



Evaluation of CMIP5 models and ensemble climate projections using a Bayesian approach: a case study of the Upper Indus Basin, Pakistan

Firdos Khan¹ · Jürgen Pilz² · Shaukat Ali³

Received: 23 April 2020 / Revised: 29 November 2020 / Accepted: 24 February 2021 /
Published online: 24 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The availability of a variety of Global Climate Models (GCMs) has increased the importance of the selection of suitable GCMs for impact assessment studies. In this study, we have used Bayesian Model Averaging (BMA) for GCM(s) selection and ensemble climate projection from the output of thirteen CMIP5 GCMs for the Upper Indus Basin (UIB), Pakistan. The results show that the ranking of the top best models among thirteen GCMs is not uniform regarding maximum, minimum temperature, and precipitation. However, some models showed the best performance for all three variables. The selected GCMs were used to produce ensemble projections via BMA for maximum, minimum temperature and precipitation under RCP4.5 and RCP8.5 scenarios for the duration of 2011–2040. The ensemble projections show a higher correlation with observed data than individual GCM's output, and the BMA's prediction well captured the trend of observed data. Furthermore, the 90% prediction intervals of BMA's output closely captured the extreme values of observed data. The projected results of both RCPs were compared with the climatology of baseline duration (1981–2010) and it was noted that RCP8.5 show more changes in future temperature and precipitation compared to RCP4.5. For maximum temperature, there is more variation in monthly climatology for the duration of 2011–2040 in the first half of the year; however, under the RCP8.5, higher variation was noted during the winter season. A decrease in precipitation is projected during the months of January and August under the RCP4.5 while under RCP8.5, decrease in precipitation was noted during the months of March, May, July, August, September, and October; however, the changes (decrease/increase) are higher than under the RCP4.5.

Keywords Bayesian Model Averaging · Ensemble Projections · General Circulation Model · Posterior Inclusion Probability · Upper Indus Basin

Handling Editor: Luiz Duczmal.

Extended author information available on the last page of the article

1 Introduction

The rapid developments in computational technology make it viable to run a complex process-based climate model for simulating a climate system. However, the output from a single model still may have uncertainties (Uusitalo et al. 2015). There have been several General Circulation Models (GCMs) developed by the scientific community for the projection of future climate changes and these provide useful information for impact assessment studies (Intergovernmental Panel on Climate Change = IPCC 2013). Different GCMs may produce different results due to internal atmospheric variability (Lorenz 1982), structural and parametric differences (Neggers 2015; Baumberger et al. 2017; Achieng and Zhu 2019). Therefore, it is better to consider the output of several GCMs instead of relying on a single GCM, which will reduce the uncertainty as different climate models may simulate different aspects of climate better (Brient 2020). Regression models can be used to fit the observed climate by using outputs from different GCMs as covariates. However, it is important to select the set of best possible GCMs. In a regression model context, the problem of uncertainty modeling has been raised by Raftery et al. (2005). In such models, covariate selection is a basic part of building a valid regression model, and the objective is then to find the “best” model for the response variable given a set of predictors. The first problem to solve is which covariates should be included in the model and how important are they?

Bayesian Model Averaging (BMA) is a regression-based approach where the dependent variable is the observed climate variable of interest, e.g., maximum temperature, minimum temperature, and precipitation in this study, and covariates are the output of GCMs. Suppose there are k covariates then the total number of possible linear regression models will be 2^k . It is important to remember that in both cases, i.e., GCM selection and ensemble projections, covariates are the outputs of GCMs. Posterior Inclusion Probability (PIP) is used as GCM selection criterion where PIP is the sum of the posterior probabilities of each covariate from each regression model included in BMA. In ensemble projection, each regression model has weight equal to posterior probability during the training period (reference or baseline duration). Consequently, the model with better performance will have higher weight as compared to the model having poor performance. Details about GCM selection and ensemble projections, as well as posterior probabilities, are presented in the methodology section.

The Atmospheric-Ocean coupled GCMs (AOCMs) are the tools that provide the basic and detailed information about the global climate. Therefore, it is important to choose better GCMs among a set of available models for impact assessment studies (Themßl et al. 2011; Maraun 2013; Mendlik and Gobiet 2016). There are only few studies that focus on this topic, e.g., Mendlik and Gobiet (2016) presented an approach based on three steps: find common patterns of climate change using principal component analysis, using cluster analysis to detect model similarities, generate subgroups of representative models by sampling from clusters. McSweeney et al. (2015) used decision matrices over three steps which have four classification categories about the model performances on

three different parts of the Globe, Southeast Asia, Europe, and Africa. Further studies include McSweeney et al. (2012), Smith and Hulme (1998), and Salathe et al. (2007). The skill of GCMs is normally assessed on the basis of some global phenomena e.g., monsoon or ENSO. In this study, initially the GCMs are selected on the basis of available literature (Ashfaq et al. 2017), skills score, warm, dry, wet, cold criteria (Lutz et al. 2016). For the final selection of GCMs, we propose BMA, a probabilistic approach where the posterior probability of each GCM in each regression model being included in model averaging is used as a model selection criterion. Ensemble climate projections are developed based on the output of selected GCMs.

Ensemble projections provide more information because of their reliability and take account of the uncertainties, which are particularly important for impact assessment studies (Leedale et al. 2016; Kumar et al. 2012; Patricola et al. 2012; Stedinger et al. 2008; Reum et al. 2020). One way of computing ensemble data is arithmetic ensemble mean (AEM), which produces better results than those based on single models. However, this approach lacks uncertainties in predictions (Min et al. 2007). BMA estimates regression models using all possible combinations of covariates and then builds a weighted average model from all of them to produce ensemble projections (Zeugner 2011). There are various studies that recently focused on BMA for different purposes. Achieng and Zhu (2019) and Qu et al. (2017) used BMA for hydrological prediction, whereas Sloughter et al. (2007) used BMA for the weather forecast in the North American Pacific Northwest in 2003–2004. Yang et al. (2011) used BMA for bias reduction in RCMs' outputs in the Eastern-Pacific-region.

In this study, the output of thirteen GCMs is considered for predicting the observed local climate under the representative concentration pathways (RCPs). According to Eckstein et al. (2019), Pakistan is listed among the top countries vulnerable to the impacts of climate change. Therefore, the medium-range emission scenario (RCP4.5) and the high range emission scenario (RCP8.5) are being considered. The target location of this study is Upper Indus Basin (UIB), Pakistan, which has abrupt spatial variation in climate due to its unique topography. Consequently, the output of the process-based climate model has more uncertainties (Khan et al. 2015, 2017; Ali et al. 2015). Before embarking on ensemble projections, the outputs of selected GCMs were downscaled from coarser to finer resolution using a statistical approach called Bias Correction and Spatial Disaggregation (BCSD). The details are presented in the methodology section of this paper and various steps of the study are schematically presented in Fig. 2. For analysis and visualization, R software with packages BAS (Clyde et al. 2020), BMA (Raftery et al. 2020), BMS (Feldkircher and Zeugner 2020) and ensembleBMA (Fraley et al. 2018) have been used.

