We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 4

Statistical Methodology for Evaluating Process-Based Climate Models

Firdos Khan and Jürgen Pilz

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.80984

Abstract

In climatology, there are mainly two types of models used, that is, global circulation/climate models (GCMs) and regional climate models (RCMs). GCMs can be run for the whole globe, while RCMs can be run only for a part of the globe. In this chapter, we provided a general statistical methodology for evaluating process-based (GCM or RCM) climate models. To bridge observed and simulated data sets, statistical bias correction was implemented. A meta-analysis technique is used for selecting a model or scenarios, which have good performance compared to others. For model selection and ensemble projection, Bayesian model averaging (BMA) is used. Posterior inclusion probability (PIP) is used as model selection criterion. Our analysis concluded with a list of best models for maximum, minimum temperature, and precipitation where the rank of the selected models is not the same for the listed three variables. The outputs of BMA closely followed the pattern of observed data; however, it underestimated the variability. To overcome this issue, 90% prediction interval was calculated, and it showed that almost all the observed data are within these intervals. The results of Taylor diagram show that the BMA projected data are better than the individual GCMs' outputs.

Keywords: bias correction, climate change, meta analysis, model selection, posterior inclusion probability

1. Introduction

This chapter is basically about statistical evaluation of climate models; however, prior to model's evaluation, it is important to highlight briefly about climate models and their types. In the literature, climate models are also known as process-based models as these models work closely to the physical process of the climate of our planet. Broadly speaking, climate models can be

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

divided into two types, global circulation/climate model (GCM) and regional climate model (RCM). As the name of these models suggests, GCM can be run for the whole globe, while RCM can be run for a particular location of interest. Due to different uncertainties, we need to evaluate these models before using their outputs for further analysis. Evaluation of a climate model means assessing a model's performance so as to articulate the grounds on which a model can be declared good enough for its anticipated use. Model's evaluation is an important step in climate change assessment and impact assessment studies. It provides guide lines to choose the best models or scenarios for further analysis. For evaluation of climate models, we need observational data and historical simulated data by using climate models for the same variables and for the same time period. This chapter presents a combination of classical and Bayesian statistical approaches for the evaluation of climate models. This chapter is structured as: right after this brief introduction, climate models in general and GCM and RCM in particular and their evaluation are briefly explained, and methodology is discussed in Section 2, Section 3 is reserved for results and discussion, and Section 4 comprised of summary and conclusion.

1.1. Climate model

A climate model is a complex system of mathematical equations which represents the physical process among various components which contribute to the climate of our globe. To run a GCM, the globe is divided into a number of grid boxes with horizontal resolution (latitude and longitude) and vertical resolution (height or pressure). The climatology is solved in each grid after providing the initial conditions to the main deriving climate variables like temperature, wind speed, humidity, pressure, etc. The brief details about GCMs and RCMs are presented in the subsequent sections, for more details about GCMs, we refer to [1].

1.2. Global climate model

GCMs are the most modern and sophisticated tools available to provide basic information about climate globally. These models comprise complex mathematical equations which represent the physical process of atmosphere, ocean, cryosphere, and land surface. GCMs are run for the whole globe, and due to the complex system, it takes long time in simulations; however, super computers can be used to make their performances more efficient. According to [1], a GCM is "Numerical models, representing physical processes in the atmosphere, ocean, cryosphere and land surface, are the most advanced tools currently available for simulating the response of the global climate system to increasing greenhouse gas concentrations." The process of simulating climate systems by using climate models is called dynamical modeling or dynamical downscaling. The nesting of dynamical modeling and simulation of climate systems is presented in **Figure 1** using GCM and RCM, while a list of some popular GCMs is presented in **Table 1** along with some basic information about each model.

1.3. Regional climate model

RCMs are also process-based climate models and comprise complex mathematical equations like GCMs; however, these models can be run for a particular location of interest. As GCMs can be run for the whole globe, therefore, the grid size is coarser, and the information in the

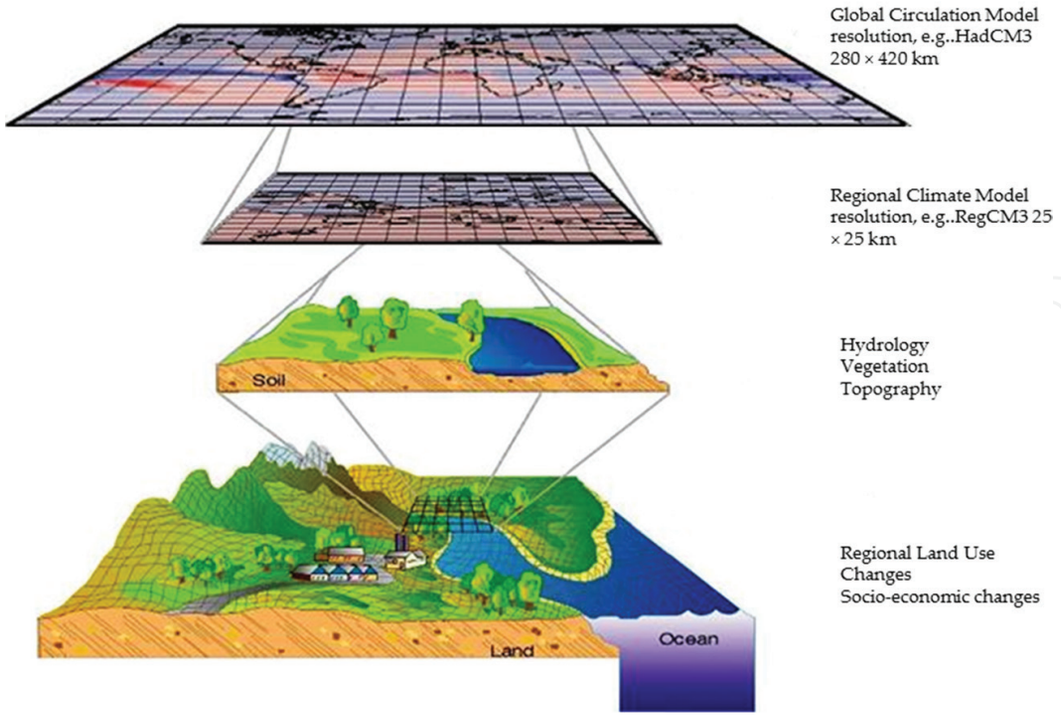


Figure 1. The chain of dynamical downscaling, global climate modeling, and then regional climate modeling to get climate information at higher resolution from coarser resolution.

