Joint Projections of Temperature and Precipitation Change from Multiple Climate Models: A Hierarchical Bayesian Approach

Claudia Tebaldi†

Climate Central, Princeton, NJ and Department of Global Ecology, Carnegie Institution of Washington, Stanford, CA, U.S.A.
`ctebaldi@globalecology.stanford.edu`

Bruno Sansó

Department of Applied Mathematics and Statistics, University of California, Santa Cruz, U.S.A.
`bruno@ams.ucsc.edu`, `www.ams.ucsc.edu/~bruno`

Summary. Posterior distributions for the joint projections of future temperature and precipitation trends and changes are derived by applying a Bayesian hierachical model to a rich dataset of simulated climate from General Circulation Models. The simulations here analysed constitute the future projections on which the Intergovernmental Panel of Climate Change based its recent summary report on the future of our planet's climate, albeit without any sophisticated statistical handling of the data. Here we quantify the uncertainty represented by the variable results of the different models and their limited ability to represent the observed climate both at global and regional scales. We do so in a Bayesian framework, by estimating posterior distributions of the climate change signals in terms of trends or differences between future and current periods, while we fully characterize the uncertain nature of a suite of other parameters, like biases, correlation terms and model-specific precisions. Besides presenting our results in terms of posterior distributions of the climate signals, we offer as an alternative representation of the uncertainties in climate change projections the use of the posterior-predictive distribution of a new model's projections. The results from our analysis can find straightforward applications in impact studies, which necessitate not only best guesses but a full representation of the uncertainty in climate change projections. For water resource and crop models, for example, it is vital to use joint projections of temperature and precipitation in order to best represent the characteristics of future climate, and our statistical analysis delivers just that.

Keywords: Climate change; Temperature and Precipitation Change; Multi-model ensembles; Probabilistic projections; IPCC AR4

†to whom all correspondence should be addressed

1. Introduction

The latest estimates based on observed records indicate that the global average annual temperature of the Earth's surface has increased by as much as 0.7 degrees Celsius since the late 19th century, with most of the warming observed in the last 50 years. The scientific consensus, recently summarized by the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC-AR4) (IPCC, 2007) is that a significant portion of this increase, and most of the warming in the last 50 years, is very likely attributable to human activities, most importantly the burning of fossil fuels, cause of a steady increase in the concentrations of greenhouse gases (GHGs) in the atmosphere. The same scientific consensus, made of hundreds of climate scientists from around the world, has vetted an ever-increasing body of peer-reviewed literature based mainly on climate model experiments. These studies unequivocally predict even larger future warming, in the absence of significant curbing of GHG emissions, with 2 degrees Celsius of additional warming being generally viewed as a lower bound estimate of the expected global change at the end of the 21st century, under so called "business as usual" emission scenarios. Global average increases in temperature are however just projections from a marginal aspect of multi-faceted regional climatic changes that are very likely to affect societies and ecosystems in adverse more than beneficial ways. If high latitude regions may see warmer temperatures as benefitial, enjoying longer growing seasons and lower heating costs, most of the world will see less positive changes, in all likelihood: climate models and scientific understanding of climate processes project changes in variables other than average temperature, like precipitation and wind, their patterns and intensities. More frequent and intense extreme events are anticipated, and the possibility is not ruled out of exceeding systemic thresholds for climate variables causing abrupt changes, with potentially much more dangerous consequences than simple gradual shifts. Associated with climatic changes, economic, biological and health-related impacts are then to be expected, ranging from changes in agricultural yields and forced abandonment and geographical shifts of crops, to disruption of ecosystems, from depletion of water resources to easier spread of vector-borne diseases.

The recent scientific and political debates have generally settled over the qualitative nature of the problem, but a quantification of the risks associated with it is far from straightforward. There exist considerable uncertainties, in terms of limits to predicting future changes in the drivers of GHG emissions (population, economic growth, technology, international collaboration, etc.), of incomplete scientific understanding of the response of the climate system to those emissions, and of intrinsic randomness, or unpredictability, of the processes involved. From the size of the changes to be expected to their timing;

from the significance of their impacts to the costs of adaptation and mitigation strategies; every link of the chain from GHG emissions to impacts to the effects of response strategies is affected by uncertainties.

In this paper we tackle a particular aspect of these uncertainties, namely the modeling uncertainties affecting, under a specified scenario of future GHG emissions, the size of expected regional changes in temperature and precipitation. Our goal is to characterize these uncertainties through a formal statistical treatment of the observed data and the ensemble of model output on which future projections are based, and produce bivariate probability distribution functions of future regional climate changes with respect to a climatological baseline, as, for example, contours of area-averaged temperature and percent precipitation change at a given time in the course of the 21st Century, for a specific region and season. Probabilistic projections of climate change at global and regional scales are a developing area of research, of late (Collins and Knight, 2007). Within it, one particular approach seeks to take advantage of ensembles of global climate model experiments, in particular those that have been occasioned and facilitated by the international activities of the IPCC. The same approach we take here, building on the already published work in Tebaldi et al. (2004, 2005); Smith et al. (2008). In those analyses univariate PDFs of average temperature *or* precipitation change at regional and seasonal scales were produced, based on multidecadal averages of the two climate variables for current conditions (usually 1961-1990 or 1980-1999 averages of observed and model simulated data) and future (2080-2099 averages of model output). Here we further that approach in two main directions. We choose to model temperature *and* precipitation jointly, and we model the entire length of the observed and simulated time series, thus explicitly quantifying trends in the data. Besides the interest in a more complete representation of the uncertainty for the entire time horizon at hand, we strive to provide this type of results for the sake of impact studies (e.g., future changes in crop yields or water resource management modeling), which usually need future climate scenarios involving both variables jointly and for which trends over time, rather than static snapshots of change may be of value. Section 2 introduces the statistical approaches that constitute a background for our analysis, and describes our data. Section 3 presents the statistical model for joint temperature and precipitation trends. Section 4 applies the method to a specific set of observations and simulations by general circulation models (GCMs) archived by the Program in Climate Model Diagnosis and Intercomparison (PCMDI; `http://www-pcmdi.llnl.gov`) and used in the IPCC-AR4. In Section 5 we summarize the most important aspects of our method and its applications and highlight some general issues that will require further treatment, if an improved approach to this problem is to be pursued.

2.   Combining multi-model ensembles

2.1.   Relevant published methods

It has been less than ten years since concerted efforts among climate modeling centers started to make available results from standardized, and thus comparable, experiments. With the IPCC activities, such multi-model data repositories have grown in size both as the number of modeling centers developing GCMs increased and as the range of variables they submit to the common archive becomes larger. There are two main approaches at combining multi-model ensemble output. One that simply considers each model as equal, and produces simple ensemble averages and measures of inter-model variability like standard deviations and ranges. The other, formalized by several published methods in different ways (Greene et al., 2006; Furrer et al., 2007), stems from the belief that not all models are to be trusted equally, but some are better than others and should receive more weight in the combination of the results. The REA (Reliability Ensemble Average) approach in Giorgi and Mearns (2002) was the first published method that tried to quantify this belief by designing – albeit rather arbitrarily – weights so that models characterized by small bias and projections that agree with the ensemble "consensus" were rewarded while models that perform poorly in replicating observed climate and that appear as outliers were discounted. The REA approach motivated the work in Tebaldi et al. (2004, 2005) and Smith et al. (2008). Here we give a brief overview of the statistical modeling in these papers, since we consider our present proposal a natural extension of their fundamental paradigm. The Bayesian analysis, summarized and discussed for a statisticians' audience in Smith et al. (2008), treats the unknown quantities of interest (current and future climate *signals*, model precisions) as random variables, for which reference prior distributions are chosen. The likelihood assumptions cause the form of the final posterior estimates of temperature change (or, separately, precipitation change) to be weighted average of the individual models' projections. Differently from the heuristic approach in Giorgi and Mearns (2002), however, the weights' formula is a direct result of the probabilistic assumptions made explicit in the likelihood – and prior – choices. Gaussian distributions are stipulated for the current (X_i's) and future (Y_i's) projections from model i in a given region and season, centered around the true climate signals, μ and ν respectively, with model-specific variances, λ_i^{-1}, and a multiplicative parameter for the future precision that accounts for the extra uncertainty of future projections. Thus,

$$
\begin{aligned}
X_i &\sim N(\mu, \lambda_i^{-1}) \\
Y_i &\sim N(\nu, (\theta\lambda_i)^{-1}).
\end{aligned}
$$

Similarly, the observed current climate, X_0 is modelled as a realization from a Gaussian distribution centered around the current climate signal μ, and whose variance is estimated through the observed record, yielding

$$X_0 \sim N(\mu, \lambda_0^{-1}). \tag{1}$$

Through Bayes theorem, evaluated numerically by Markov Chain Monte Carlo (MCMC) methods, a posterior distribution for the true climate signals is derived, and straightforwardly translated into a probability distribution for climate change, defined as $\nu - \mu$. As a consequence of the distributional assumptions the criteria of bias and convergence, in an analytical form similar to the form of the REA weights in Giorgi and Mearns (2002), shape the posterior distributions. In fact, conditionally on $\lambda_1, \ldots, \lambda_M$, the posterior means for μ and ν are

$$\widetilde{\mu} = \left(\sum_i \lambda_i X_i \right) \Big/ \left(\sum_i \lambda_i \right)$$

and

$$\widetilde{\nu} = \left(\sum_i \lambda_i Y_i \right) \Big/ \left(\sum_i \lambda_i \right),$$

where the posterior mean of the model-specific λ_i's is approximately

$$\widetilde{\lambda_i} \approx \frac{a+1}{b + \frac{1}{2}[(X_i - \widetilde{\mu})^2 + \theta(Y_i - \widetilde{\nu})^2]}. \tag{2}$$

The first term in the denominator in equation (2) is a measure of bias, being the distance of the present climate average as simulated by model i from the optimal estimate, $\widetilde{\mu}$, of current climate. The second term is a measure of convergence, computing a distance between the model's future projection and the future climate's posterior mean ($\widetilde{\nu}$). The terms a and b in (2) are the parameters of the common prior for all λ_i's, a modeling choice that ensures a balanced assignment of weight across GCMs. They are in turn modeled as random variables with their own (hyper-) prior distributions.