The major aims of this study are: to select a set of best possible GCM(s) for the area under study from the Coupled Model Intercomparison Project Phase-5 (CMIP5); to calculate ensemble projections using BMA on the basis of selected GCMs' outputs. In addition, this study may be helpful for impact assessment studies e.g. in agriculture, reservoir management, flood forecasting, hydropower generation and consequently may help to achieve some of the Sustainable Development Goals (SDGs) of the United Nations Development Program (UNDP).

2 Data and study area

Daily observed and GCMs simulated meteorological data, including maximum, minimum temperature and precipitation over the UIB, Pakistan, have been used. Observational data of climate variables have been collected from the Pakistan Meteorological Department (PMD) for the duration of 1981–2010. The UIB has unique topography and contains the third largest glaciers in the world, where the elevation range is 2000–8247 m. Therefore, the meteorological stations are dispersed as it is challenging to operate it in many parts of this region. The total area of the target location is approximately 134,621 square kilometers. The GCMs' output was acquired from the World Climate Research Programme (WCRP; <https://www.wcrp-climate.org/>). For GCM selection and ensemble projections, the outputs of thirteen GCMs have been considered. The GCM's data lie on a grid; therefore, the data were extracted for the selected meteorological stations and then averaged and converted to time series for each model. Similarly, observed climate data was averaged over the study area using the data from eleven meteorological stations. The details about meteorological stations are given in Fig. 1 and Table 1, and brief information about each GCM is presented in Table 2.

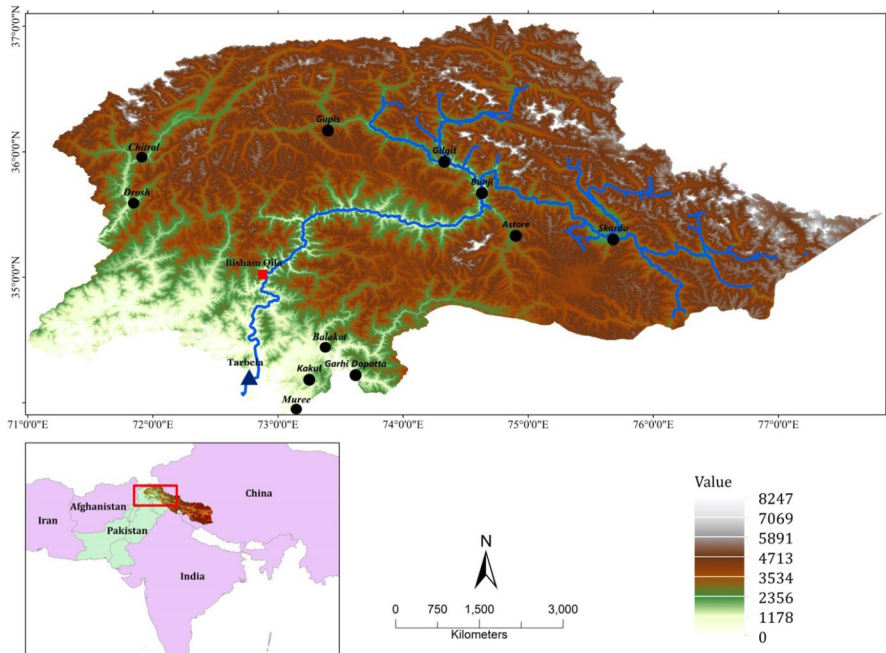


Fig. 1 The area under study comprises Northern Pakistan (Upper Indus Basin, Pakistan). The considered meteorological stations are marked with black filled circles, Indus River is indicated by a line with blue color and Tarbela Reservoir is marked with a blue color triangle

Table 1 Brief information about the meteorological stations (latitude, longitude and elevation) included in this study

S. No	Name	Latitude	Longitude	Elevation
1	Astore	35° 20'	74° 54'	2168.0 m
2	Balakot	34° 33'	72° 21'	995.4 m
3	Bunji	35° 40'	74° 38'	1372.0 m
4	Chitral	35° 51'	71° 50'	1497.8 m
5	Drosh	35° 34'	71° 47'	1463.9 m
6	Gari Dupatta	34° 13'	73° 37'	813.5 m
7	Gilgit	35° 55'	74° 20'	1460.0 m
8	Gopis	36° 10'	73° 24'	2156.0 m
9	Kakul	34° 11'	73° 15'	1308.0 m
10	Muree	33° 54'	73° 26'	2291.0 m
11	Skardu	35° 18'	75° 41'	2317.0 m

3 Methodology

The methodology is divided into three subsections, GCM(s) selection, downscaling of selected GCM(s) and ensemble projections. The details about various phases of this study are presented graphically in Fig. 2.

3.1 GCM(s) selection

Bayesian model averaging is used for model selection and producing ensemble projections which is based on posterior probability evaluations. Therefore, 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.

3.1.1 Posterior probability

Bayes' theorem states that the posterior probability of j^{th} regression model, $p(M_j|D)$, is calculated as the likelihood of observed data given j^{th} model, $p(D|M_j)$, multiplied by the prior probability $p(M_j)$ of the j^{th} regression model, divided by the (marginal) probability of having the current observation realization, $p(D)$. The posterior probability is thus given in Eq. (1):

$$P(M_j|D) = \frac{p(D|M_j) \cdot p(M_j)}{p(D)} \quad (1)$$

In Eq. (1), $p(D)$ is used as a normalizing constant given in Eq. (2),

$$p(D) = \sum_{j=1}^s p(D|M_j) \cdot p(M_j) \quad (2)$$

Table 2 Details about GCMs used in this study with their resolutions and other brief information

S. No	Name	Resolution (Degree ^a)	Institute/Centre
1	CanESM2	2.7906 × 2.8125	Canadian Centre for Climate Modelling 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