S. No.	Name	Resolution (Degree*)	Institute/Center
1	CanESM2	2.7906 × 2.8125	Canadian Centre for Climate Modeling and Analysis
2	CCSM4	1.250 × 0.942	National Center for Atmospheric Research
3	CESM1-CAM5	1.250 × 0.942	National Center for Atmospheric Research
4	CMCC-CMS	1.875 × 1.865	Centro Euro-Mediterraneo per I Cambiamenti Climatici
5	CNRM-CM5	1.406 × 1.401	Centre National de Recherches Meteorologiques
6	EC-EARTH	1.1215 × 1.125	Irish Centre for High-End Computing (ICHEC), European Consortium
7	GFDL-ESM2G	2.500 × 2.023	Geophysical Fluid Dynamics Laboratory
8	GFDL-ESM2M	2.500 × 2.023	Geophysical Fluid Dynamics Laboratory
9	INM-CM4	2.000 × 1.500	Institute for Numerical Mathematics
10	MIROC-ESM-CHEM	2.7906 × 2.8125	National Institute for Environmental Studies, The University of Tokyo
11	MPI-ESM-LR	1.875 × 1.865	Max Planck Institute for Meteorology (MPI-M)
12	MPI-ESM-MR	1.875 × 1.865	Max Planck Institute for Meteorology (MPI-M)
13	NorESM1-M	2.500 × 1.895	Norwegian Climate Centre

*1 degree is approximately equal to 111.32 km.

Table 1. Details about GCMs used in this study with their resolutions and other brief information.

form of output of GCMs is at lower resolution. For impact assessment studies like impact of climate change on water resources, agriculture, urban planning, etc., we need information at higher resolution. Toward this end, we need to do regional climate modeling which provides climatic information at higher resolutions. Due to the rapid development in computational technology, the modern RCMs can be run with a resolution of 10 km or even higher resolution.

1.4. Evaluation of climate models

Evaluation of the process-based climate models is an important step in climate change assessment and impact assessment studies. Different methods can be used for this purpose; however, we presented a combination of advanced statistical methods for model evaluation including classical and Bayesian approaches. Evaluation of process-based climate model will give guide lines about model or scenario selection to researchers in climate changes assessment and studies related to the impact assessment of climate change in different areas. This will help researchers to use specified and representative models rather than randomly selected model and to get realistic results. Statistical bias correction is important to reduce the gap between observed and model's simulated data and considers initial step toward climate change assessment. Meta analysis is used for scenario analysis, which assigns higher weights on the basis of precision of a particular scenario. For model (GCM in our case), we used Bayesian model averaging (BMA) technique to choose a set GCMs performing better than others. On the basis of chosen GCMs, ensemble projections were calculated using BMA technique and further evaluated using Taylor diagram along with individual GCMs' outputs. Further details are presented in methodology section about each method and examples with discussion in results and discussion section.

1.5. Objectives of this study

This study aims to present statistical methods including frequentist as well as Bayesian which can be used for the evaluation of process-based models with detailed examples and discussion using the real world data.

2. Data and methodology

2.1. Data

Two types of data sets have been used in this research, observed and climate model's simulated data. Observational data were acquired from Pakistan Meteorological Department (PMD) on daily frequency for three climate variables including maximum, minimum temperature, and precipitation for different locations of Pakistan. RCM's simulated data were collected from COordinated Regional climate Downscaling Experiment (CORDEX), Met. Office of the United Kingdom (UK), and Global Change Impact Studies Centre (GCISC), Pakistan. For evaluation purpose, we used data for the baseline time period, which is 1960–1990 for both observed and simulated data, but for calculating ensemble projections and their evaluation, the baseline is considered as 1975–2005.

2.2. Statistical bias correction

According to [2], all models are wrong but some of them are useful. Climate models provide useful information; however, there are various sources of uncertainties which have influence on the outputs of these models [3, 4]. To bridge the difference between observed and model's simulated data, we need to utilize statistical methods. In order to carry out statistical bias correction, different statistical methods were developed starting from simple to most sophisticated ones. For detailed literature about statistical bias correction methods, we refer to [5, 6]. We present the methodology of latest developed methods by [7], which preserve trend and climate extremes in future climate model's simulations called quantile delta mapping (QDM).

A four step methodology is required to implement the QDM method, starting from the cumulative distribution function (CDF) of model projected series $Y_{m,f}$. We assume that f, h, m, and o stand for future, historical, model, and observed data, respectively. Further, F and Y represent CDF of the data and original data, respectively.

$$F_{m,f}\big(y(t)\big) = P\big(Y_{m,f}(t) \le y(t)\big), F_{m,f}(t) \in [0, 1] \tag{1}$$

To proceed to the second step, we need to find the relative change using the ratio of the inverse CDF of model predicted data applied to the CDF of model predicted data and the inverse CDF of historical observed data applied to model predicted data. Mathematically, this can be written by Eq. (2).

$$\Delta_m\big(y(t)\big) = \frac{F_{m,f}^{-1}\big(F_{m,f}\big(y(t)\big)\big)}{F_{m,h}^{-1}\big(F_{m,f}\big(y(t)\big)\big)} = \frac{y(t)}{F_{m,h}^{-1}\big(F_{m,f}\big(y(t)\big)\big)} \tag{2}$$

The quantiles of model's predicted data $F_{m,f}(y(t))$ can now be bias corrected by implementing the inverse CDF estimated from historical observational data.

$$\hat{Y}_{o,m}(t) = F_{o,h}^{-1}\big(F_{m,f}\big(y(t)\big)\big) \tag{3}$$

Finally, the bias corrected future projections can be obtained by applying the relative changes to the historical bias corrected data presented in Eq. (4).

$$\hat{Y}_{m,f}(t) = \hat{Y}_{o,m}(t).\Delta_m\big(y(t)\big) \tag{4}$$

$\hat{Y}_{m,f}(t)$ is the future model's bias corrected data which can be used now for further analysis. To preserve absolute changes in the data, Eqs. (2) and (4) can be applied additively rather than multiplicatively [7]. The multivariate counterpart of the method presented in [7] is available and can be found in [8]. One advantage of multivariate quantile mapping bias correction (presented in [8]) is that it preserves spatial dependence structures between climate variables when we are applying this method to more than one variable simultaneously. This is

especially important when we are dealing with impact assessment studies like hydrological modeling, agricultural production, etc. under the changing climate.