2.2. Data

In Tebaldi et al. (2004, 2005) and Smith et al. (2008) the data consist of observed and modeled quantities, averaged over regions and seasons. An estimate of the natural variability of the observed averages is used to fix the value of the precision component λ_0 in the likelihood of the observed data (1). Modeled multi-decadal averages of current and future temperature or precipitation, as

simulated by a number of GCMs, are then considered. Current climate simulations are usually labeled "all-forcings" runs, because modelers impose the best approximation to the historic external forcings, like solar cycle estimates, volcano eruptions (known to have a short-lived cooling effect on the global temperature) and anthropogenic emissions of GHGs and sulfate aerosols leading to the observed concentrations in the atmosphere. When future climate modeling is concerned, standard scenarios of future emissions are adopted in concert and run by the different modeling centers. The scenarios vary in their assumptions of GHGs production and they usually span a range from low to mid to high emissions. Our analysis will be conducted conditional on a specific scenario, so we are not concerned with the uncertainty pertaining to future economic growth and social development that may lead to the alternative scenarios.

Differently from Tebaldi et al. (2004, 2005) and Smith et al. (2008), we work with time series of decadal averages, i.e., $\tau_0 = 6$ decades of observed temperature and precipitation records, and $\tau^* = 15$ decades of simulated temperature and precipitation model output. The number of models contributing output to the archive varies between 15, for the high-emissions scenario, and 18, for low and mid emissions. Their output has been registered to a common grid, whose horizontal resolution is on the order of 250 Km. along the longitude and latitude directions. For our analysis we average the grid point output within the regions of interest. We then analyse one region and one season at a time.

Let $\boldsymbol{O}_t$, $t = 1, \ldots 6$ a two-component vector of observed temperature and log-precipitation averaged over a given region and a given season for all the years of a given decade. The time index, t, corresponds to the decades centered at $1955, 1965, \ldots, 2005$. Let $\boldsymbol{X}_{jt}$, $t = 1, \ldots, 15$ be the vector similar temperature and log-precipitation averages (to handle the positive domain and the skewed distribution of precipitation) derived from the j-th GCM output. Here the time index corresponds to the decades centered at $1955, 1965, \ldots, 2005, 2015, \ldots, 2095$, so that both historical and future periods are considered.

3.   A joint model for temperature and precipitation

Our fundamental approach is that of Bayesian hierarchical models. The basic assumptions are the following:

- The vector of observed values $\boldsymbol{O}_t$ is a noisy version of the underlying temperature and precipiatation process, with correlated Gaussian noise.

- The true process is piecewise linear, for both temperature and precipitation. We hypothesize an "elbow" at year 2000, which accomodates our expectation that future trends will be steeper than the observed ones. This

expectation is informed by the behavior of model simulations of, where the increasing future GHG concentrations enhance the rate of change in the computer-generated time series. A more sophisticated model could use a random change point approach, but given the coarse resolution of our time dimension and the limited amount of data at our disposal we choose to fix the position of the elbow here.

- The model output $\boldsymbol{X}_{jt}$ is a biased and noisy version of the truth. We assume an additive bias and a bivariate Gaussian noise.

- We expect the model biases to be related across the population of models Our approach provides an estimate of the overall bias for the ensemble of model simulations.

To formulate a statistical model we use superscripts T and P to refer to the temperature and precipitation components of the different vectors. Thus,

$$
\begin{aligned}
O_t^T &\sim N[\mu_t^T; \eta^T] \qquad \text{for } t = 1, \ldots, \tau_0 \\
O_t^P &\sim N[\mu_t^P + \beta_{xo}(O_t^T - \mu_t^T); \eta^P] \qquad \text{for } t = 1, \ldots, \tau_0 \\
\text{where } \beta_{xo} &\sim N[\beta_0, \lambda_o], \\
X_{jt}^T &\sim N[\mu_t^T + d_j^T; \xi_j^T] \qquad \text{for } t = 1, \ldots, \tau^* \text{ and } j = 1, \ldots, M \\
X_{jt}^P &\sim N[\mu_t^P + \beta_{xj}(X_{jt}^T - \mu_t^T - d_j^T) + d_j^P; \xi_j^P] \\
&\qquad\qquad \text{for } t = 1, \ldots, \tau^* \text{ and } j = 1, \ldots, M.
\end{aligned} \tag{3}
$$

Here all distributional assumptions for the data are conditional on the quantities appearing in the right-hand side of the $\sim$ symbol. In Equations (3) we specify bivariate normal distributions for $\boldsymbol{O}_t$ and $\boldsymbol{X}_{jt}$ using conditionality. After accounting for the underlying trends and biases terms, $\beta_{xo}, \beta_{x1}, \ldots, \beta_{xM}$ are used to model the correlation between temperature and precipitation. We assume that all the parameters are random with the exception of β_{xo}, η^T abd η^P, which we estimate on the basis of the observed records.

We assume that the time evolution of *true climate process* $\boldsymbol{\mu}_t' = (\mu_t^T, \mu_t^P)$, consists of a piecewise linear trend in both components, so

$$
\begin{pmatrix} \mu_t^T \\ \mu_t^P \end{pmatrix} \equiv \begin{pmatrix} \alpha^T + \beta^T t + \gamma^T (t - \tau_0)\mathcal{I}_{\{t \geq \tau_0\}} \\ \alpha^P + \beta^P t + \gamma^P (t - \tau_0)\mathcal{I}_{\{t \geq \tau_0\}} \end{pmatrix}. \tag{4}
$$

In Equation (4) we account for the fact that we expect the future trend over the period 2000–2100 to change slope. Since we expect the trend for temperature to be steeper in the future, one could go as far as to hypothesize that $\gamma^T > 0$. This could be incorporated in the prior for this parameter. In this paper we adopt a non-informative prior.

The priors for the parameters in Model (3) are specified hierarchically by assuming that $\beta_{xj} \sim N[\beta_0, \lambda_B]$, $d_j^T \sim N[a^T; \lambda_D^T]$, $d_j^P \sim N[a^P; \lambda_D^P]$ for $j = 1, \ldots, M$ $\xi_j^T \sim G[a_{\xi^T}, b_{\xi^T}]$ and $\xi_j^P \sim G[a_{\xi^P}, b_{\xi^P}]$. λ_o is fixed to a value estimated on the basis of the observed record. All the other quantities in this priors are assigned vaguely informative priors $\beta_0, a^T, a^P \sim U[-\infty, +\infty]$ and $\lambda_B, \lambda_D^T, \lambda_D^P, a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P} \sim G[g, h]$, where $g = h = 0.01$. Similarly, for the parameters in (4), we assume $\alpha^T, \beta^T, \gamma^T, \alpha^P, \beta^P, \gamma^P \sim U[-\infty, +\infty]$.

Notice that the correlation coefficients, β_{xo} and β_{xj}, have a common mean, β_0, and the biases, d_j^T and d_j^P, have prior means, a^T and a^P possibly different from zero. The likelihood and priors form a conjugate model, and a Gibbs sampler can be programmed to explore the posterior and predictive distributions for this model, with the only complication of a Metropolis-Hastings step which is used to generate sample values for $a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P}$. In an appendix we describe the full conditional distributions of all the random quantities in our model, and thus the implementation of the Markov Chain Monte Carlo algorithm.

We are assuming that each model has its own precision in simulating the true temperature and precipitation time series, but we impose common priors to ξ_j^T and ξ_j^P $\forall j$, whose parameters are in turn estimated by the data. This has been shown in Smith et al. (2008) to produce robust estimates of the relative precisions of the different GCMs, not overly sensitive to small perturbations in the GCM's trajectories. It also provides a data-driven sampling distribution for a new GCM, allowing us to perform cross-validatory exercises and to offer the predictive distribution of a new bivariate time series $\{\boldsymbol{X}_{*t}\}$ as a probabilistic quantification of the uncertainty in future projections. We will discuss this in detail in Section 4.

The model-specific bias terms d_j^T, d_j^P are assumed constant over the length of the simulation. They model systematic errors in each GCM simulated variables. All the GCM biases for temperature, like all GCM biases for precipitation, are realization from a common Gaussian distribution, whose mean (a^T, or a^P) may be different from zero, when the set of model trajectories is distributed around the truth non-symmetrically. We do not expect a systematic behavior across models when it comes to precipitation versus temperature biases, that is, we do not expect that models having relatively larger temperature biases would show relatively larger precipitation biases, so we do not model a correlation structure between d_j^T, d_j^P. In fact, this correlation structure, if there at all, would not to be identifiable/separable from the correlation modeled through $\beta_{xo}, \beta_{x1}, \ldots, \beta_{xM}$, given the configuration of the present dataset.

All the remaining parameters of the model have non-informative, conjugate distributions. Notice that we use improper priors for the location parameters of the Gaussian distributions and linear regression parameters in the correlation structure and in the trend structure, and proper but diffuse priors for the

precision parameters and as hyper-priors of the ξ. parameters.