^a1 degree is approximately equal to 111.32 km

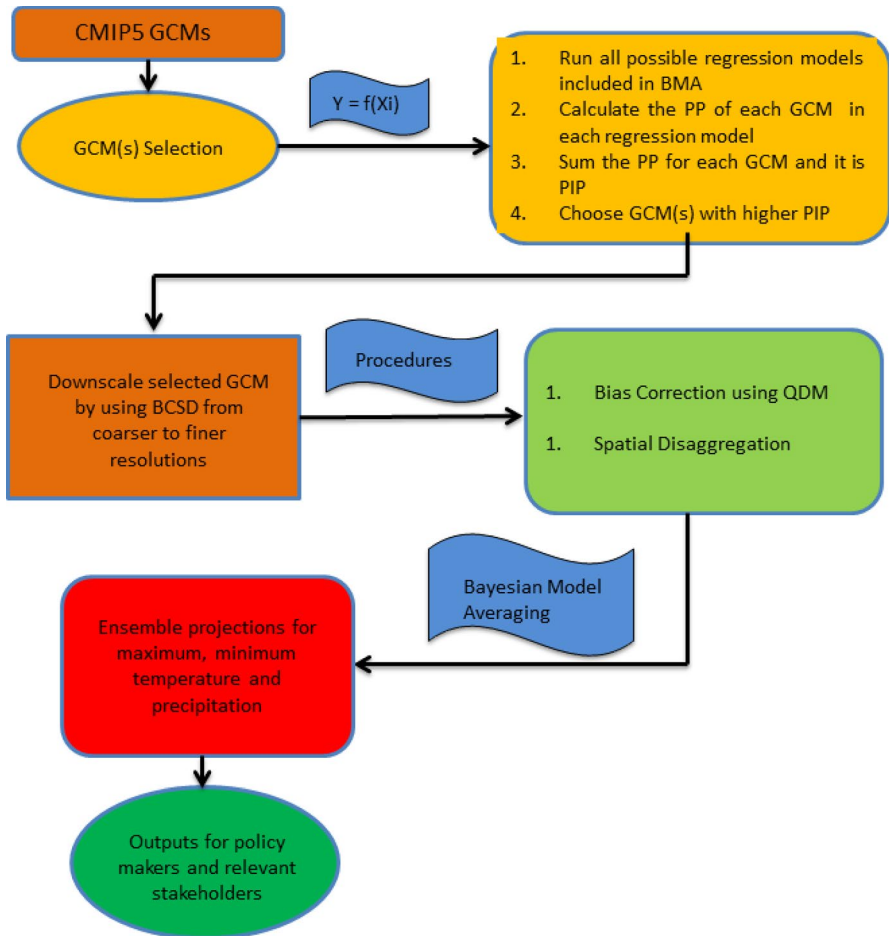


Fig. 2 The schematic diagram represents various steps followed in conducting this research work

and hence Bayes’ rule can be simply stated as in Eq. (3).

$$p(M_j|D) \propto p(D|M_j).p(M_j) \tag{3}$$

In this study we have $M_j; j = 1, 2, \dots, s = 2^k$ possible regression 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 build their own customized priors for their analysis (Zeugner 2011). In this study, a uniform prior distribution (i.e. $p(M_j) \propto 1/2^k$) has been used, which assigns equal prior weight to all models. Then the posterior model probability can be expressed as in Eq. (4).

$$p(M_j|D) \propto p(D|M_j) \quad (4)$$

Equation (4) shows that in this case the posterior probability of a model is only determined by the likelihood of observational data. The likelihood of a model is the ability to reproduce the given observed data. Different likelihood functions have been proposed in the literature, $p(D|M_j)$, see for example Blasone et al. (2001), Zheng et al. (2007), Freer et al. (1996) and Beven et al. (1992). A Gaussian likelihood proposed by Stedinger et al. (2008) is used in this study.

The same procedures are used here for GCM selection, where GCM now stands for a covariate x_k ; $k=1, \dots, m=13$; and a uniform prior (over the whole real line) was used as a prior probability distribution for the intercept term and the regression parameters β_k ($k=0, 1, \dots, 13$) corresponding to the covariates: $p(\beta_k)=1$, which corresponds to so-called prior ignorance (non-informativity) about the regression parameters.

The PIP (posterior inclusion probability) is the sum of posterior probabilities of each covariate's contribution from all possible regression models included in BMA. The PIP has a range between zero and one, where a value close to one means that the GCM contributes to explain the variability in the observed data while a value close to zero means that the corresponding GCM does not really contribute to explain the variability. The selection of GCMs was carried out in the following steps:

- i. Estimate the regression models included in BMA (with full or customized number of regression models)
- ii. Observe the posterior probability of each regression parameter in each regression model included in BMA
- iii. Calculate PIP
- iv. Decide about GCM, select if it has higher PIP compared to other GCMs.

Since precipitation usually follows a gamma distribution, therefore, Anscombe transformation (Anscombe 1948) was used to achieve approximate normality. Let Y_{or} and Y_{tr} denote original recorded and transformed precipitation, respectively, then the relationship is given by Eq. (5).

$$Y_{tr} = \sqrt[3]{Y_{or}} \quad (5)$$

3.2 Downscaling the selected GCM(s)

For downscaling the outputs of selected GCM(s), the statistical approach BCSD, which comprises of two steps, has been used. The details about both steps of BCSD are presented in the following subsections:

3.2.1 Statistical bias correction

Due to the inherent uncertainties in the models and input data, climate model outputs are biased and unable to reproduce the variation of observed data (Räty et al.

2014; Rivington et al. 2006). Various statistical methods are available in the literature to reduce the gap between simulated and observed data; here we employed the Quantile Delta Mapping (QDM) method, which preserves the relative and absolute climate change in future (Cannon et al. 2015). Before starting bias correction of GCMs' outputs, daily climate observations at the target location were first assigned to the fine-scale ($1/8^\circ \times 1/8^\circ$) grid cells. The GCM's outputs were re-gridded to $1^\circ \times 1^\circ$ resolution to make it uniform for all GCMs. The fine-scale observations were then resampled to coarse-scale regrided GCMs' resolution ($1^\circ \times 1^\circ$ for CMIP5) using a bilinear interpolation. Once the observed and simulated data were available at the same resolution, then QDM was used for bias correction.

The QDM requires four steps to implement, starting from the time dependent cumulative distribution function (CDF) of the model projected series. For the details of this method, we refer to Cannon et al. (2015), Khan and Pilz (2018) and Ali et al. (2015).

