1}]$. A natural spline is a spline where in addition the second derivative vanishes at the boundary knots s_0 and s_{n+1} . Natural splines with n interior knots can be written as linear combinations of $n + 2$ basis functions $\psi_1, \dots, \psi_{n+2}$. The larger the values of n , the more flexible the splines are, but the price to pay is an increased estimation error due to the increasing number of unknown parameters.

In section 3c, we use natural splines with equispaced knots at $s_j = 1961 + j \times 23$ to represent $\mu_{o,t}$ and $\mu_{g,t}$. To represent $\mu_{\text{clim},t}$ we use natural splines with only 1 interior knot at $s_3 = 2030$. As basis functions we choose a constant $\psi_1(t) = 1$, a linear function $\psi_2(t) = (t - 2030)/100$, and a natural cubic spline with one interior knot at $s_3 = 2030$ that is uncorrelated (orthogonal) with ψ_2 and satisfies $\psi_3(2030) = 0$. With these choices the coefficient of ψ_1 is the true mean in the year 2030 and the coefficient of ψ_2 is the increase per century.

Our choice of basis functions implies that $\mu_{\text{clim},t}$ is either convex or concave. If we wanted to allow $\mu_{\text{clim},t}$ to be also S-shaped, we would need more than one interior knot, but then separating the different scales becomes more delicate.

The remaining four basis function $\psi_4, \dots, \psi_7$ are used to represent multidecadal variations. They are chosen such that together $\psi_1, \dots, \psi_7$ are a basis for the natural splines with 7 knots at $s_0, \dots, s_6$ and $\psi_4, \dots, \psi_7$ are uncorrelated (orthogonal) with ψ_1, ψ_2 , and ψ_3 . The last property is needed to make the separation of the climate signal and the multidecadal variation unique.

We choose the scale of the functions ψ_k for $k > 2$ such that $\sum_{t=1}^{T_2} \psi_k^2(t) = 1$. Together with orthogonality, this implies that

$$\begin{aligned} \sum_{t=1}^{T_2} (v_{g,t})^2 &= \sum_{t=1}^{T_2} \sum_{j,k=4}^7 c_{k,g} c_{j,g} \psi_k(t) \psi_j(t) \\ &= \sum_{j,k=4}^7 c_{k,g} c_{j,g} \sum_{t=1}^{T_2} \psi_k(t) \psi_j(t) = \sum_{k=4}^7 c_{k,g}^2. \end{aligned} \quad (\text{A1})$$

REFERENCES

- Annan, J. D., and J. C. Hargreaves, 2010: Reliability of the CMIP3 ensemble. *Geophys. Res. Lett.*, **37**, L02703, doi:10.1029/2009GL041994.
- Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär, 2013: Physical constraints for temperature biases in climate models. *Geophys. Res. Lett.*, **40**, 4042–4047, doi:10.1002/grl.50737.
- Boberg, F., and J. H. Christensen, 2012: Overestimation of Mediterranean summer temperature projections due to model