4.   Current and future climate trends: a regionally aggregated perspective under a mid-range emission scenario

We apply the statistical model described in Section 3 to a set of climate model simulations prepared for the IPCC-AR4. We present results at the global average scale (land only, from now on indicated as GLOB) and at two different regional scales. Starting with the work in Giorgi and Mearns (2002) regional assessment of climate change has been presented on standard subcontinental regions, that have become known as "Giorgi regions". These regions are very large, on the order of 10^7Km^2 and the usefulness of treating jointly temperature and precipitation at this level of aggregation is questionable. Nevertheless, for consistency with the main strain of analyses already published, we adopt one of them here, the region of Western North America (WNA), defined as the land area in the large solid box in Figure 1. But since nothing prevents us from applying the same statistical model to smaller or larger regions, and in particular to region tailored to specific impact analysis, where projections of joint temperature and precipitation change could be used to drive impacts models (e.g., crop models, as in Lobell and Field (2007) or water resource management models, as in Groves et al. (2008)), we are also going to define a subset of this region over California (CAL) shown as the smaller dashed box in Figure 1, and run the same analysis at this finer scale.

We will produce two kinds of probabilistic projections. One consists of the bivariate *posterior* distribution of the joint change in temperature and precipitation signal, which will be derived as a marginal projection from the joint posterior distribution of $\boldsymbol{\mu_t}$ by computing the difference between time averages from two periods. We choose the last 20 years of the 20th century as representative of current climate, and the last 20 years of the 21st century as our future, and we indicate the two-dimensional random vector by $(\Delta T, \Delta P)$. We expect the posterior distribution of temperature and precipitation change to have a much smaller width than the range of the individual GCM projections, since it represents the estimate of the models' central tendency, and its uncertainty. The width of the posterior distributions is an inverse function of the number of GCMs used to estimate it, consistently with the assumption of independence among the individual GCMs. However, it is a widespread belief in the climate change community, as we discuss in Section 5, that the individual model projections are to be regarded as spanning the actual range of uncertainty of what is expected as future climate. In a way, such perspective invites us to present the posterior-*predictive* distribution of a new GCM projection as an alternative representation of this uncertainty. The width of the predictive distribution of
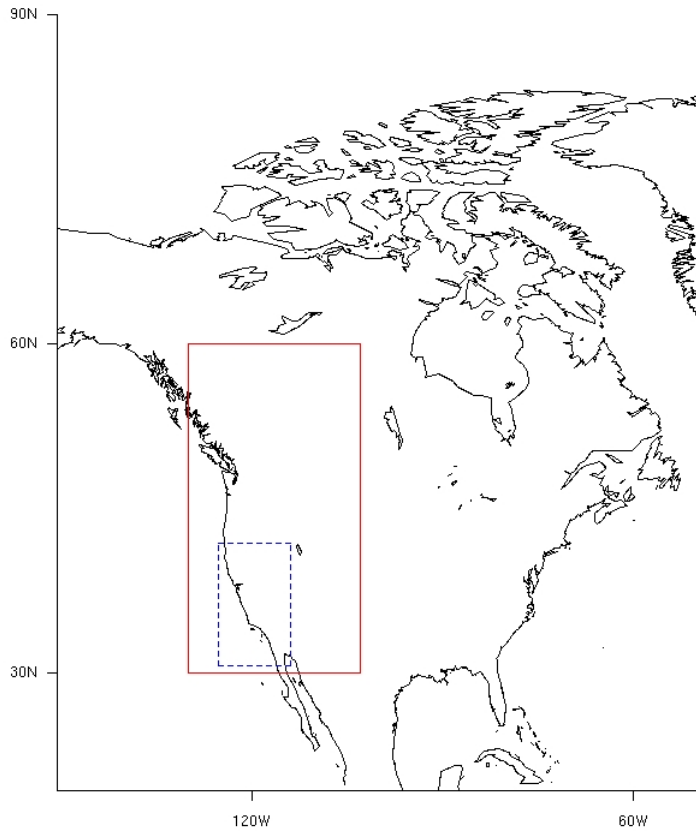
Fig. 1. The two regional areas chosen to exemplify the results our statistical model. The larger box delimits the WNA region while the smaller box delimits the CAL. Note that only land points within this geographical areas are averaged. We estimate the parameters of our model separately for the regions WNA, CAL and GLOB, repeating the analysis for Boreal winter (DJF) seasonal averages and Boreal summer (JJA) seasonal averages.

the quantity $(\Delta X_*^T, \Delta X_*^P)$, if the statistical model is consistent with the data, will be similar to the range of GCM projections and will not depend on the number of GCMs included in the analysis.

Other aspects that will be explored for this dataset are: the significance of the trend in the temperature and precipitation, established by analyzing the posterior distribution of β^T and β^P; the significance of the change in the trend at 2000, which can be assessed by considering the posterior distributions of γ^T and γ^P; comparison of the posterior distribution of the correlation coefficient β_0 to the observed value β_{xo}, in order to gauge the overall consistency of the ensemble in simulating the joint behavior of temperature and precipitation; evaluation of the posterior distribution of the parameters a^T and a^P to assess the value of the systematic bias common to the GCMs; evaluation of the posterior distributions of the individual parameters d_j^T, d_j^P to quantify model-specific biases. Finally, we should be concerned with testing the goodness of fit of our model. We do so by leave-one-out cross validation, thereby fitting the model on all but one GCM's projections, calculating the predictive distribution of the left-out $(\Delta X_*^T = \Delta x_*^T, \Delta X_*^P = \Delta x_*^P)$, and validating the bivariate distribution by testing that $P(\Delta X_*^P <= \Delta x^P | \Delta X_*^T = \Delta x^T)$ is distributed as a uniform distribution, independent of $P(\Delta X_*^T = \Delta x^T)$ (Rosenblatt, 1952).

Figures 2 through 7 represent examples of the results of our estimates for GLOB, WNA and CAL and two seasons, December-January-February (DJF) and June-July-August (JJA). The six panels of each figure show bivariate and univariate distributions and trends estimates for average temperature and precipitation in a region/season combination. Let us focus at first on the wide contour lines of each figure, and the corresponding univariate PDFs, representing marginals of the predictive distribution for a new GCM projection. The two sets of contours in the first panel describe the predictive distribution of average current and future climate. In all cases, the contours shift unambiguously to the right because of a significant increase of the average temperatures. In some of the six examples the contours also shift upward, signaling a significant increase in precipitation. In the second panel the difference in current and future climate is described by a single set of contours. The shape of these contours are suggesting a negative correlation only in the case of WNA and CAL in JJA, otherwise the shape is not appreciably different from spherical, indicating absence of a significant correlation at this level of regional and seasonal aggregation. The dots on these pictures show the individual GCMs' projections contributing to the estimation (18 of them). They are consistent with the large contours and the marginal PDFs by eye-inspection.

Our model is fitting the entire length of the GCM trajectories as they simulate temperature and precipitation along the 15 decades centered at 1955, 1965, and so on up to 2095. The two panels in the center row show in thin lines the
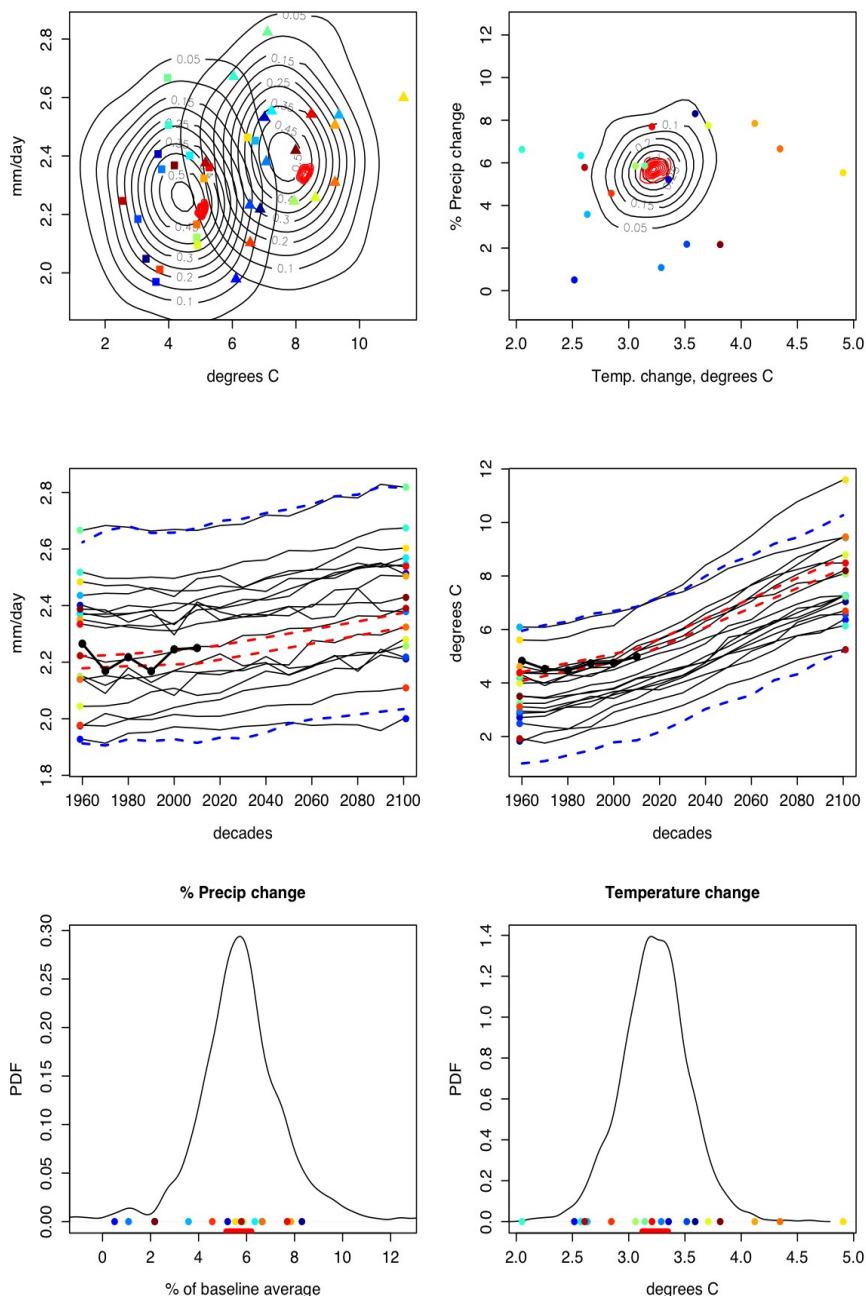
Fig. 2. Global average (GLOB) temperature and precipitation projections in DJF. Top left: distributions of current and future temperature and precipitation, and models' projections for the same periods (one color per model). Wide contours correspond to the predictive distributions of a new GCM's projections, the much tighter small contours correspond to the posterior distributions of the climate signal. Top right: Analogous plot for temperature and percent precipitation *changes*. Center left: Trajectories of average precipitation from the models (thin lines) and observations (thick line with marks), 95% posterior probability intervals (tight pair of dashed lines) and 95% predictive intervals (wider