3.2.2 Spatial disaggregation

The approach of spatial disaggregation comprises of various steps. At the initial step, subtract the coarse-scale spatial pattern from the bias-corrected data which will yield a temporal scaling factor at coarser resolution of $1^\circ \times 1^\circ$. Then interpolate the temporal scaling factor from a coarser resolution of $1^\circ \times 1^\circ$ to a finer resolution of $1/8^\circ \times 1/8^\circ$. Finally, merge the temporal scaling factor with fine-scale observed climatology, as a result this will yield the spatially downscaled climate projections at fine-scale resolution ($1/8^\circ \times 1/8^\circ$). For precipitation, the fine-scale temporal factor is multiplied with fine-scale observed precipitation, while for temperature, the fine-scale temporal factor is added with the fine-scale observed climatology. For further details about the BCSD approach, we refer to Brekke et al. (2013). To obtain climate data at daily frequency, disaggregate monthly outputs to daily time scales by resampling a historical month and scaling daily data to match the monthly projections (Wood et al. 2004).

3.3 Ensemble projections

In standard regression modeling a single model is assumed to be the true model to investigate the response variable given covariates, but other plausible models could give different results for the same problem. The typical approach, which means conditioning on a single model deemed to be true, does not take account of model uncertainty. BMA is a statistical technique that estimates the regression model using all possible combinations of covariates and then builds a weighted average over all of them. Thus, it provides probabilistic projections (Zeugner 2011; Hoeting et al. 1999; Draper 1995). The BMA predictive probability density function of a quantity of interest is the weighted average of PDFs of individual forecasts, and the weights are the posterior model probabilities directly tied to the performance of the model during the training period (Raftery et al. 2005). The BMA's performance has been studied in various areas such as hydrological predictions, weather forecast,

groundwater modeling, and model uncertainty (Li and Shi 2010; Zhang et al. 1999; Rojas et al. 2008; Verugt et al. 2007; Duan et al. 2007; Ajami et al. 2006; Fernandez et al. 2001; Raftery and Zheng 2003, Slaughter et al. 2007).

More recently, Achieng and Zhu (2019) and Qu et al. (2017) used BMA for hydrological prediction, Yang et al. (2011) used BMA for bias reduction in RCMs' outputs in the Eastern Pacific-region and Huang (2014) used BMA for model uncertainty evaluations.

Suppose, we have k covariates (GCMs in this study), then there are $s = 2^k$ regression models M_1, \dots, M_s , the conditional forecast PDF of the variable to be forecast on the basis of training data D is given by Eq. (6).

$$p(y|D) = \sum_{j=1}^s p(y|M_j, D) \cdot p(M_j|D) \quad (6)$$

where $p(y|M_j)$ is the forecast PDF based on model M_j and $p(M_j|D)$ is the corresponding posterior probability used as a weight; consequently it reflects how the model fits the training data. The weights are posterior probabilities, therefore $\sum_{j=1}^s p(M_j|D) = 1$. The posterior mean and variance can be easily calculated and are given in Eqs. (7) and (8), respectively (Li and Shi 2010; Zhang et al. 2009; Rojas et al. 2008).

$$E(y|D) = \sum_{j=1}^s E(y|M_j, D) \cdot p(M_j|D) = \sum_{j=1}^s \mu_j \cdot w_j \quad (7)$$

$$Var(y|D) = \sum_{j=1}^s \left(\mu_j - \sum_{i=1}^s w_i \cdot \mu_i \right)^2 + \sum_{j=1}^s w_j \cdot \sigma_j^2 \quad (8)$$

Here μ_j denotes the predictive mean based on the j^{th} model, w_j stands for the posterior probability of M_j and σ_j^2 denotes the variance associated with the j^{th} model's prediction. It is seen from Eq. (8) that the predictive variance has two parts; one is the between model variance and the other one is the within model variance (Raftery 1993). 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 (Rojas et al. 2008, Draper 1995). The within model variance represents the individual model contributions weighted by the corresponding posterior model probability. Therefore, in this study, we are not taking into account the second term as the objective is to ensemble assessments of climate change under recent emission scenarios as suggested by Huang (2014). As the prediction mean and variance are available now, the confidence interval for the predictive mean can be constructed using Eq. (9),

$$\mu_{pr} \pm z_{\left(1-\frac{\alpha}{2}\right)} \cdot \sqrt{Var_{pr}} \quad (9)$$

Here μ_{pr} , Var_{pr} and $z_{(1-\frac{\alpha}{2})}$ are the ensemble prediction mean and variance as given in Eqs. (7) and (8), respectively, and $z_{(1-\frac{\alpha}{2})}$ denotes the $(1 - \frac{\alpha}{2})$ -quantile of the standard normal distribution.

4 Results and discussion

The results and discussion part is divided into three subparts, GCM selection, evaluation of the climate models and future climate projections.

4.1 GCM selection

The results about model selection are presented in Table 3 and Fig. 3, using 6 experiments with different sample sizes. In Fig. 3a-c the results of model selection

Table 3 This table presents the selection of GCM(s) for three climate variables under six experiments on the basis of Posterior Inclusion Probabilities of each GCM

Variable	Model/Experiment GCM	1	2	3	4	5	6
		Posterior Inclusion Probabilities					
Maximum Temp	canESM2	1	1	1	1	1	1
	EC.EARTH	1	1	1	1	1	1
	GFDL.ESM2G	1	1	1	1	1	1
	Inmcm4	1	1	1	1	1	1
	MPI.ESM.LR	1	1	1	1	1	0.98
	MPI.ESM.MR	1	1	1	1	1	1
	CESM1.CAM5	0.78	0.76	0.74	0.74	0.74	0.74
Minimum Temp	canESM2	1	1	1	1	1	1
	EC.EARTH	1	1	1	1	1	1
	GFDL.ESM2G	1	1	1	1	1	1
	Nor.ESM.M	1	1	1	1	1	1
	CCSM4	1	1	1	1	1	0.99
	CESM1.CAM5	1	1	1	1	1	1
	GFDL.ESM2M	0.99	1	1	1	0.99	1
Precipitation	canESM2	1	1	1	0.99	1	1
	CCSM4	1	1	1	0.99	1	1
	Inmcm4	1	1	1	1	1	1
	MICRO.ESM.CHEM	1	1	1	1	1	1
	MPI.ESM.LR	1	1	1	1	1	1
	NorESM.M	0.93	0.99	0.94	0.98	0.99	0.99
	CMCC.CMS	0.90	0.88	0.83	0.90	0.90	0.89