- deficiencies. *Nat. Climate Change*, **2**, 433–436, doi:[10.1038/nclimate1454](#).
- Buser, C. M., H. R. Künsch, D. Lüthi, M. Wild, and C. Schär, 2009: Bayesian multi-model projection of climate: Bias assumptions and interannual variability. *Climate Dyn.*, **33**, 849–868, doi:[10.1007/s00382-009-0588-6](#).
- , —, and C. Schär, 2010a: Bayesian multi-model projections of climate: Generalization and application to ENSEMBLES results. *Climate Res.*, **44**, 227–241, doi:[10.3354/cr00895](#).
- , —, and A. Weber, 2010b: Biases and uncertainty in climate projections. *Scand. J. Stat.*, **37**, 179–199, doi:[10.1111/j.1467-9469.2009.00686.x](#).
- Chandler, R. E., 2013: Exploiting strength, discounting weakness: Combining information from multiple climate simulators. *Philos. Trans. Roy. Soc. London*, **371A**, 20120388, doi:[10.1098/rsta.2012.0388](#).
- Christensen, J. H., and O. B. Christensen, 2007: A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change*, **81**, 7–30, doi:[10.1007/s10584-006-9210-7](#).
- , and F. Boberg, 2012: Temperature dependent climate projection deficiencies in CMIP5 models. *Geophys. Res. Lett.*, **39**, L24705, doi:[10.1029/2012GL053650](#).
- Déqué, M., and S. Somot, 2010: Weighted frequency distributions express modelling uncertainties in the ENSEMBLES regional climate experiments. *Climate Res.*, **44**, 195–209, doi:[10.3354/cr00866](#).
- Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, doi:[10.1038/nclimate1562](#).
- Fischer, A. M., A. P. Weigel, C. M. Buser, R. Knutti, H. R. Künsch, M. A. Liniger, C. Schär, and C. Appenzeller, 2012: Climate change projections for Switzerland based on a Bayesian multi-model approach. *Int. J. Climatol.*, **32**, 2348–2371, doi:[10.1002/joc.3396](#).
- Fischer, E. M., J. Rajczak, and C. Schär, 2012: Changes in European summer temperature variability revisited. *Geophys. Res. Lett.*, **39**, L19702, doi:[10.1029/2012GL052730](#).
- Furrer, R., S. R. Sain, D. Nychka, and G. A. Meehl, 2007: Multivariate Bayesian analysis of atmosphere–ocean general circulation models. *Environ. Ecol. Stat.*, **14**, 249–266, doi:[10.1007/s10651-007-0018-z](#).
- , S. Geinitz, and S. R. Sain, 2012: Assessing variance components of general circulation model output fields. *Environmetrics*, **23**, 440–450, doi:[10.1002/env.2139](#).
- Geinitz, S., R. Furrer, and S. R. Sain, 2015: Bayesian multilevel analysis of variance for relative comparison across sources of global climate model variability. *Int. J. Climatol.*, **35**, 433–443, doi:[10.1002/joc.3991](#).
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013: *Bayesian Data Analysis*. 3rd ed. CRC Press, 675 pp.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, 1996: *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 512 pp.
- Giorgi, F., and L. O. Mearns, 2002: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *J. Climate*, **15**, 1141–1158, doi:[10.1175/1520-0442\(2002\)015<1141:COAURA>2.0.CO;2](#).
- , C. Jones, and G. R. Asrar, 2009: Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bull.*, **58**, 175–183.
- Greene, A. M., L. Goddard, and U. Lall, 2006: Probabilistic multimodel regional temperature change projections. *J. Climate*, **19**, 4326–4343, doi:[10.1175/JCLI3864.1](#).
- Grimit, E. P., T. Gneiting, V. J. Berrocal, and N. A. Johnson, 2006: The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quart. J. Roy. Meteor. Soc.*, **132**, 2925–2942, doi:[10.1256/qj.05.235](#).
- Hastie, T., R. Tibshirani, and J. Friedman, 2009: *The Elements of Statistical Learning—Data Mining, Inference and Prediction*. 2nd ed. Springer-Verlag, 745 pp.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, doi:[10.1175/2009BAMS2607.1](#).
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, doi:[10.1029/2008JD010201](#).
- Heaton, M., T. Greasby, and S. R. Sain, 2013: Modeling uncertainty in climate using ensembles of regional and global climate models and multiple observation-based data sets. *SIAM/ASA J. Uncertainty Quantif.*, **1** (1), 535–559, doi:[10.1137/12088505X](#).
- Hewitt, C. D., and D. J. Griggs, 2004: Ensembles-based predictions of climate changes and their impacts. *Eos, Trans. Amer. Geophys. Union*, **85**, 566, doi:[10.1029/2004EO520005](#).
- Ho, C. K., D. B. Stephenson, M. Collins, C. A. T. Ferro, and S. J. Brown, 2012: Calibration strategies: A source of additional uncertainty in climate change projections. *Bull. Amer. Meteor. Soc.*, **93**, 21–26, doi:[10.1175/2011BAMS3110.1](#).
- IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., doi:[10.1017/CBO9781107415324](#).
- Jun, M., R. Knutti, and D. W. Nychka, 2008: Spatial analysis to quantify numerical model bias and dependence. *J. Amer. Stat. Assoc.*, **103**, 934–947, doi:[10.1198/016214507000001265](#).
- Kang, E. L., N. Cressie, and S. R. Sain, 2012: Combining outputs from the North American Regional Climate Change Assessment Program by using a Bayesian hierarchical model. *J. Roy. Stat. Soc.*, **61C**, 291–313, doi:[10.1111/j.1467-9876.2011.01010.x](#).
- Kaufman, C. G., and S. R. Sain, 2010: Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Anal.*, **5**, 123–149, doi:[10.1214/10-BA505](#).
- Kerkhoff, C., H. R. Künsch, and C. Schär, 2014: Assessment of bias assumptions for climate models. *J. Climate*, **27**, 6799–6818, doi:[10.1175/JCLI-D-13-00716.1](#).
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in combining projections from multiple climate models. *J. Climate*, **23**, 2739–2758, doi:[10.1175/2009JCLI3361.1](#).
- Laprise, R., 2014: Comment on “The added value to global model projections of climate change by dynamical downscaling: A case study over the continental U.S. using the GISS-ModelE2 and WRF models” by Racherla et al. *J. Geophys. Res. Atmos.*, **119**, 3877–3881, doi:[10.1002/2013JD019945](#).
- Lenderink, G., 2010: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. *Climate Res.*, **44**, 151–166, doi:[10.3354/cr00946](#).
- Maraun, D., 2012: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophys. Res. Lett.*, **39**, L06706, doi:[10.1029/2012GL051210](#).