GCM trajectories for precipitation (left panel) and temperature (right panel) together with the observed time series in a thick line joining 6 dots (corresponding to the six decades centered at 1955, 1965 and so on up to 2005, the last being estimated only on the basis of data up to 2006). Two sets of dashed lines draw the 95% probability intervals of the posterior and the predictive distribution for – respectively – the trends and a new GCM's trajectories. One can appreciate how the posterior distribution heavily weights the observations to position the absolute values of the climate signals' trajectories, and uses the models only to estimate the slopes of the piecewise linear process. This helps explaining the position of the two small contours in the first panel of the figures They are the contours of the posterior distributions of the bivariate climate signal, for current and future periods, which in some cases results significantly misaligned with the consensus of the models' projections, and thus the mode of the predictive distribution. In these instances, the statistical model has estimated a large systematic bias in the ensemble. Notice though that even in the cases when that is true, the misalignment of the modes of posterior and predictive disappears in the contours of the climate *change* distribution. This result gives support to the practice, popular among climate researcher, to always consider the difference projected by each model, not the absolute value of their current and future projections, according to the belief that the systematic bias would thus cancel out.

Figure 8 answers the questions about the significance of the trends in the climate signal time series. Recall that we model the trends as piecewise linear, i.e., simplifying the notation of Section 3, as $\beta t + \gamma(t - \tau_0)\mathcal{I}_{\{t \geq \tau_0\}}$. The three panels in the left column show posterior distributions (as boxplots) of the coefficients of temperature trends in the three regions. Three boxplots for each season are shown side by side. The first three boxplots in each panel describe distributions of β, γ and their sum, for DJF temperature, thus representing the current period trend, its increment (or decrement) at the beginning of the future period, and the future trend. In all cases the boxplots are significantly above the zero line, leading to the conclusion that all the trends are positive, significant and they increase significantly in the future.

In the case of precipitation trends the conclusions are mixed. For the GLOB region in DJF all three coefficients are positive and significantly different from zero, suggesting increasing trends in the historic global averages, and increasing trend, significantly larger, in the future. In JJA the trend becomes significantly positive only in the future. For WNA there is a positive and significant increase in precipitation in store for future winters, while current DJF trends and JJA trends, both current and future, are estimated as not significantly different from zero. For CAL in DJF the trends are all constant, while in current and future JJA the trends appear significantly negative, but with no significant change in
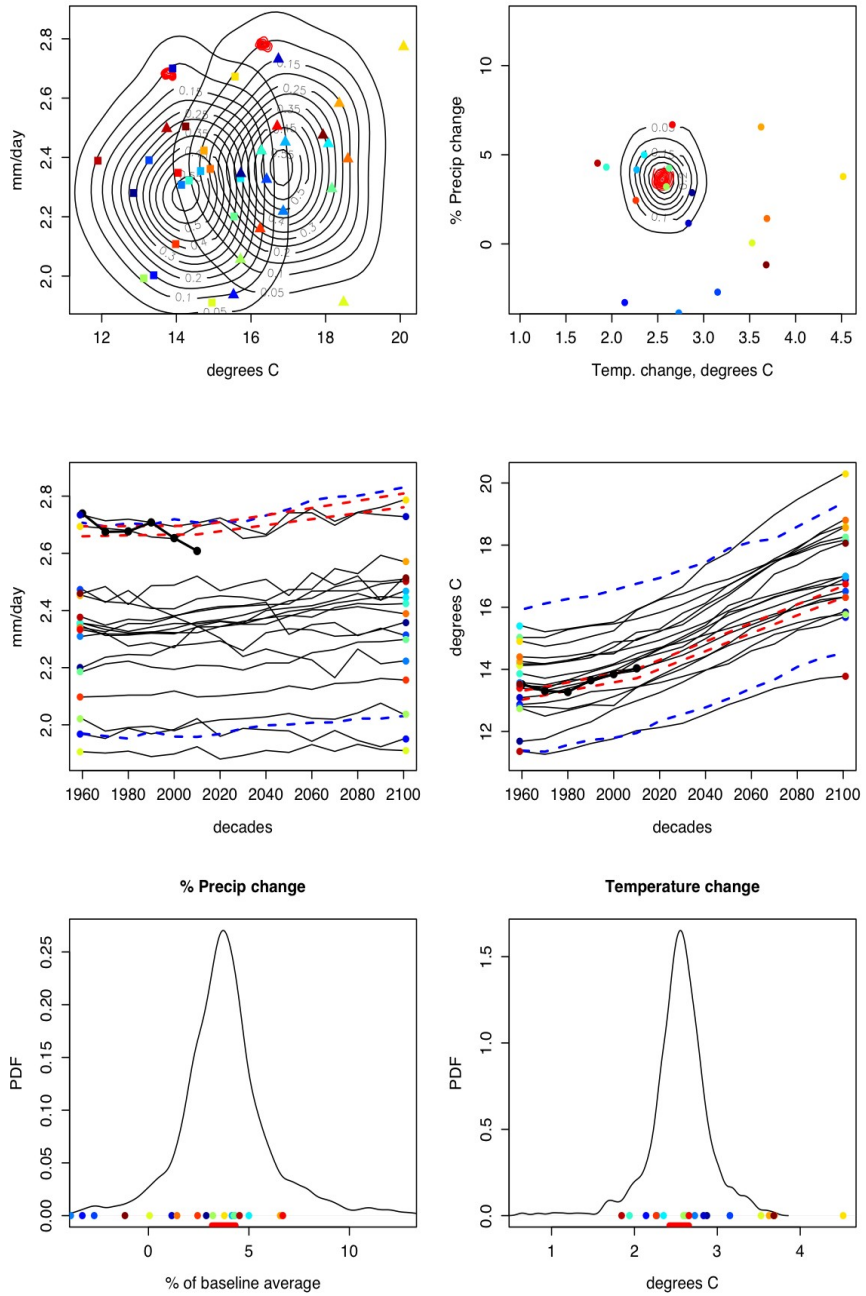
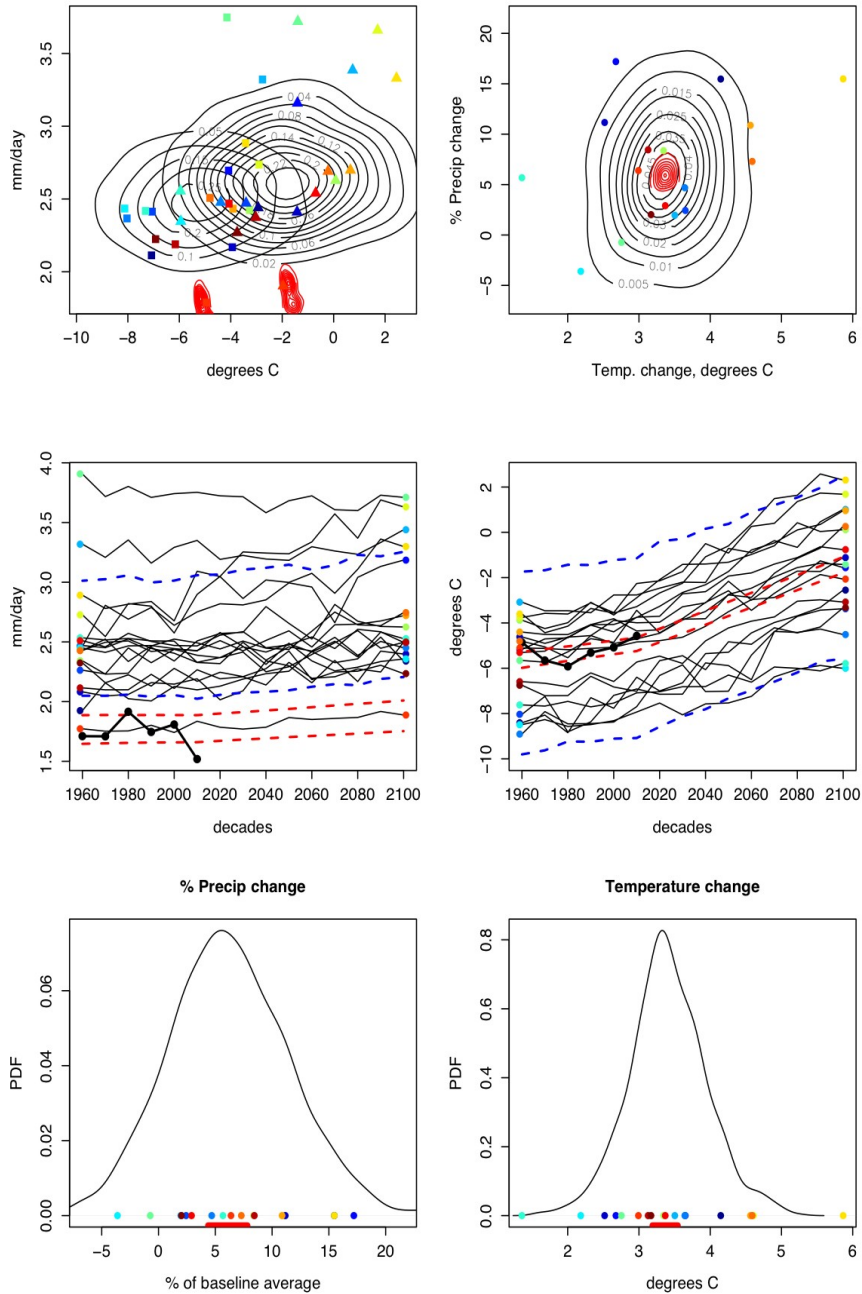**Fig. 3.** As in Figure 2, for global averages (GLOB) in summer (JJA).