2.3. Meta analysis

Scenario's or model's assessment is an essential part of climate change analysis as it provides valuable information about a particular scenario or model. Meta analysis is a statistical method which can be used for this purpose; however, it is also a useful technique to produce a combined estimate of projections from individual model outputs or different scenarios. It gives weight to each study on the basis of its precision and, consequently, provides confidence in future projections which have higher precision. Usually, researchers prefer models, scenarios, studies, laboratories' outputs, etc., which have higher weights than those with lower weights. In order to accomplish the evaluation of models or scenarios using meta analysis, the three step methodology is explained briefly in the following subsections.

2.3.1. Selection of the model

There are two basic models to perform a meta analysis: the fixed effect model (FEM) and the random effect model (REM) [9]. The FEM assumes that all the studies included in the meta analysis come from a single identical population or share a common effect (mean or average), while a REM assumes that the effects of the studies included in the meta analysis form a random sample from a population following a specified distribution. The observed effects in the FEM and REM are mathematically presented in Eqs. (5) and (6), respectively. Suppose we have k studies, and let θ denote the (true) intervention effect in the population, which we would like to estimate. Further, let θ_k denote the kth study effect, and ζ_k the random effect in this study; $k = 1, 2, \ldots, K$.

$$\theta_k = \theta + \varepsilon_k,\ \varepsilon_k \sim N\left(0, v_k^2\right) \tag{5}$$

$$\theta_k = \theta + \zeta_k + \varepsilon_k,\ \zeta_k \sim N(0, \tau^2) \tag{6}$$

Here, ε_k describes the variation within the kth study, and the random effects ζ_k reflect the variations between the considered studies. FEM is a special case of REM when the variations between studies are equal to zero:

$$\zeta_1 = \zeta_2 = \ldots = \zeta_K = 0, \tag{7}$$

then the random effect model reduces to the fixed effect model. Model selection is mainly based on the nature and objectives of the study [9–11]. This is an important step because the remaining two steps depend on model selected. Therefore, model selection must be made carefully. This chapter presented both models for explanation, and in results and discussion section, we present an example to make clear the differences between these two models.

2.3.2. Weighting schemes for parameter estimation

Different weighting schemes are available for the estimation of the effect size in meta analysis; however, it depends on the nature of the study to choose one of them [9]. We proceed with the so-called inverse-variance weighting technique for quantifying the effect size in our analysis. For details about different weighting schemes, we refer to [10]. According to [9], all the available schemes are efficient because they assign higher weights to more precise studies. In case of a fixed-effect model, the weights are calculated by Eq. (8).

$$\omega_k = \frac{1}{v_k^2} \tag{8}$$

where ω_k and v_k^2 are the weight and variance, respectively, of the kth study. In a random effect model, the weights are calculated by Eq. (9).

$$\begin{aligned} \omega_k^* &= \frac{1}{v_k^*} \\ v_k^* &= v_k^2 + \tau^2 \end{aligned} \tag{9}$$

where ω_k^* is the weight for kth study, and v_k^* is the combined variance of within-study and between-studies (τ^2). As we had already discussed, the weights for estimating the effect size depend on the model chosen in the model specification stage.

2.3.3. Estimation of parameters

The next step is to estimate the unknown parameters of the specified model by incorporating the weighted least squares method given by Eq. (10).

$$\begin{aligned} \theta_c &= \frac{\sum_{k=1}^{K} W_k.\theta_k}{\sum_{k=1}^{K} W_k} \\ W_k &= \frac{1}{var(\theta_k)} \end{aligned} \tag{10}$$

The $(1-\alpha)\times 100\%$ confidence interval of the combined estimator is given by

$$\theta_c \pm Z_{\left(1-\frac{\alpha}{2}\right)} \times SE(\theta_c)$$

where θ_c is the combined size effect, $SE(\theta_c)$ is the standard error, and $Z_{\left(1-\frac{\alpha}{2}\right)}$ is the $\left(1-\frac{\alpha}{2}\right)$-quantile of the standard normal distribution.

2.4. Model (GCM) selection, ensemble projections, and their evaluation

In this part, Bayesian model averaging (BMA) is used for two purposes, that is, model selection and producing ensemble projections by using the outputs of selected GCMs and finally the evaluation of models' outputs. **Table 1** lists some popular GCMs with brief details about

their resolution, and the institutes where each model was developed. The outputs of these models are used in the subsequent sections of this chapter. However, before embarking on this journey, it is important to discuss briefly the concept of posterior probability which is at the core of the Bayesian approach. Bayesian model averaging is discussed afterwards.

2.4.1. Posterior probability

Bayes' theorem states that the posterior probability of j^{th} model, $p(M_j \mid D)$, is calculated as the likelihood of observed data given j^{th} model, $p(D \mid M_j)$, multiplied by the prior probability of the j^{th} model, and divided by the probability of having the current observation realization, $p(D)$. The posterior probability is thus calculated as follows:

$$P(M_j \mid D) = \frac{p(D \mid M_j).p(M_j)}{p(D)} \tag{11}$$

In Eq. (11), $p(D)$ is used as a normalizing constant given in Eq. (12), and hence the Bayes' rule can be simply stated as in Eq. (13).

$$p(D) = \sum_{j=0}^{s} p(D \mid M_j).p(M_j) \tag{12}$$

$$p(M_j \mid D) \propto p(D \mid M_j).p(M_j) \tag{13}$$

The prior distribution of a model shows the probability allocated to a statistical model. In this study, we have $M_j; j = 0, 1, 2, \ldots, s = 2^k$-1 possible statistical models. The likelihood of observation represents the probability of getting the current model realization. The posterior probability of a model represents the probability of the model to realize the current model given observations. Different choices for the prior are available; however, the users can also implement their own customized priors for their analysis. In case of using a uniform prior distribution (i.e., $p(M_j) \propto 1/2^k$), assigning equal prior weight to all models then the posterior model probability can be expressed by Eq. (14).

$$p(M_j \mid D) \propto p(D \mid M_j) \tag{14}$$

Eq. (14) shows that in this case the posterior probability of a model is only determined by the likelihood of observational data. Likelihood of a model reflects the ability to reproduce a given system of observed data. Different likelihood functions have been proposed to calculate the likelihood, $p(D \mid M_j)$, for example, see [13–15, 17]. A Gaussian likelihood proposed by [16] is used in this chapter.