- Masson, D., and R. Knutti, 2011: Climate model genealogy. *Geophys. Res. Lett.*, **38**, L08703, doi:[10.1029/2011GL046864](https://doi.org/10.1029/2011GL046864).
- Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. *Manage. Sci.*, **22**, 1087–1096, doi:[10.1287/mnsc.22.10.1087](https://doi.org/10.1287/mnsc.22.10.1087).
- Mearns, L. O., W. J. Gutowski, R. Jones, L.-Y. Leung, S. McGinnis, A. M. B. Nunes, and Y. Qian, 2009: A regional climate change assessment program for North America. *Eos, Trans. Amer. Geophys. Union*, **90**, 311, doi:[10.1029/2009EO360002](https://doi.org/10.1029/2009EO360002).
- Mueller, B., and S. I. Seneviratne, 2012: Hot days induced by precipitation deficits at the global scale. *Proc. Natl. Acad. Sci. USA*, **109**, 12 398–12 403, doi:[10.1073/pnas.1204330109](https://doi.org/10.1073/pnas.1204330109).
- Nakicenovic, N., and R. Swart, Eds., 2000: *Special Report on Emission Scenarios*. Cambridge University Press, 570 pp.
- Nychka, D., and C. Tebaldi, 2003: Comments on “Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the ‘reliability ensemble averaging’ (REA) method.” *J. Climate*, **16**, 883–884, doi:[10.1175/1520-0442\(2003\)016<0883:COCOAU>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0883:COCOAU>2.0.CO;2).
- Piani, C., J. O. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, doi:[10.1007/s00704-009-0134-9](https://doi.org/10.1007/s00704-009-0134-9).
- Rougier, J., M. Goldstein, and L. House, 2013: Second-order exchangeability analysis for multimodel ensembles. *J. Amer. Stat. Assoc.*, **108**, 852–863, doi:[10.1080/01621459.2013.802963](https://doi.org/10.1080/01621459.2013.802963).
- Sain, S. R., R. Furrer, and N. Cressie, 2011: A spatial analysis of multivariate output from regional climate models. *Ann. Appl. Stat.*, **5**, 150–175, doi:[10.1214/10-AOAS369](https://doi.org/10.1214/10-AOAS369).
- Salazar, E., B. Sansó, A. O. Finley, D. Hammerling, I. Steinsland, X. Wang, and P. Delamater, 2011: Comparing and blending regional climate model predictions for the American Southwest. *J. Agric. Biol. Environ. Stat.*, **16**, 586–605, doi:[10.1007/s13253-011-0074-6](https://doi.org/10.1007/s13253-011-0074-6).
- Sanderson, B. M., and R. Knutti, 2012: On the interpretation of constrained climate model ensembles. *Geophys. Res. Lett.*, **39**, L16708, doi:[10.1029/2012GL052665](https://doi.org/10.1029/2012GL052665).
- Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability for European summer heat waves. *Nature*, **427**, 332–336, doi:[10.1038/nature02300](https://doi.org/10.1038/nature02300).
- Scinocca, J. F., D. B. Stephenson, T. C. Bailey, and J. Austin, 2010: Estimates of past and future ozone trends from multimodel simulations using a flexible smoothing spline methodology. *J. Geophys. Res.*, **115**, D00M12, doi:[10.1029/2009JD013622](https://doi.org/10.1029/2009JD013622).
- Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land-atmosphere coupling and climate change in Europe. *Nature*, **443**, 205–209, doi:[10.1038/nature05095](https://doi.org/10.1038/nature05095).
- Smith, R. L., C. Tebaldi, D. Nychka, and L. O. Mearns, 2009: Bayesian modeling of uncertainty in ensembles of climate models. *J. Amer. Stat. Assoc.*, **104**, 97–116, doi:[10.1198/jasa.2009.0007](https://doi.org/10.1198/jasa.2009.0007).
- Stephenson, D. B., M. Collins, J. C. Rougier, and R. E. Chandler, 2012: Statistical problems in the probabilistic prediction of climate change. *Environmetrics*, **23**, 364–372, doi:[10.1002/env.2153](https://doi.org/10.1002/env.2153).
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1).
- Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc. London*, **365A**, 2053–2075, doi:[10.1098/rsta.2007.2076](https://doi.org/10.1098/rsta.2007.2076).
- , and B. Sansó, 2009: Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *J. Roy. Stat. Soc.*, **172A**, 83–106, doi:[10.1111/j.1467-985X.2008.00545.x](https://doi.org/10.1111/j.1467-985X.2008.00545.x).
- , R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *J. Climate*, **18**, 1524–1540, doi:[10.1175/JCLI3363.1](https://doi.org/10.1175/JCLI3363.1).
- van der Linden, P., and J. F. B. Mitchell, 2009: ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre Tech. Rep., 160 pp.
- Vidale, P. L., D. Lüthi, R. Wegmann, and C. Schär, 2007: European summer climate variability in a heterogeneous multi-model ensemble. *Climatic Change*, **81**, 209–232, doi:[10.1007/s10584-006-9218-z](https://doi.org/10.1007/s10584-006-9218-z).
- Weigel, A. P., R. Knutti, M. A. Liniger, and C. Appenzeller, 2010: Risks of model weighting in multimodel climate projections. *J. Climate*, **23**, 4175–4191, doi:[10.1175/2010JCLI3594.1](https://doi.org/10.1175/2010JCLI3594.1).