Fig. 4. As in Figure 2, for Western North America (WNA) in winter (DJF).
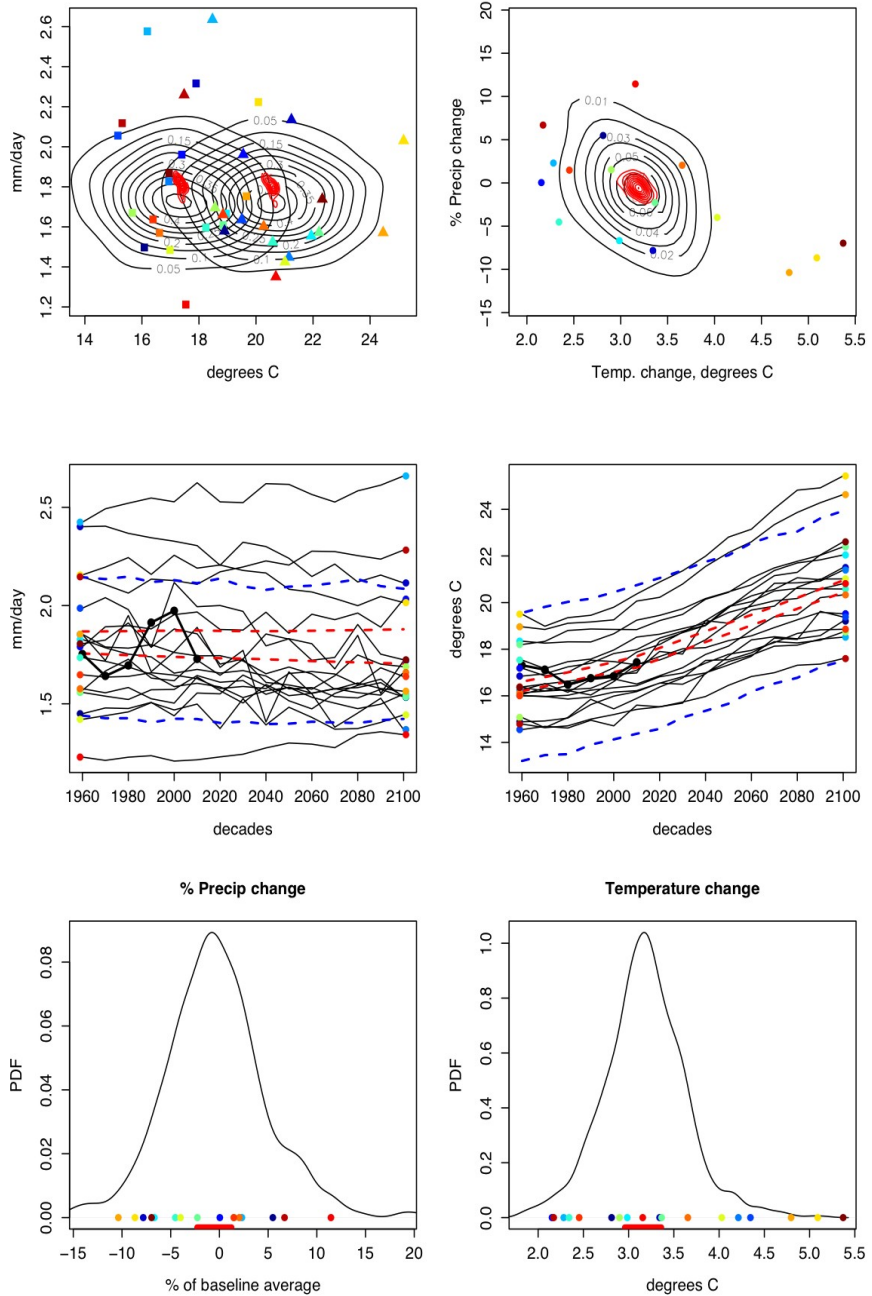
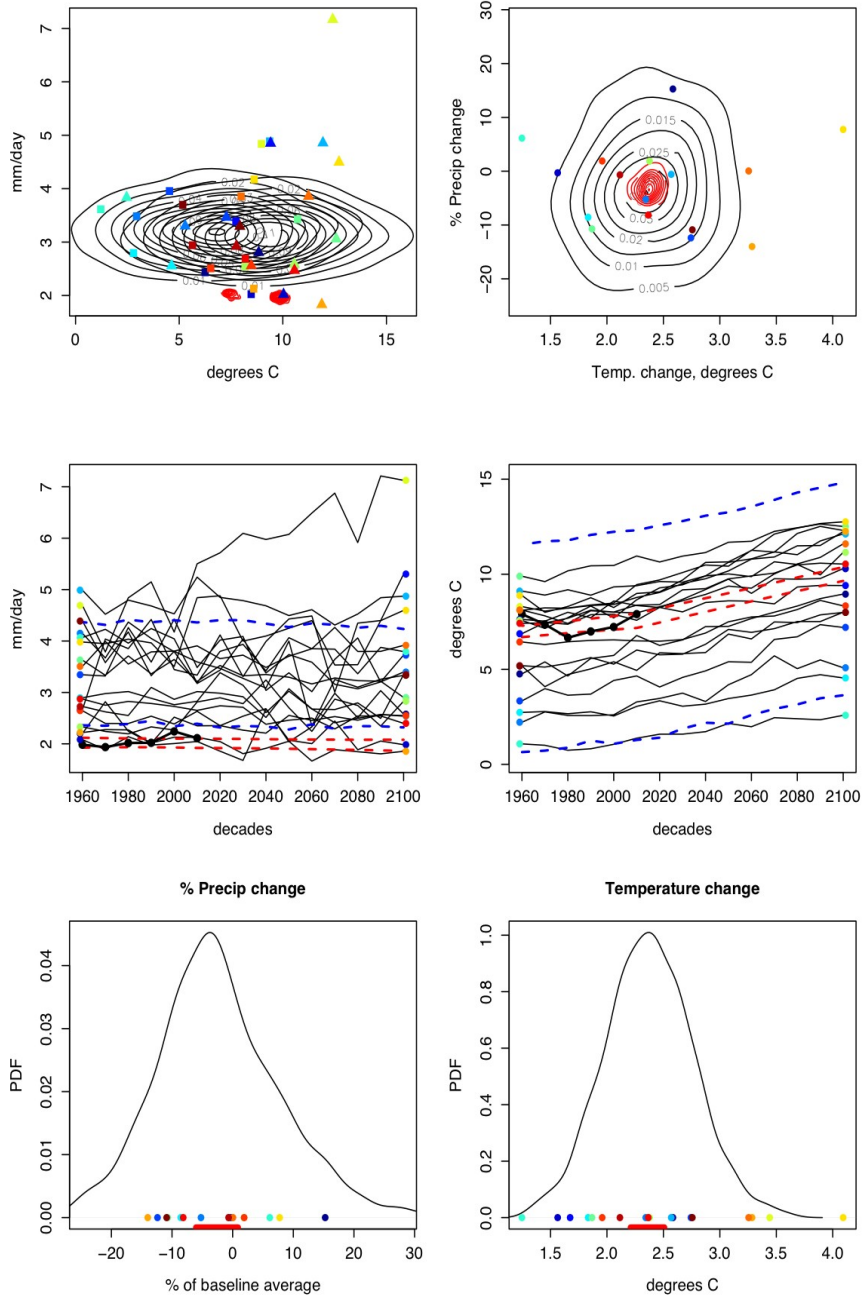**Fig. 5.** As in Figure 2, for Western North America (WNA) in summer (JJA).