2.4.2. GCM selection

The rapid developments in the computational technology make it practicable to run complex process-based climate models for simulating complex climate systems of our planet;

however, the output from a single model still may have uncertainties [18]. There has been a number of GCMs developed to project the future global climate change and use their output for impact assessment studies in different areas [1]. Due to different parameterization schemes of GCMs, internal atmospheric variability [19], and uncertainties in input data, different GCMs may produce quite different results. Therefore, it is important to consider more GCMs instead of relying on a single GCM. Regression models can be used to estimate the observed climate by using outputs from different GCMs as covariates. In a regression model context, the problem of uncertainty modeling has been raised by Raftery et al. [20]. In such models, covariate (GCM here) selection is a basic part to build a valid regression model, and the objective is then to find the "best" model for the response variable and a given set of predictors. The first problem to solve is which covariates should be included in the model and how important are they? Suppose we have a response variable Y and set of covariates, $X_1, \ldots, X_k$, and E represents the expected value, then there are 2^k different linear regression models expressing the relationship between the response variable and the potential predictors as follows:

$$\begin{gathered} M_0: EY = \beta_0, M_1: EY = \beta_0 + \beta_1 X_1, \ldots, M_k: EY = \beta_0 + \beta_k X_k \\ M_{k+1}: EY = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \ldots, M_{k+l}: EY = \beta_0 + \beta_{k-1} X_{k-1} + \beta_k X_k; l = k(k-1)/2 \\ M_{k+l+1}: EY = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \ldots, M_{k+l+m}: EY = \beta_0 + \beta_{k-2} X_{k-2} + \beta_{k-1} X_{k-1} + \beta_k X_k, \\ m = \frac{k(k-1)(k-2)}{6} \\ M_{2^k-1}: EY = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \end{gathered} \tag{15}$$

The same procedures are used here for GCM selection, where GCM now stands for a model M_j; $j = 1, \ldots, m = 13$; and a uniform prior was used as a prior probability distribution for GCMs ($p(GCM) \propto \frac{1}{m}$). The posterior inclusion probability (PIP) is the sum of posterior probabilities of each covariate (GCM) from all possible models included in BMA used as model selection criterion. The PIP has a range between zero and one, where a value close to one means that the GCM closely reproduces the observed data, while a value close to zero means that the corresponding GCM's output does not agree at all with the observed data.

It has a strong and solid background in Bayesian statistics, and there is a rich body of literature on BMA and PIP. GCM selection was made along the following four steps:

1. Run BMA as presented in Eq. (15).
2. Calculate the posterior probability for each GCM included in all possible regression models in step 1.
3. Sum the posterior probabilities for each GCM from all possible models called PIP in step 2.
4. Decide about the models having higher PIPs.

As the criterion is probability; therefore, we prefer models with higher PIPs than those with lower PIPs. Similarly, this procedure can be used for model selection in other areas like hydrology, ecology, forestry, etc.

2.4.3. Ensemble projections

Normally, it is assumed in standard regression modeling that a single model be the true model to examine the response variable given a set covariates, but other probable models could give different outcomes for the same problem at hand. The typical approach, which means conditioning on a single model supposed to be true, nevertheless, it does not take account of model uncertainties. One way is to compute an arithmetic ensemble mean (AEM) as a prediction as this could provide better results than any of the single model's output; however, this approach gives no information about the uncertainty that the predictions have [21]. BMA overcomes this issue by estimating the regression models using all possible combinations of covariates given in Eq. (15) and then builds a weighted average model from all possible models. Thus, it provides probabilistic projections where the weights are the posterior probabilities during the training period, and these are directly tied to the performance of the models [20, 22–24]. The predictive probability density function (PPDF) of BMA of a variable of interest is the weighted average of PDFs of individual forecasts where the weights are the posterior model probabilities [20]. The performance of BMA is considered better in different areas such as ground water modeling, weather forecast, hydrological predictions, and model uncertainty analysis [25–31]. Suppose we have a set of k covariates (different GCMs' outputs in this study), then there are 2^k statistical models $M_0, \ldots, M_s$. Then, the conditional forecast PDF of the variable of interest on the basis of training data D (observational data) is presented in Eq. (16).

$$p(y \mid D) = \sum_{j=0}^{s} p\left(y \mid M_j, D\right).p\left(M_j \mid D\right) \tag{16}$$

where $p\left(y \mid M_j\right)$ is the forecast PDF based on model M_j, and $p\left(M_j \mid D\right)$ is the corresponding posterior probability used as a weight; consequently, it reflects how well the model fits the data during the training time period. As the weights are posterior probabilities presented in Eq. (16), therefore $\sum_{j=0}^{s} p\left(M_j \mid D\right) = 1$. The posterior mean and variance of PDF in Eq. (16) can be easily calculated and are given in Eqs. (17) and (18), respectively [24].

$$E(y \mid D) = \sum_{j=0}^{s} E\left(y \mid M_j, D\right).p\left(M_j \mid D\right) = \sum_{j=0}^{s} \mu_j.w_j \tag{17}$$

$$Var(y \mid D) = \sum_{j=0}^{s} \left(\mu_j - \sum_{i=0}^{s} w_i.\mu_i\right)^2 + \sum_{j=0}^{s} w_j.\sigma^2{}_j \tag{18}$$

$$\text{Predictive variance} = \text{Between Model Variance} + \text{Within Model Variance} \tag{19}$$

In Eq. (18), the predictive variance has two parts, one is the between models variance, and the other one is the within model variance [20]. The variance $\sigma^2{}_j$ is associated with the j^{th} model's prediction. The between model variance indicates how the individual model mean predictions deviate from the ensemble prediction, with all contributions to deviations weighted by posterior model weights [26]. The within model variance represents the individual model contributions weighted by the corresponding posterior model probability. Whether to consider only one of them or both depends on the objectives of the study. In the example

discussed in results and discussion section, we are not taking into account the second term as the objective is the ensemble assessment of climate change as suggested by [26]. As the prediction mean and variance of the forecasted PDF are available now, the prediction interval can be constructed using Eq. (20).

$$\mu_{pr} \pm z_{\left(1-\frac{\alpha}{2}\right)} \cdot \sqrt{Var_{pr}} \tag{20}$$

Here, μ_{pr}, Var_{pr}, and $z_{\left(1-\frac{\alpha}{2}\right)}$ are ensemble mean, variance, and the $\left(1-\frac{\alpha}{2}\right)$-quantile of the standard normal distribution, respectively. The subscript pr with μ_{pr}, Var_{pr} means that these statistics are about the predicted PDF.