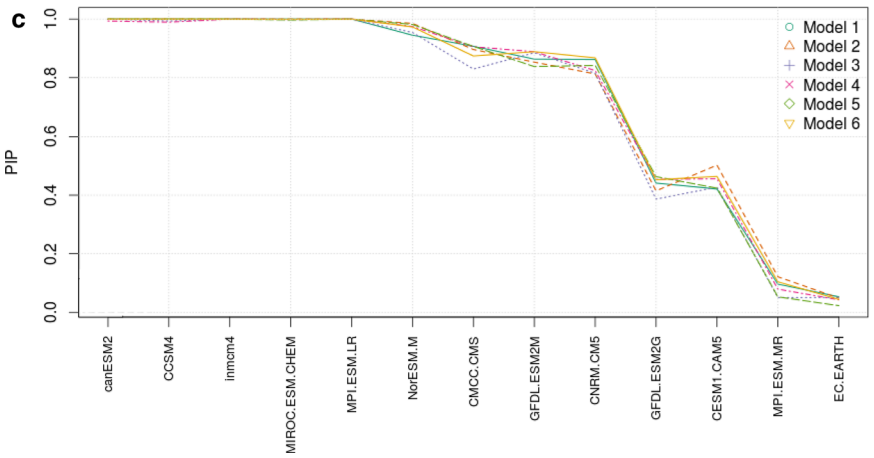
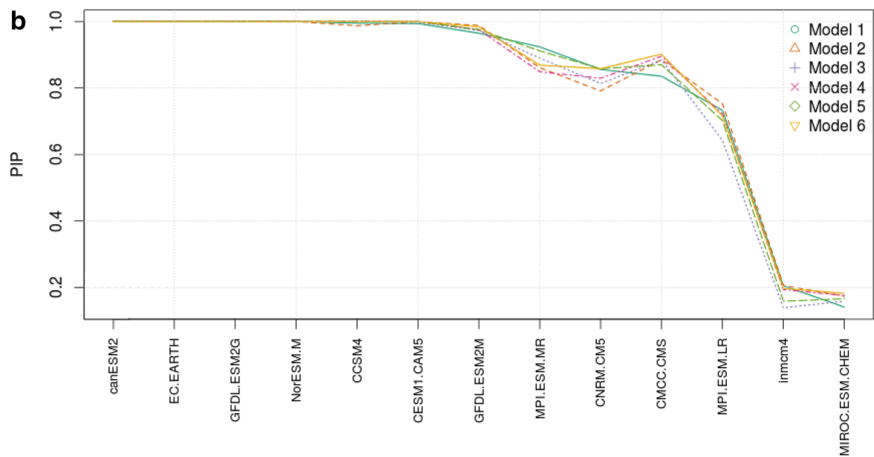
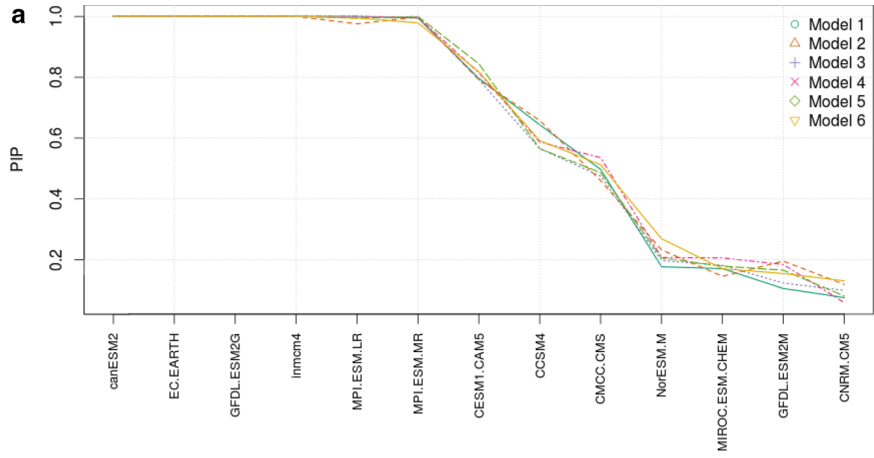
PIP takes values between zero and one where values near to one mean high probability and those near to zero indicate low probability

Fig. 3 The posterior inclusion probabilities using different variables to select a suitable GCM. **a** maximum temperature, **b** minimum temperature, **c** precipitation with six different models (experiments). On the x axis different GCMs are presented while on the y axis the posterior inclusion probability (PIP) is given. *Note* Users can customize their own prior: <http://bms.zeugner.eu/custompriors.php>

among all $m=13$ GCMs, whereas in Table 3, the best seven models are given. The results in Fig. 3a-c and Table 3 suggest that CanESM2, EC.EARTH, inmcm4 and GFDL-ESM2G are the models which performed well for maximum and minimum temperature. Precipitation shares CanESM2 as the best model jointly with temperature while the other GCMs which performed well are CCSM4, inmcm4, MICRO.ESM.CHEM and MPI.ESM.LR. The results are quite interesting and suggest that the performance of different GCMs varies, depending on the variable of interest. For downscaling temperature, the same GCM can be used for maximum and minimum temperature, while for precipitation, we might be well advised to choose a different one. Nevertheless, this is an effort in a new direction for model (GCMs) selection and we believe that it will help researchers dealing with model or scenario selection for their studies of climate change and impact assessment.

4.2 Evaluation of BMA's and GCMs' output

Figure 4 shows a comparison of BMA's results for the baseline duration and individual GCMs' output with observational data by using three different statistical criteria, i.e. correlation coefficient, Root Mean Square Error (RMSE), standard deviation of individual data set. The results in Table 4 present correlation coefficients of BMA predicted data with observed data and only maximum and minimum values of correlation coefficients are presented for individual GCMs' outputs with observed data. Table 4 shows that the BMA predicted climate data has stronger correlation with observed data compared to that of individual GCMs' outputs. Figures 4a-c show evaluations of individual GCMs output and BMA predicted data with observational data by utilizing the Taylor (Taylor 2001) diagram for maximum, minimum temperature and precipitation, respectively. The results of these GCMs' outputs are close to each other, however, CMCC.CMS has larger deviation from observational data as compared to other GCMs. CMCC.CMS has higher RMSE and weaker correlation with observed data in comparison to other GCMs. It also gives clear indications that the models' output has larger deviations under maximum temperature as compared to minimum temperature and, subsequently, the BMA output follows the same trend. Inter-comparison of GCMs output shows that there is less variation in the outputs of different GCMs for maximum and minimum temperature, except CMCC.CMS, while for precipitation, the variation is higher. All GCMs' outputs underestimate the observational variability slightly for maximum and minimum temperature, and considerably for precipitation. Figure 4 shows that the BMA predicted data has higher correlation ($r=0.95$ for maximum and minimum temperature and 0.2 for precipitation) with observational data than individual GCMs output. The Root Mean Square Error (RMSE) results show that the BMA predicted climate data has smaller values for all climate variables than individual GCMs output, where we have values of