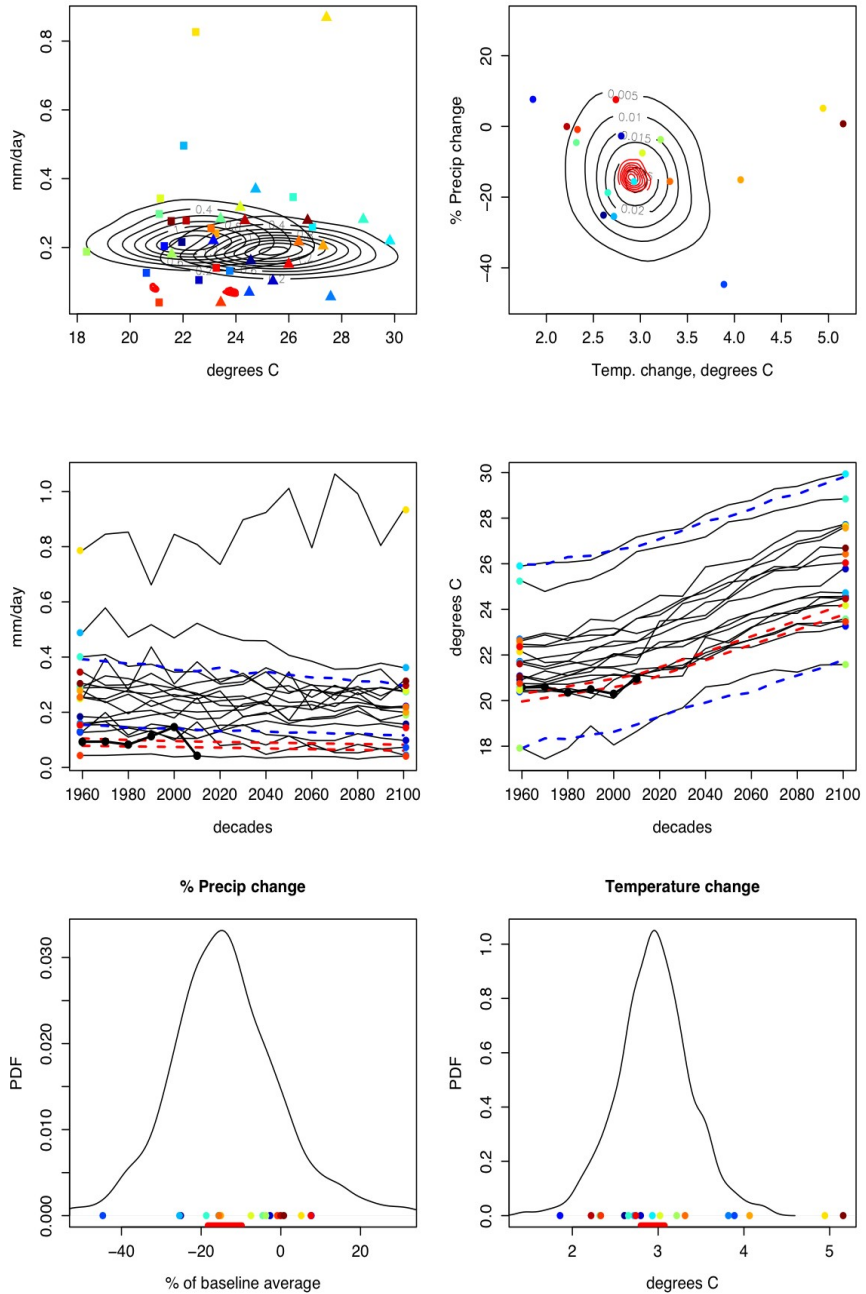**Fig. 6.** As in Figure 2, for California (CAL) in winter (DJF).

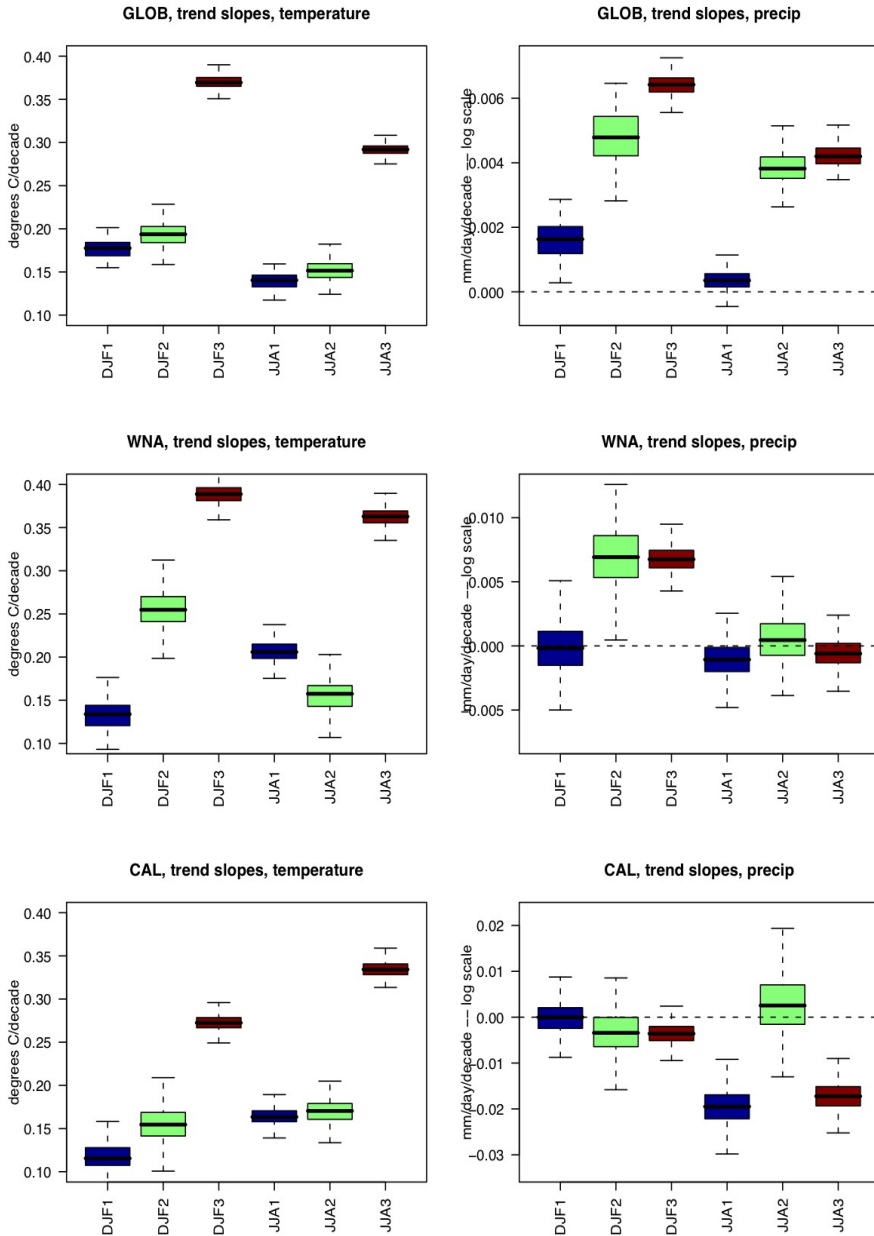**Fig. 7.** As in Figure 2, for California (CAL) in summer (JJA).

Fig. 8. Posterior distributions of the coefficients modeling the trend in the time series of the signal of temperature and precipitation. Each panel shows for one of the regions, and either temperature (left column) or precipitation (right column), three trend coefficients. The first three boxplots in each panel show trends in DJF, the last three show trends in JJA. For each triplet, the first boxplot represents the posterior distribution of the baseline trend β, the second boxplot reproduces the distribution of the incremental value starting at 2000, γ, the last boxplot shows the quantity $\beta + \gamma$, representing the future trend during the period 2000 to 2100. The boxes delimit the 25th and 75th

slope along the length of the time series. These results are also reflected in the position of the posterior distributions of the climate signals in the contour plots of Figures 2 through 7. There, the x-axis always shows only positive values, confirming a significant trend for temperature in all cases. The y-axis is instead always including zero, since the predictive contours are in all cases but one (GLOB, JJA) straddling the zero line in the vertical dimension. The red contours of the posterior distributions are however contained in the positive region of the y-axis for GLOB DJF and JJA, and WNA DJF; they are contained in the negative region for CAL JJA, and only for WNA JJA and CAL DJF the signal of change for precipitation is symmetrically distributed around the zero line.

In Figure 9 we show results regarding the estimate of correlation between temperature and precipitation. In our model the current climate correlation and a measure of its uncertainty are estimated through the observed record, and the corresponding parameters are held fixed. The posterior distribution is then obtained for all the correlation parameters that are model specific, β_{xj}, and for the mean parameter of the prior distribution, β_0, and its precision, that is hypothesized to be common to both modeled and observed climate. By doing so we let the observed value of the correlation drive the posterior estimates, whenever the observed record has enough information to provide a firm estimate of the correlation, with small uncertainty. Figure 9 presents the posterior density of the parameter β_0, as the first boxplot to the left, and those of the suite of model-specific parameters β_{xj}, together with the observed value that we indicate by a solid horizontal line. The band delimited by the dashed lines corresponds to two standard deviations around the estimate of β_{xo}. In all but two cases the posterior distribution of β_0 lies within the dashed lines, an indication that the observed value of the coefficient was effective in guiding the overall estimate. There are two region/season combinations, however, where all posterior distributions lie outside of the dashed band, at least for a significant part of their domain. We interpret these results by noticing how in both cases the model specific estimates are in overall agreement with one another, and fairly tight (i.e. precise) as suggested by the relatively narrow extent of the whiskers. In this cases the one fixed parameter is not enough to overcome the consensus estimate of the models, which pulls the posterior distribution of the overall correlation coefficient β_0 towards itself. Sensitivity analysis results – not shown here – indicate that this behavior can be overcome by increasing the apriori precision of the observed correlation coefficient, as should be expected.