2.4.4. Taylor diagram

Taylor diagrams are rather sophisticated diagrams for graphical evaluation of a system, process or phenomenon. It was invented by [31] in 1996; however, it was published later in 2001 to aid researchers in comparative assessment of the performance of different models. The diagram is used to quantify the degree of correspondence between observational and modeled data sets in term of three statistics, standard deviation (SD), root mean square error (RMSE), and correlation coefficient (CC). We used the R software system to create the Taylor diagrams (a) for maximum temperature, (b) for minimum temperature, and (c) for precipitation presented in **Figure 6**. The data on all these variables are taken from northern Pakistan.

The interpretation of Taylor diagrams is straight forward; however, sometimes it feels tricky and needs basic understanding of statistics. The model's performance is considered better if the modeled and observed data have strong correlation, and the modeled data have low RMSE and have closer standard deviation to that of the observational data.

3. Results and discussion

This section provides analysis and results about the methodology presented in Section 2. Each section presented in methodology section is explained with examples and detailed discussion.

3.1. Statistical bias correction

Figure 2 presents the results about evaluation of statistical bias correction methods applied to temperature and rainfall data taken from Northern Pakistan. In **Figure 2**, observed represents observed data, Sim-baseline is model's simulated data for the base line period (1960–1990), and Sim-BC stands for model's simulated data after the application of statistical bias correction techniques. For comparison, we keep the time duration of both data sets (simulated and observed) same (1960–1990). It can be seen from **Figure 2** that there are marked differences between observed and regional climate model's simulated data for temperature and precipitation in Northern Pakistan. The left panel displays maximum temperature, while the right panel is about precipitation. In both parts, original simulated results deviated from observational data; however, after the application of statistical bias correction techniques, the

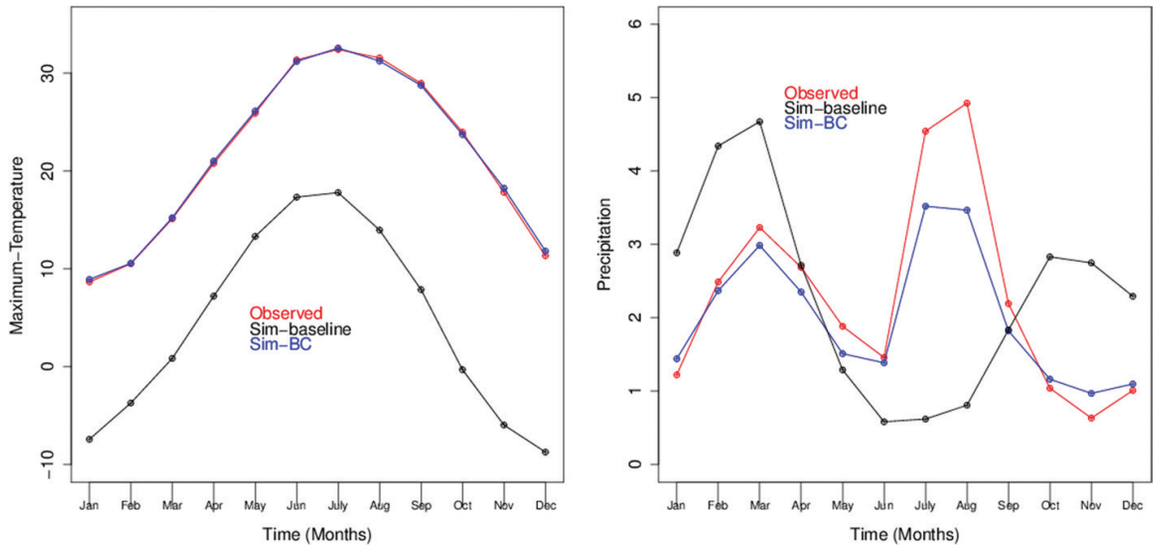


Figure 2. Comparison of observed, simulated, and bias corrected maximum temperature and precipitation for Northern Pakistan. In both parts, the red, black, and blue colors show observed, model simulated, and bias corrected data.

differences are reduced, and the pattern is followed in a better way, particularly for precipitation, where we have quite a large difference between observed and simulated precipitation data. This example is about 30 years averaged data for each month; however, these techniques can be applied to data on other frequencies like daily, hourly, etc.

3.2. Meta analysis

In **Figure 3**, climate change scenarios were analyzed, and the same procedures can be performed for model selection depending on the researcher's objective. In **Figure 3**, on the left side under the heading, studies 1, 2, 3, and 4 represent the A2, B2, RCP4.5, and RCP8.5 scenarios, respectively. The former two are chosen from the fourth assessment report (AR4), while the latter are the two scenarios stem from the fifth assessment report (AR5) of the Intergovernmental Panel on Climate Change (IPCC). In this study, scenario analysis was performed for the mean difference between baseline and future time period. The subheadings total, mean, and SD under experimental and control stand for total number of observation included in a particular scenario, mean value of each scenario, and standard deviation of each scenario, respectively. Experimental and control stand for the baseline period and future time period, respectively. The thick black vertical line shows no difference between the mean values of experimental and control periods. The dotted vertical line shows the combined mean difference. Against study 1, there is outcome effect mean difference for scenario 1, similarly for other scenarios. The length of the line on each box shows the width of the confidence interval, and the size of the box (square shape) shows the weight assigned to a particular study. As the weights are assigned on the basis of precision of a scenario (in our case), therefore, the scenarios receiving higher weights have less variance and consequently exhibit shorter confidence intervals. The bigger the square box the higher the weights assigned to a particular scenario. The two diamond shapes represent combined mean differences where the upper one is for FEM, while the lower one is for REM. The width of the diamond shows the

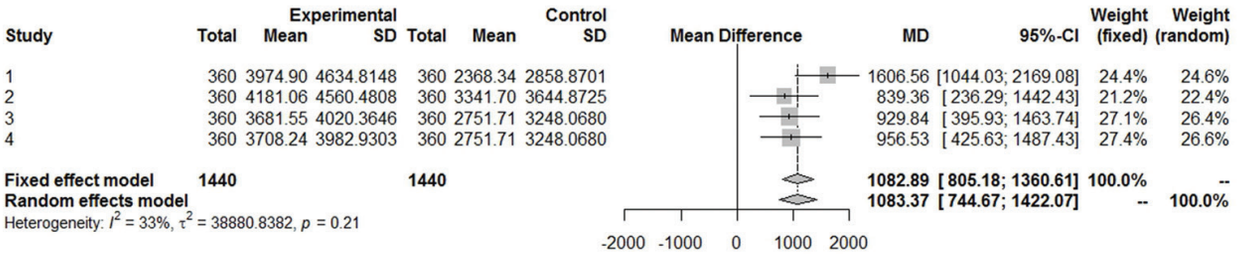


Figure 3. Here, studies 1, 2, 3, and 4 represent A2, B2, RCP4.5, and RCP8.5 scenarios, respectively. Meta analysis was conducted for mean difference between experimental and control time periods.

confidence interval under the selected model, and it is obvious that the width of the diamond for REM is larger than that of FEM. The reason for this difference is that REM also considered the variation between studies (Eq. (9)), while a FEM does not consider this variation (Eq. (8)). Bear in mind that two R packages "metafor" and "meta" have been used for this analysis. For details about these packages, we refer to [11, 12], respectively.