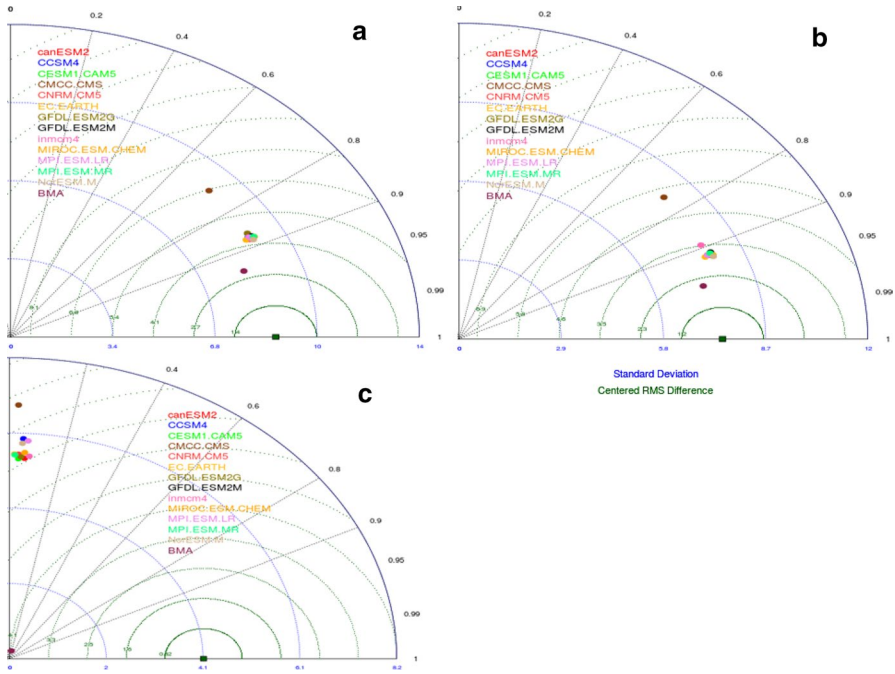


Fig. 4 Taylor diagram presenting a comparison between models and observational data on the basis of the correlation coefficient, Root Mean Square Error (RMSE) and standard deviation. Where **a** maximum temperature, **b** minimum temperature **c** precipitation. The comparison included thirteen different GCMs’ outputs and BMA predicted climate data

Table 4 Comparison of correlation coefficients and Nash–Sutcliffe efficiency coefficients between observed, BMA predicted climate variables and individual GCM bias corrected outputs

Correlation	MaxTMP	MinTMP	Precipitation
<i>Correlation coefficients</i>			
BMA	0.94	0.96	0.28
Max	0.89	0.92	0.13
Min	0.72	0.74	0.09
<i>Nash–Sutcliffe efficiency coefficients</i>			
BMA	0.88	0.93	0.08
Max	0.76	0.83	– 1.01
Min	0.41	0.46	– 1.06

Where Max and Min indicate the lower and higher correlation coefficients between individual GCM outputs and the observed climate data

2.9 °C, 2.3 °C for maximum temperature, minimum temperature, respectively, and 4.1 mm day⁻¹ for precipitation. The results show that CMCC.CMS stands last w.r.t performance as compared to other GCMs. It was observed during the comparison that both statistics, correlation coefficient and Nash–Sutcliffe coefficient, have different trends for temperature and precipitation. CMCC-CMS has lower correlation

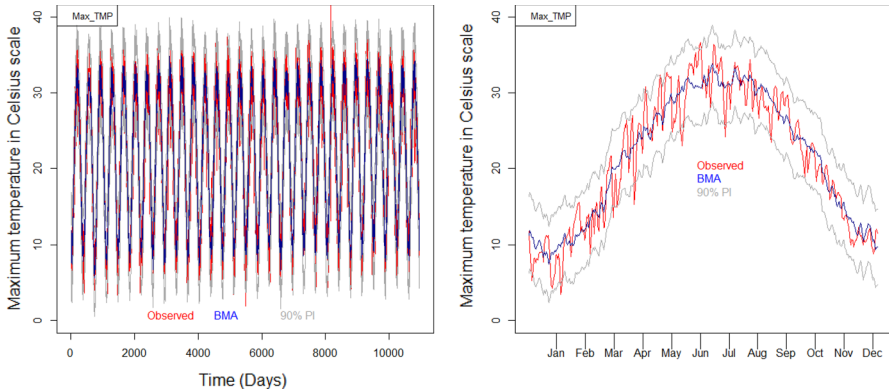


Fig. 5 Comparison of predicted and observed values for maximum temperature for the duration of 1980–2010 on the basis of daily frequencies. Red, blue and dark-grey colors represent observed values, BMA predicted values and 90% prediction intervals, respectively. Left side panel is for thirty years and right side is for one year to depict detailed information

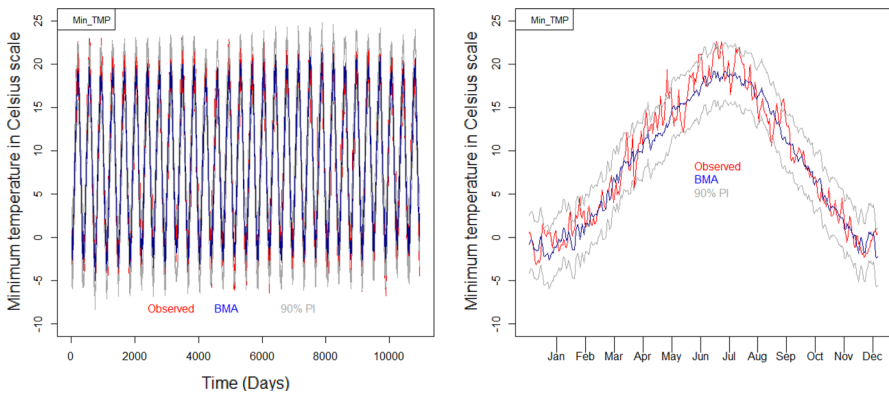


Fig. 6 Comparison of predicted and observed for minimum temperature for the duration of 1980–2010 on the basis of daily frequencies. Red, blue and dark-grey colors represent observed, BMA predicted and 90% prediction interval, respectively. Left side panel is for thirty years and right side is for one year to depict detailed information

and Nash–Sutcliffe coefficients for maximum and minimum temperature, while it behaves differently for precipitation. The reason is possibly the parameterization scheme of CMCC.CMS, which is better for precipitation than for temperature, but it needs further experiments and tests to validate this point. The results illustrate that the trend of observed data is followed better by BMA predicted data than by using individual GCM output.