Our model delivers many other marginal posterior distributions of parameters that may be of interest, in isolation or jointly, or of any deterministic function of theirs. For example one may be interested in analyzing the posterior distribution of climate changes at a closer time in the future, or we could
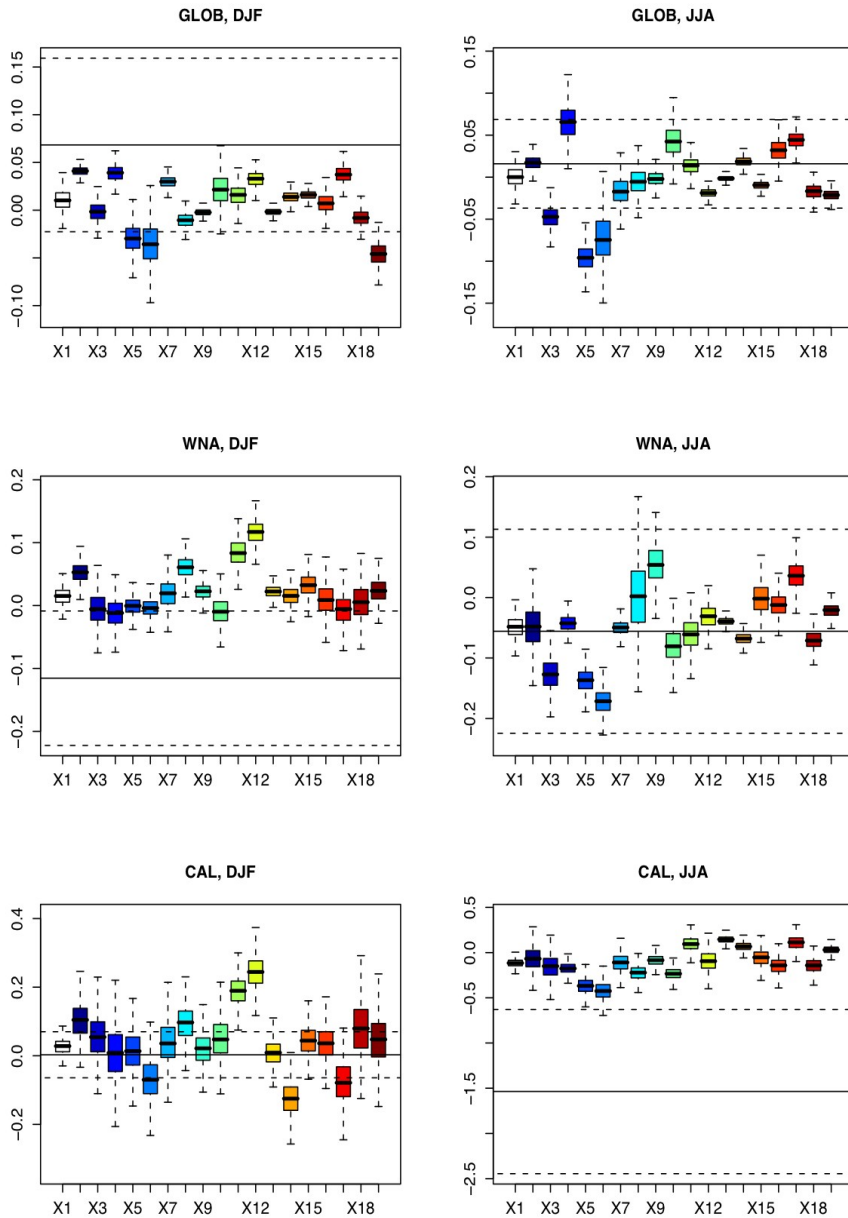
Fig. 9. Posterior distributions of the coefficients introducing correlation between the (detrended) means of temperature and precipitation, for the six regional/seasonal analyses. The first boxplot in each panel represents the posterior of the mean parameter in the prior distribution, β_0, from which we hypothesize that both the observed (indicated by the horizontal solid line, together with its two-standard deviations bounds by dashed lines) and the GCM-specific coefficients are drawn. The GCM-specific coefficients' posterior distributions are shown by the suite of colored boxplots to the right of the observed.

ask about model-specific, or overall biases, and model-specific precisions in simulating temperature or precipitation, or both.

Hindcasting of observed temperature trends could be used to assess the reliability of the "projections" (in this case aimed to the past) derived from the GCMs by our method. The small numbers of observed decadal means, however, and the fact that we are estimating trends makes us look for a different approach at model validation. In fact, we are not aware of any study that has used hindcast from GCM for the oberved record of the last few decades as an independent test of reliability, probably due to the fact that GCM development uses observed data in many fashions and makes comparison with recent past trends less of a rigorous test than we would like. Obviously, we do not have a straightforward way to validate our probabilistic projections of future climate, either, but we have a natural way to confirm that our modeling assumptions are consistent with the data at hand, namely, that our predictive distributions are consistent with our suite of GCM projections. Thus, separately for each of the six analyses, we left one GCM out in turn, computed three bivariate predictive distributions, for current climate, future climate and climate change (corresponding to the three kinds of large contours we presented in Figures 2 through 7). We then computed the two sets of pairs $(Z_1 = F_T(X_*^T = x_*^T), Z_2 = F_{P|T}(X_*^P = x_*^P))$ for both current and future time windows, and the pair $(Z_1 = F_T(\Delta X_*^T = \Delta x_*^T), Z_2 = F_{P|T}(\Delta X_*^P = \Delta x_*^P))$ where the first univariate CDF is simply the marginal predictive distribution of temperature change, while the second CDF is the predictive distribution of precipitation change, conditional on the corresponding simulated temperature change. We do this in turn for all 18 models and we test the null hypothesis that the pairs of (z_{1j}, z_{2j}) for $j = 1, \ldots, 18$ are independent and identically distributed random variates, sampled from uniform distributions on the $(0, 1)$ interval. For all tests (we have 36 of them, since we test the uniform distribution assumption for two samples of 18 values, for three different bivariate distributions and 6 different region/season combination) the null hypothesis was rejected only three times out of 36, a result not surprising in the context of multiple testing. The correlation between the two variables of the pair (Z_1, Z_2) resulted significant (above 0.46) only in one case out of 18, again not enough to raise concern on the validity of the independence assumption.

5. Standing issues and conclusions

We have modelled in a Bayesian hierarchical framework a complex dataset composed of decadal means of temperature and precipitation for specific regions and seasons, as observed and as simulated by a suite of state-of-the-art GCMs running experiments under specified GHG emission scenarios. Our model jointly

estimates a suite of uncertain parameters giving a full characterization of the climate evolution over the recent past and the future decades. The statistical model also provides a measure of the uncertainty that resides in the observed trends, confounded by natural variability, in the observed relation between temperature and precipitation anomalies, and in the limited ability of GCMs to simulate the evolution of our climate. The latter are due to necessary approximations of the fundamental climate processes that cannot be explicitly represented at this time in the evolution of climate modelling and computational resources, and thus require a full characterization if the results of the GCMs are to be used to inform adaptation and mitigation decisions and policy. Our statistical treatment produces a wealth of information, but perhaps most importantly delivers joint PDFs of temperature and precipitation change that can be naturally applicable to impact studies, overcoming the usual empasse experienced by other methods that provide only separate (marginal) projections of the two variable, while for most applications it is their joint behavior to be eminently relevant. From the statistical point of view we had to impose a number of simplifying assumption, by modeling the trend as a piecewise linear function, by allowing only an additive bias for the model simulations and by hypothesizing that it remains constant over the length of the simulation. We are also letting the observed trends and correlation heavily inform the final posterior estimates, choosing to overlook the problems inherent to observational uncertainties. These would be most compelling for climate change projections in under-developed regions of the world where data sparsity challenges robust estimates of historic trends and other parameters, and model validation is difficult to perform as a consequence. Another issue worth highlighting resides in the choice of treating the GCMs as independent pieces of information, while one could argue that there exist "families" of GCMs sharing modules and solutions to the representation of some of the processes at play. Modeling the dependence among GCMs would probably make the posterior distribution of the parameters less precise, since in our model each GCM contributes an independent piece of information to the estimation. Last but not least, we will not stress enough the limitations in the representation of the uncertainty based on this specific ensemble of models (Tebaldi and Knutti (2007)). Climate models use parameters to describe subgrid processes, and the parameter values are not exactly known, they are not derivable by physics law, but they are approximated by observational studies and modelers' expert judgment, and thus are rather uncertain. The proper way to deal with such uncertainties would be to introduce simulation loops where for each loop (i.e., climate model run) those parameter values would be varied. At present such many climate model runs (in the order of several thousands because many unknown parameters are involved) are computationally not feasible, even if experiments in so called perturbed physics ensembles are being

conducted for at least one climate model, at the Hadley center of the MetOffice in the UK (Murphy et al. (2007)) and we await the time when this approach will be shared by all climate model development centers. In the meantime, we see this approach as intuitive in the modeling assumptions, and easily modifiable in the presence of richer datasets, that may permit to model more sophisticated biases and GCM interdependencies and uncertainties. Thus, for the time being, we are confident in the results of the cross-validation exercise, which does not suggest any obvious shortcoming in the statistical assumptions.

Appendix: Full conditionals for MCMC sampling

Considering each random variable in turn, conditionally on the remaining variables, we can derive the full conditional distributions to implement the MCMC sampling algorithm. Note that in the following, the "prime" symbol denotes the operation of centering a variable (O_t or X_{jt}) by the respective climate signal $\mu_t = \alpha + \beta t + \gamma t \mathcal{I}_{\{t >= \tau_0\}}$.

Coefficients of the piecewise linear model:
Define

$$A = \tau_0 \eta^T + \tau_0 \eta^P \beta_{xo}^2 + \tau^* \sum_j \xi_j^T + \tau^* \sum_j \xi_j^P \beta_{xj}^2$$

and

$$
\begin{aligned}
B \;=\; & \eta^T \sum_{t \le \tau_0} (O_t^T - \beta^T t) + \eta^P \sum_{t \le \tau_0} \beta_{xo}^2 (O_t^T - \beta^T t - \beta_{xo} O_t^{P'}) + \sum_j \xi_j^T \sum_{t \le \tau_0} (X_{jt}^T - \beta^T t - d_j^T) \\
& \sum_j \xi_j^T \sum_{t > \tau_0} (X_{jt}^T - \beta^T t - \gamma^T(t - \tau_0) - d_j^T) + \sum_j \xi_j^P \beta_{xj}^2 \sum_{t \le \tau_0} (X_{jt}^T - \beta^T t - d_j^T - \beta_{xj}(X_{jt}^P \\
& \sum_j \xi_j^P \beta_{xj}^2 \sum_{t > \tau_0} (X_{jt}^T - \beta^T t - \gamma^T(t - \tau_0) - d_j^T - \beta_{xj}(X_{jt}^{P'} - d_j^P)).
\end{aligned}
$$