3.3. GCM selection and ensemble projection

3.3.1. GCM selection

Figure 4 presents the output of model selection results for 13 different GCMs where the criterion of model selection is PIP. To our knowledge, this technique is used for the first time in atmospheric sciences for the purpose of GCMs selection. The model selection procedure was run multiple times (six times here) with small changes in sample size; however, no significant changes have been noted in the end results. In **Figure 4**, the results are presented only for maximum temperature; nevertheless, the same procedures can be used for other variables.

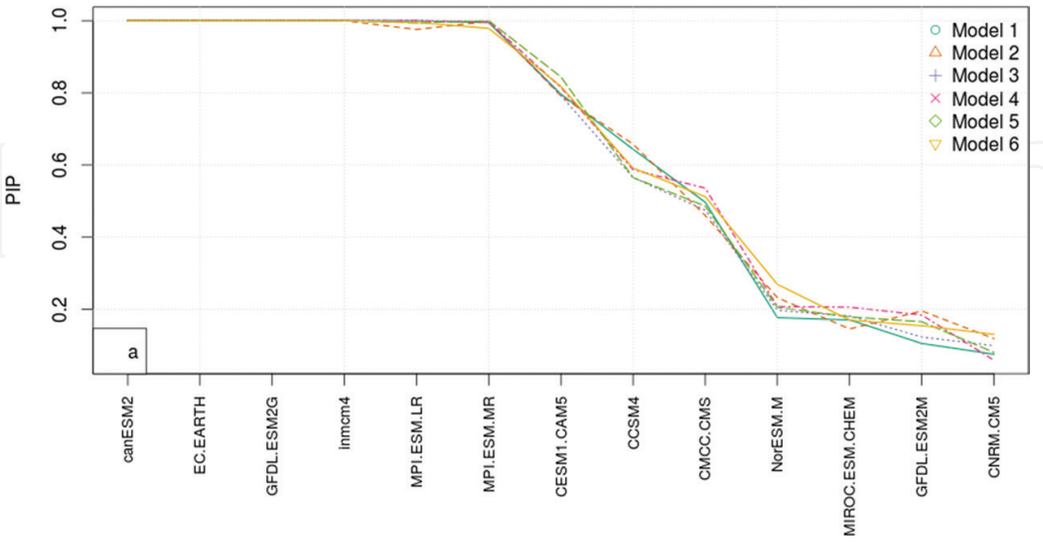


Figure 4. GCM selection among competing models for maximum temperature using posterior inclusion probability (PIP). PIPs are given on the vertical axis, and models (GCMs) are indexed on the horizontal axis. Model 1-model 6 stand for different runs of BMA with different sample sizes.

For maximum temperature, top five selected GCMs are canESM2, CESM-CAM5, EC-EARTH, GFDL-ESM2G, and INM-CM4 in ascending order. It was investigated that different variables (minimum temperature and precipitation in our case) have different lists of top GCM; however, they shared some models in the top list. The list of top five models has maximum PIPs and can be used for further analysis rather than to use all GCMs.

3.3.2. Ensemble projections

Figure 5 elaborates the results of BMAs' outputs calculated from selected GCMs compared with observational data sets for three key climate variables, maximum temperature, minimum temperature, and precipitation for the duration of 30 years (1975–2005) considered as baseline period here. As BMA is a regression-based approach, it estimates the mean value and underestimates variation. To cope with this issue, 90% prediction intervals were calculated for each variable and plotted. In **Figure 5**, red, blue, and deep gray colors represent observational, BMAs' outputs, and 90% prediction intervals for each variable. From the upper panel, it is difficult to get a clear conclusion as it is for 30-year data; however, from the bottom panel, clear information can be inferred. We can see from parts d (maximum temperature) and e (minimum temperature) that BMAs' outputs follow the pattern nicely; however, it does not capture well the variability in both cases. The 90% prediction intervals cover almost all the variation for both variables. For precipitation, we calculated upper 90% prediction intervals, and it can be seen that it covers the observational values; however, quite a few values are still outside because precipitation is a more complex phenomenon than the temperature. Similarly, we can calculate ensemble projections from various models or climate change scenarios in hydrology, agriculture, ecology, etc. to address the uncertainty while relying on the single model's output.

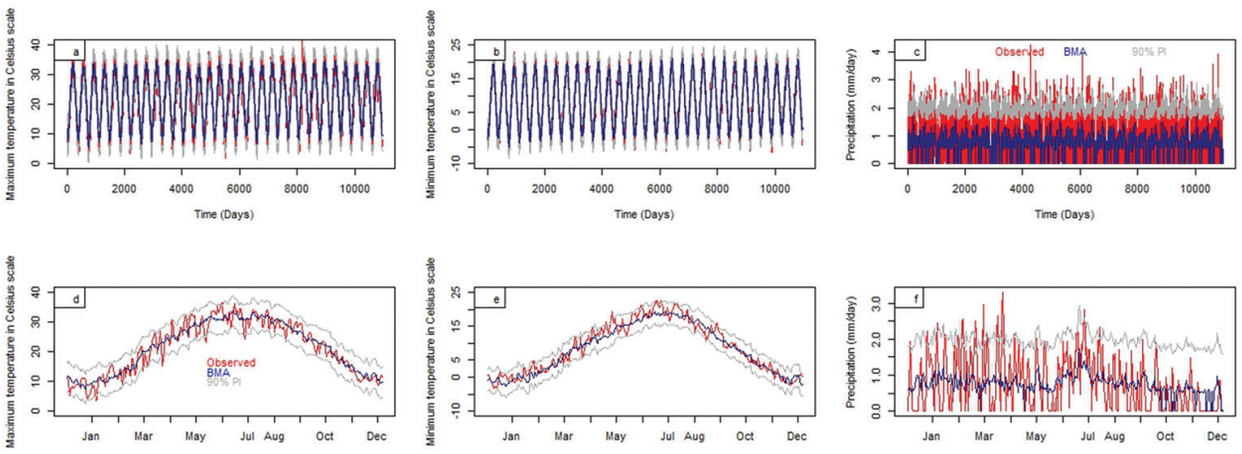


Figure 5. The upper panel demonstrates the evaluation of BMAs' outputs from selected GCMs for a period of 30 years (1975–2005) (a) maximum temperature, (b) minimum temperature, and (c) precipitation. Red, blue, and deep gray colors represent observed, BMA, and 90% prediction interval for each variable. The bottom panel demonstrates the same as in the upper panel for just 1 year.