Figures 5, 6, and 7 show the results about observed and BMA predicted maximum, minimum temperature and precipitation, respectively, along with their 90% prediction intervals for the duration of 1981–2010 on a daily frequency. In each of these figures the red, blue and dark grey colors represent observed

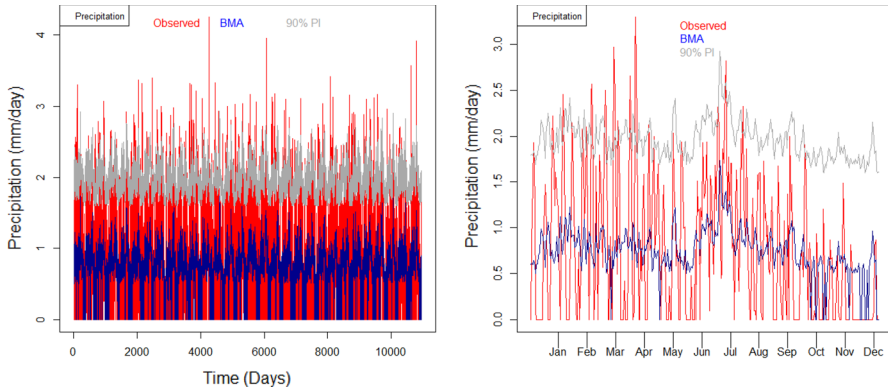


Fig. 7 Comparison of predicted and observed values for precipitation for the duration of 1980–2010 on the basis of daily frequencies. Red, blue and dark-grey colors represent observed, BMA predicted and 90% prediction intervals, respectively. Left side panel is for thirty years and right side is for one year to depict detailed information

data, BMA predicted data and 90% prediction intervals, respectively. It can be seen from Fig. 5 that BMA's predicted maximum temperature follows the trend of observed temperature, but the extreme values are not captured well. To overcome this issue, 90% prediction intervals have been calculated which covered almost all values of observed maximum temperature. Figure 6 represents the observed, BMA predicted minimum temperature, along with their 90% prediction intervals. The extreme values are mostly covered by the 90% prediction intervals. Figure 7 displays observed and BMA predicted precipitation, along with 90% prediction intervals. In Figures 5, 6, and 7 on the right side, the results are presented for one year of observed and BMA predicted climate variables, along with their 90% prediction intervals, which makes it easier and clearer to understand the predictive power of BMA. Figure 5 shows that the trend is followed precisely and prediction intervals covered most of the observed values of maximum temperature. There is more variation during the first half of the year as compared to the second half, and consequently, some values of observed maximum temperature are still outside the 90% prediction interval. Also, most of the observed values of minimum temperature are covered by the 90% prediction intervals. From Figures 5, 6, and 7 and Table 4, it can be concluded that the performance of BMA is better for minimum temperature as compared to that for maximum temperature. During the evaluation, it was noted that BMA performs better for temperature as compared to precipitation in target locations.

4.3 Future's climate projections

The results in Figs. 8, 9, and 10 show the comparison of baseline climatology to the ensemble projections by using BMA for maximum temperature, minimum temperature and precipitation, for RCP4.5 and RCP8.5, respectively. The red line shows the average climatology for the baseline (30 years) and the thick black line inside

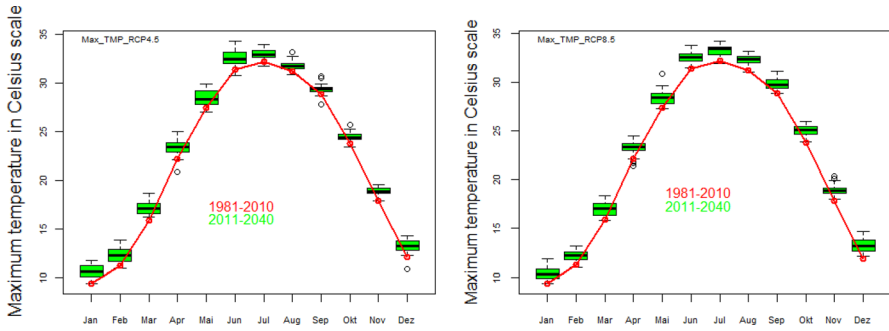


Fig. 8 Comparison of baseline (1981–2010) and future’s (2011–2040) projected maximum temperature using Bayesian Model Averaging. The left side panel is for RCP4.5 and the right side panel is for RCP8.5

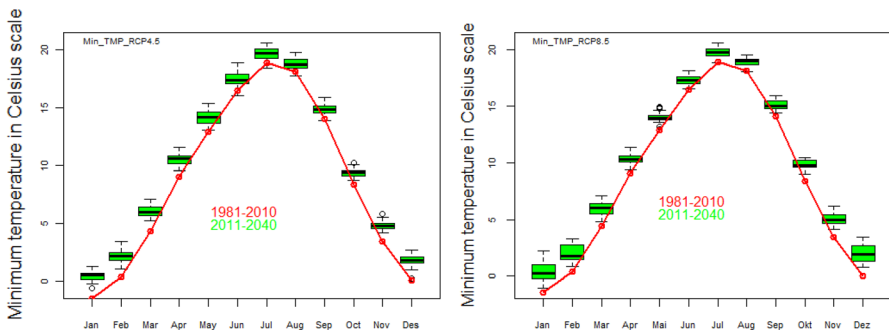


Fig. 9 Comparison of baseline (1980–2010) and future’s projected minimum temperature using Bayesian Model Averaging. The left side panel is for RCP4.5 and the right side panel is for RCP8.5

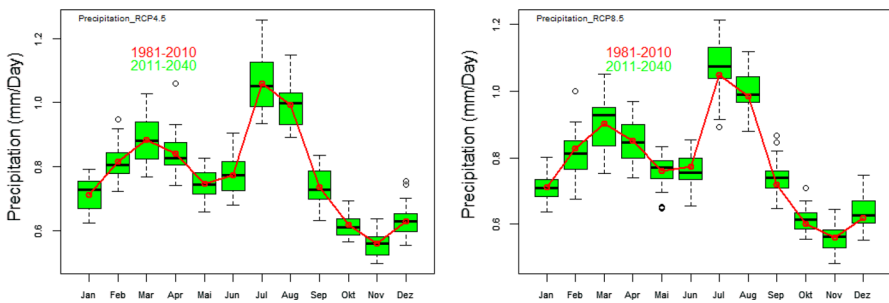


Fig. 10 Comparison of baseline (1980–2010) and future’s projected precipitation using Bayesian Model Averaging. The left side panel is for RCP4.5 and the right side panel is for RCP8.5

the whisker plot shows the average climatology for each month for thirty years of the future period (2011–2040). By comparing these two data sets we can assess the changes in future climatology with respect to the baseline time period. Additionally, the Figs. 8, 9, and 10 also offer a comparison between the two emission scenarios for the UIB Pakistan.