Then

$$\alpha^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \tau_0 \eta^P + \tau^* \sum_j \xi_j^P$$

and

$$
\begin{aligned}
B \;=\; & \eta^P \sum_{t \le \tau_0} (O_t^P - \beta^P t - \beta_{xo} O_t^{T'}) + \sum_j \xi_j^P \sum_{t \le \tau_0} (X_{jt}^P - \beta^P t - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)) + \\
& \sum_j \xi_j^P \sum_{t > \tau_0} (X_{jt}^P - \beta^P t - \gamma^P(t - \tau_0) - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)).
\end{aligned}
$$

Then
$$\alpha^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define
$$A = \eta^T \sum_{t \le \tau_0} t^2 + \eta^P \beta_{xo}^2 \sum_{t \le \tau_0} t^2 + \sum_j \xi_j^T \sum_{t \le \tau^*} t^2 + \sum_j \xi_j^P \beta_{xj}^2 \sum_{t \le \tau^*} t^2$$

and
$$B = \eta^T \sum_{t \le \tau_0} t(O_t^T - \alpha^T) + \eta^P \sum_{t \le \tau_0} \beta_{xo}^2 t(O_t^T - \alpha^T) - \beta_{xo} t O_t^{P'}) + \sum_j \xi_j^T \sum_{t \le \tau_0} t(X_{jt}^T - \alpha^T - d_j^T) +$$
$$\sum_j \xi_j^T \sum_{t > \tau_0} t(X_{jt}^T - \alpha^T - \gamma^T(t - \tau_0) - d_j^T) + \sum_j \xi_j^P \sum_{t \le \tau_0} t(\beta_{xj}^2 (X_{jt}^T - \alpha^T - d_j^T) - \beta_{xj}(X_{jt}^{P'} - $$

Then
$$\beta^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define
$$A = \eta^P \sum_{t \le \tau_0} t^2 + \sum_j \xi_j^P \sum_{t \le \tau^*} t^2$$

and
$$B = \eta^P \sum_{t \le \tau_0} t(O_t^P - \alpha^P - \beta_{xo} O_t^{T'}) + \sum_j \xi_j^P \sum_{t \le \tau_0} t(X_{jt}^P - \alpha^P - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)) +$$
$$\sum_j \xi_j^P \sum_{t > \tau_0} t(X_{jt}^P - \alpha^P - \gamma^P(t - \tau_0) - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)).$$

Then
$$\beta^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define
$$A = \sum_j \xi_j^T \sum_{t > \tau_0} (t - \tau_0)^2 + \sum_j \xi_j^P \beta_{xj}^2 \sum_{t > \tau_0} (t - \tau_0)^2$$

and
$$B = \sum_j \xi_j^T \sum_{t > \tau_0} (t - \tau_0)(X_{jt}^T - \alpha^T - \beta^T t - d_j^T) + \sum_j \xi_j^P \sum_{t > \tau_0} (t - \tau_0)(\beta_{xj}^2 (X_{jt}^T - \alpha^T - \beta^T t - d_j^T) - \beta_{xj}(X_{jt}^{P'} - d_j^P$$

Then
$$\gamma^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \sum_j \xi_j^P \sum_{t > \tau_0} (t - \tau_0)^2$$

and

$$B = \sum_j \xi_j^P \sum_{t > \tau_0} (t - \tau_0)(X_{jt}^P - \alpha^P - \beta^P t - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)).$$

Then

$$\gamma^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Bias terms and their priors' parameters:
Define

$$A = \tau^* \xi_j^T + \tau^* \xi_j^P \beta_{xj}^2 + \lambda_D^T$$

and

$$B = \xi_j^T \sum_{t \leq \tau^*} X_{jt}^{T'} + \xi_j^P \sum_{t \leq \tau^*} (\beta_{xj}^2 X_{jt}^{T'} - \beta_{xj}(X_{jt}^{P'} - d_j^P)) + \lambda_D^T a^T.$$

Then

$$d_j^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \tau^* \xi_j^P + \lambda_D^P$$

and

$$B = \xi_j^P \sum_{t \leq \tau^*} (X_{jt}^{P'} - \beta_{xj}(X_{jt}^{T'} - d_j^T)) + \lambda_D^P a^P.$$

Then

$$d_j^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define $A = M\lambda_D^T$ and $B = \lambda_D^T \sum_j d_j^T$, then

$$a^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define $A = M\lambda_D^P$ and $B = \lambda_D^P \sum_j d_j^P$, then

$$a^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

$$\lambda_D^T \sim \mathcal{G}(1 + \frac{M}{2}; 1 + \frac{\sum_j (d_j^T - a^T)^2}{2}).$$

$$\lambda_D^P \sim \mathcal{G}(1 + \frac{M}{2}; 1 + \frac{\sum_j (d_j^P - a^P)^2}{2}).$$

The correlation coefficients between temperature and precipitation in the models, and their prior parameters:

Define $A = \xi_j^P \sum_t (X_{jt}^{T'} - d_j^T)^2 + \lambda_B$ and $B = \xi_j^P \sum_t (X_{jt}^{T'} - d_j^T)(X_{jt}^{P'} - d_j^P) + \lambda_B \beta_0$, then

$$\beta_{xj} \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define $A = M\lambda_B + \lambda_o$ and $B = \lambda_B \sum_{j>0} \beta_{xj} + \lambda_o \beta_{xo}$, then

$$\beta_0 \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

$$\lambda_B \sim \mathcal{G}(0.01 + \frac{M}{2}; 0.01 + \frac{\sum_j (\beta_{xj} - \beta_0)^2}{2}).$$

Precision terms for the models:

$$\xi_j^T \sim \mathcal{G}(a_{\xi^T} + \frac{\tau^*}{2}; b_{\xi^T} + \frac{\sum_t (X_{jt}^{T'} - d_j^T)^2}{2}).$$

$$\xi_j^P \sim \mathcal{G}(a_{\xi^P} + \frac{\tau^*}{2}; b_{\xi^P} + \frac{\sum_t (X_{jt}^{P'} - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T))^2}{2}).$$

Only the full conditionals of the hyperparameters $a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P}$ cannot be sampled directly, and a Metropolis step is needed. We follow the solution described in Smith et al. (2008). The algorithm works identically for the two pairs, and we describe it for a_{ξ^T} and b_{ξ^T} (the sampling is done jointly for the pair). We define U_1, U_2 as independent random variables, uniformly distributed over the interval $(0, 1)$, and we compute two proposal values $a'_{\xi^T} = a_{\xi^T} e^{(\delta(u_1 - \frac{1}{2}))}$ and $b'_{\xi^T} = b_{\xi^T} e^{(\delta(u_2 - \frac{1}{2}))}$, where δ is an arbitrary increment, that we choose as $\delta = 1$. We then compute

$$
\begin{aligned}
\ell_1 =\ & M a_{\xi^T} \log b_{\xi^T} - M \log \Gamma(a_{\xi^T}) + (a_{\xi^T} - 1) \sum_j \log \xi_j^T - b_{\xi^T} \sum_j \xi_j^T + \\
& 0.01 \log(a_{\xi^T} b_{\xi^T}) - 0.01(a_{\xi^T} + b_{\xi^T}),
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
\ell_2 =\ & M a'_{\xi^T} \log b'_{\xi^T} - M \log \Gamma(a'_{\xi^T}) + (a'_{\xi^T} - 1) \sum_j \log \xi_j^T - b'_{\xi^T} \sum_j \xi_j^T + \\
& a \log(a'_{\xi^T} b'_{\xi^T}) - b(a'_{\xi^T} + b'_{\xi^T}).
\end{aligned}
\tag{6}
$$

In (5) and (6) we are computing the log likelihoods of (a_{ξ^T}, b_{ξ^T}) and (a'_{ξ^T}, b'_{ξ^T}). Then, within each iteration of the Gibbs/Metropolis algorithm, the proposed values (a'_{ξ^T}, b'_{ξ^T}) are accepted with probability $e^{\ell_2 - \ell_1}$ if $\ell_2 < \ell_1$, or 1 if $\ell_2 \geq \ell_1$.

Acknowledgements

References

Collins, M. and Knight, S. (2007) *Phil. Trans. of the Royal Soc. Series A, Theme Issue Ensembles and probabilities: a new era in the prediction of climate change*, vol. 365. n. 1857, Royal Society Publishing. Pp. 1957-2191.

Furrer, R., Sain, S., Nychka, D. and Meehl, G. (2007) Multivariate Bayesian analysis of atmosphere-ocean general circulation models. *Environmental and Ecological Statistics*, **14**, 249–266.

Giorgi, F. and Mearns, L. (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.

Greene, A., Goddard, L. and Lall, U. (2006) Probabilistic multimodel regional temperature change projections. *J. Clim.*, **19**, 4326–4343.

Groves, D. G., Yates, D. and Tebaldi, C. (2008) Uncertain global climate change projections for regional water management planning. *Water Resources Research*. (submitted).

IPCC (2007) *Climate Change 2007 – The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC*. Solomon, S. et al. (eds.), Cambridge University Press. 996 pp.

Lobell, D. B. and Field, C. B. (2007) Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, **2**, 1–7.

Murphy, J., Booth, B., Collins, M., G.R. Harris, G., Sexton, D. and Webb, M. (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Phil. Trans. R. Soc. A*, **1857**, 1993–2028.

Rosenblatt, M. (1952) Remarks on a multivariate transformation. *Annals of Statistics*, **23**, 470–472.

Smith, R., Tebaldi, C., Nychka, D. and Mearns, L. (2008) Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*. (forthcoming).

Tebaldi, C. and Knutti, R. (2007) The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A*, **1857**, 2053–2075.

Tebaldi, C., Mearns, L., Nychka, D. and Smith, R. (2004) Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophys. Res. Let.*, **31**, doi:10.1029/2004GL021276.

Tebaldi, C., Smith, R., Nychka, D. and Mearns, L. (2005) Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multi-model ensembles. *J. Clim.*, **18**, 1524–1540.