http://dx.doi.org/10.5772/intechopen.80984

3.4. Taylor diagram

The green lines inside the graph (**Figure 6**) show RMSE between observed and modeled data sets, while the blue lines show standard deviation of each data set. The straight lines inside the graph show the correlation coefficient between modeled and observational data sets. The dots on the horizontal lines represent observed data in each part of **Figure 6**. Look at part (a) of **Figure 6** which is about maximum temperature and the purpose is to assess which model's simulated data is best as compared to other models. We calculated ensemble projections by using the outputs from all selected GCMs by using BMA and plotted them in each part of **Figure 6**. We can see that the BMA performance is superior to the performance of individual model's output. The output of BMA has higher correlation with observed data than that of individual GCM's outputs. The BMA's result also has less standard deviation than individual model outputs and smaller RMSE than the individual GCMs' outputs. The other parts of **Figure 6** can be interpreted similarly. In the same way, other models can be evaluated with other variables mentioned in the above example and can help in model's selection in different areas.

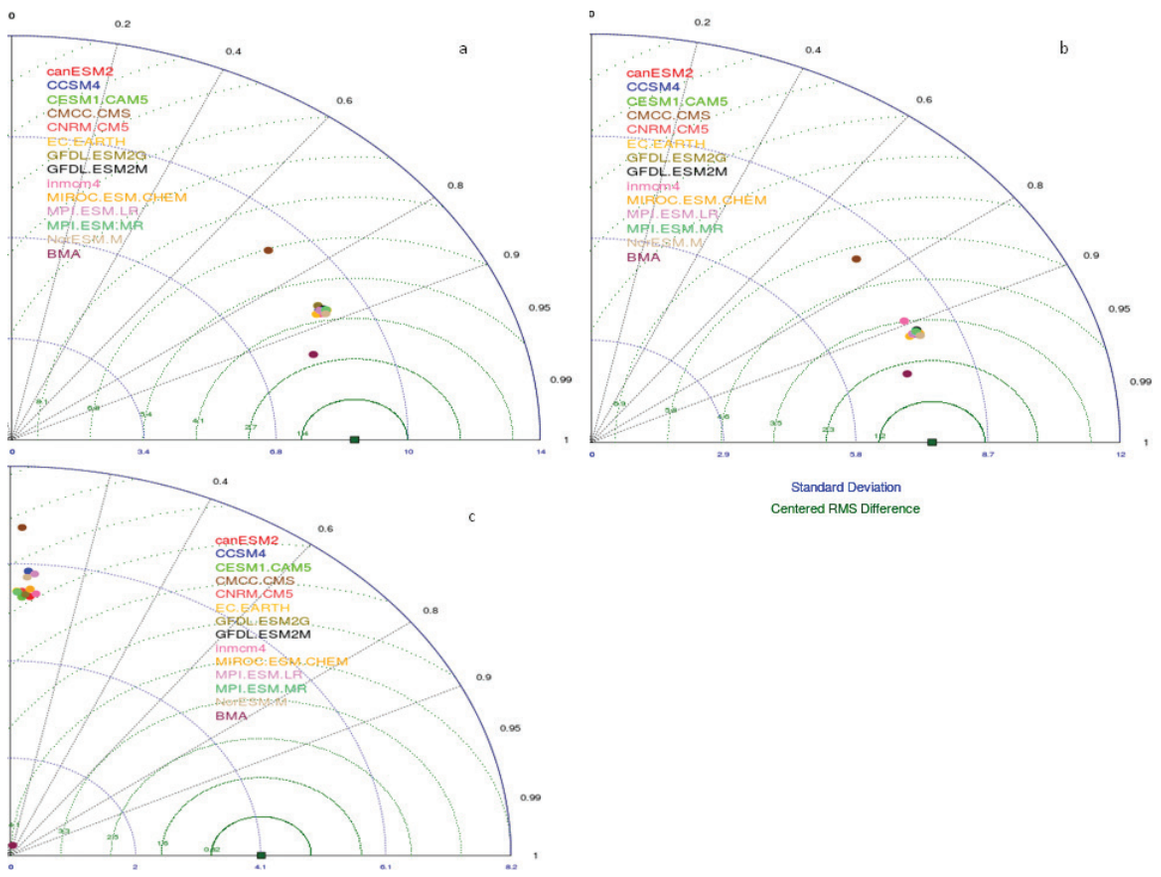


Figure 6. Taylor diagrams for evaluation of model and observational data for (a) maximum temperature, (b) minimum temperature, and (c) precipitation. Thirteen different GCMs' outputs and BMA's outputs are evaluated in this study for the Northern part of Pakistan.

4. Summary and conclusion

This study aims to present statistical methodology for evaluating process-based climate models. Different techniques have been presented for this purpose including statistical bias correction, meta analysis, model selection, ensemble projections, and Taylor diagram. The application of statistical bias correction bridged regional climate model's simulated and observational data. The performance of bias correction technique is better for temperature than precipitation; however, bias-corrected precipitation follows the observed precipitation's pattern nicely. Meta analysis can be used for different purposes like model selection, scenario analysis, etc. In this study, meta analysis is used for scenario analysis by considering four different scenarios two each from fourth assessment report (AR4) and fifth assessment report (AR5) of the IPCC. Meta analysis shows higher confidence in RCP projections and assigned higher weights on the basis of their precision. GCM's selection is of course important part in climate change assessment as there are many GCMs available. BMA is used for this purpose, and the results show that different variables have different ranks for different GCMs; however, they shared some GCMs in the list of best models. On the basis of selected GCM, ensemble projections were calculated using BMA technique. The results of GCMs and BMA's outputs were then evaluated by using Taylor diagram. Evaluation statistics used in Taylor diagram are root mean square error, correlation coefficient, and standard deviation of each data set. The evaluation confirms that ensemble projections are better than individual GCMs' outputs; nevertheless, we need to conduct this type of studies at different locations and then can make recommendations on the basis of their results.

Glossary

AEM	arithmetic ensemble mean
BMA	Bayesian model averaging
CC	coefficient of correlation
CORDEX	COordinated Regional climate Downscaling Experiment
FEM	fixed effect model
GCISC	Global Change Impact Studies Centre
GCM	global circulation/climate model
IPCC	Intergovernmental Panel on Climate Change
PDF	probability density function
PIP	posterior inclusion probability
PMD	Pakistan Meteorological Department
PPDF	predictive probability density function
RCM	regional climate model

REM	random effect model
RMSE	root mean square error
SD	standard deviation
UK	the United Kingdom

Author details

Firdos Khan[1,2]* and Jürgen Pilz[2]

*Address all correspondence to: fkyousafzai@gmail.com

1 Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan

2 Department of Statistics, University of Klagenfurt, Klagenfurt, Austria

References

[1] Intergovernmental Panel on Climate Change. Available from: http://www.ipcc-data.org/guidelines/pages/gcm_guide.html [Accessed: 06-05-2018]

[2] Box GEP, Jenkins S. Time Series Analysis, Forecasting and Control. San Francisco: Olden Day; 1970

[3] Räty O, Räisänen J, Ylhäisi JS. Evaluation of delta change and bias correction methods for future daily precipitation: Intermodel cross-validation using ENSEMBLES simulations. Climate Dynamics. 2014;**42**:2287-2303. DOI: 10.1007/s00382-014-2130-8

[4] Rivington M, Matthews KB, Bellocchi G, Buchan K. Evaluating uncertainty introduced to process-based simulation model estimates by alternative sources of meteorological data. Agricultural Systems. 2006;**88**(2006):451-471. DOI: 10.1016/j.agsy.2005.07.004

[5] Lofan T, Dadson S, Buys G, Prudhomme C. Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods. International Journal of Climatology. 2013;**33**:1367-1381. DOI: 10.1002/joc.3518

[6] Khan F, Pilz J, Amjad M, Wiberg DA. Climate variability and its impacts on water resources in the Upper Indus Basin under IPCC climate change scenarios. International Journal of Global Warming. 2015;**8**(1):46-69. DOI: 10.1504/IJGW.2015.071583

[7] Cannon AJ, Sobie SR, Murdock TQ. Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes. Journal of Climate. 2015;**38**:6938-6959. DOI: 10.1175/JCLI-D-14-00754.1

[8] Cannon AJ. Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. Climate Dynamics. 2018;**50**(1-2):31-49. DOI: 10.1007/s00382-017-3580-6

[9] Borenstein M, Hedges LV, Higgins PT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Research Synthesis Methods. 2010;**1**: 97-111. DOI: 10.1002/jrsm.12

[10] Chen DG, Peace KE. Applied Meta-Analysis with R. Boca Raton, Florida: A Chapman & Hall Book, CRC Press, Taylor & Francis Group; 2013

[11] Schwarzer G, Carpenter JR, Rücker G. Meta-analysis with R. New York: Springer International Publishing; 2015

[12] Viechtbauer W. Conducting meta-analyses in R with the meta for package. Journal of Statistical Software. 2010;**36**(3):1-48

[13] Beven K, Binley A. The future of distributed models—Model calibration and uncertainty prediction. Hydrological Processes. 1992;**6**(3):279-298. DOI: 0.1002/hyp.3360060305

[14] Blasone RS, Verugt JA, Madsen H, Rosbjerg D, Robinson BA, Zyvoloski GA. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. Advances in Water Resources. 2008;**31**(4):630-648. DOI: 10.1016/j.advwatres.2007.12.003

[15] Freer J, Beven K, Ambroise B. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research. 1996;**32**:2161-2173. DOI: 10.1029/95WR03723

[16] Stedinger JR, Vogel RM, Lee SU, Batchelder R. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research. 2008;**44**:W00B06. DOI: 10.1029/2008WR006822

[17] Zheng Y, Keller AA. Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach. Water Resources Research. 2007;**43**:W08407. DOI: 10.1029/2006WR005345

[18] Uvsitalo L, Lehikoinen A, Helle I, Myrberg K. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling & Software. 2015;**63**:24-31. DOI: 10.1016/j.envsoft.2014.09.017

[19] Lorenz EN. Atmospheric predictability experiments with large numerical model. Tellus. 1982;**34**:505-513

[20] Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review. 2005;**133**:1155-1174. DOI: 10.1175/MWR2906.1

[21] Min SK, Simonis D, Hense A. Probabilistic climate change predictions applying Bayesian model averaging. Philosophical Transactions of the Royal Society A. 2007;**365**:2103-2116. DOI: 10.1098/rsta.2007.2070

[22] Zeugner S. Bayesian model averaging with BMS. Technical Report. Available from: http://bms.zeugner.eu/tutorials/bms.pdf; 2011

[23] Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: A tutorial. Statistical Science. 1999;**14**(4):382-417

[24] Draper D. Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society: Series B. 1995;**57**(1):45-97

[25] Huang Y. Comparison of general circulation model outputs and ensemble assessment of climate change using Bayesian approach. Global and Planetary Change. 2014;**122**:362-370

[26] Vrungt JA, Robinson BA. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research. 2007;**43**:W01411. DOI: 10.1029/2005WR004838

[27] Duan Q, Ajami NK, Gao X, Sorooshian S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources. 2007;**30**(5):1371-1386

[28] Ajami NK, Duan Q, Gao X, Sorooshian S. Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. Journal of Hydrometeorology. 2006;**7**:755-768

[29] Fernandez C, Ley E, Steel MFJ. Model uncertainty in cross-country growth regression. Journal of Applied Econometrics. 2001;**16**:563-576. DOI: 10.1002/jae.263

[30] Raftery AE, Zheng Y. Long-run performance of Bayesian model averaging. Journal of the American Statistical Association. 2003;**98**(464):931-938. DOI: 10.1198/016214503000000891

[31] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research. 2001;**106**:7183-7192. DOI: 10.1029/2000JD900719

IntechOpen

IntechOpen

IntechOpen