Figure 8 displays the results for maximum temperature under the RCP4.5 and RCP8.5 scenarios. It can be seen from both scenarios that, on average, the increase in temperature is larger in the second half of each year as compared to the first half. It is obvious from Fig. 8 that the maximum temperature has higher increasing trend in winter season as compared to that in summer season, under both scenarios. Figure 8 shows that the increase is higher under the RCP8.5 in comparison to the baseline period. The results also show that there is more variability in projected temperature in the first half of the year than in the second half, under both scenarios. On average, there is less variation in the projected temperature under the RCP8.5.

Figure 9 shows the projected minimum temperature from selected GCMs under the RCP4.5 and RCP8.5 scenarios. The projected minimum temperature indicates that there is more variation in the first half of the year than in the second one (monthly average climatology for 2011–2040). In Figs. 8, 9, and 10, one thing is common under both scenarios: the increase occurs more often during the winter season as compared to the summer season. On average, the results of RCP4.5 are looking more stable during the whole year as compared to that of RCP8.5 with respect to variability. However, in some months there is more variability under the RCP4.5 scenario, e.g. during the Summer (May, June, July, August), whereas during winter season (December, January, February, March) it is larger under the RCP8.5 scenario.

Figure 10 represents a comparison of precipitation of baseline and projected future periods under the RCP4.5 and RCP8.5 scenarios. The general trend in precipitation is looking similar, under both scenarios, but, if we look carefully at the results, we can see some differences, particularly in different months. The differences are obvious in the months of January, March, April, June, July, September, October and December. Under RCP4.5, precipitation increases in the months of January and August while it decreases in February, April, May, July, September, October and remains similar in the remaining months. Under RCP8.5, precipitation increases in the months of February and June while no changes in January, April and decreases in the remaining months occur as compared to the baseline precipitation. The results under RCP4.5 show more variations in most of the months. Overall, there is a clear increasing/decreasing trend in most of the months under the RCP8.5 scenario as compared to the baseline duration. On the other hand, under the RCP4.5 scenario, no such obvious increasing/decreasing trend is noted in future's projected precipitation.

5 Conclusion and recommendations

It was noted that some climate models are best and consistent in their performance for maximum, minimum temperature and precipitation. For precipitation, GCMs behave somewhat differently compared to maximum and minimum temperature, particularly when using different prior choices. The results give significant directions about using GCMs for different variables. It was also noted that the trend of observed data is captured better by BMA predictions than by individual GCMs outputs, which is supported by higher values of correlation coefficients and Nash–Sutcliffe efficiency coefficients. Important results have been noted during the

comparison of correlation coefficients and Nash–Sutcliffe efficiency coefficients for different variables; both statistics behave differently for temperature and precipitation. One possible reason may be that the parameterization scheme of some models is better suited for precipitation, and the other ones are better when it comes to predicting temperature. However, we need more studies to confirm this. Future projections of maximum temperature, minimum temperature and precipitation under RCP4.5 and RCP8.5 show mixed results: in some months, these variables have an increasing trend and a decreasing trend in the other months. It was observed that there is more increase in maximum and minimum temperature under the RCP8.5 scenario as compared to RCP4.5. Further, it was noted that there are more changes in results under RCP8.5 but the projections under RCP4.5 show more variability as compared to the RCP8.5. The application of this methodology is location specific, but the readers can easily apply it to the area of their interest for model selection as well as for ensemble projections. Furthermore, the interested researchers can use this approach to select RCM(s) and scenario(s). We are looking forward to conducting a study covering all climate zones of Pakistan to evaluate the performance of BMA and explore the preferences of GCMs for each climate zone.

References

- Achieng KO, Zhu J (2019) Application of Bayesian framework for evaluation of streamflow simulations using multiple climate models. *J Hydrol* 574:1110–1128. <https://doi.org/10.1016/J.JHYDROL.2019.05.018>
- Ajami NK, Duan Q, Gao X, Sorooshian S (2006) Multimodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results. *J Hydrometeorol* 7:755–768
- Ali S, Li D, Fu C, Khan F (2015) Twenty first century climate and hydrological changes over the Himalayan region of Pakistan. *Environ Res Lett*. <https://doi.org/10.1088/1748-9326/10/1/014007>
- Ancombe FJ (1948) The transformation of Poisson, Binomial and Negative-Binomial data. *Biometrika* 35(3–4):246–254. <https://doi.org/10.2307/2332343>
- Ashfaq A, Rastogi D, Mei R, Touma D, Leung LR (2017) Sources of errors in the simulation of South Asian Summer Monsoon in the CMIP5 GCMs. *Clim Dyn* 49(1–2):193–223. <https://doi.org/10.1007/s00382-016-3337-7>
- Baumberger C, Knutti R, Hirsch HG (2017) Building confidence in climate model projections: an analysis of inferences from fit. *Wiley Interdiscip Rev Clim Change*. <https://doi.org/10.1002/wcc.454>
- Beven K, Binley A (1992) The future of distributed models—model calibration and uncertainty prediction. *Hydrol Process* 6(3):279–298. <https://doi.org/10.1002/hyp.3360060305>
- Blasone RS, Verugt JA, Madsen H, Rosbjerg D, Robinson BA, Zyvoloski GA (2001) Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Adv Water Resour* 31(4):630–648. <https://doi.org/10.1016/j.advwatres.2007.12.003>
- Brekke L, Thrasher BL, Maurer EP, Pruitt T (2013) Downscaled CMIP3 and CMIP5 Climate Projections. http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/. Accessed 26 June 2019
- Brient F (2020) Reducing uncertainties in climate projections with emergent constraints: concepts, examples and prospects. *Adv Atmos Sci* 37:1–15. <https://doi.org/10.1007/s00376-019-9140-8>
- Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of GCM precipitation by quantile mapping: How well do the methods preserve changes in quantiles and extremes? *J Clim* 28(17):2385–2404. <https://doi.org/10.1175/JCLI-D-14-00754.1>

- Clyde M, Littman M, Wang Q, Ghosh J, Li Y, Bergh DVD (2020) BAS: Bayesian variable selection and model averaging using Bayesian adaptive sampling. <https://cran.r-project.org/web/packages/BAS/index.html>. Accessed 27 September 2020
- Draper D (1995) Assessment and propagation of model uncertainty. *J R Stat Soc B* 57(1):45–97
- Duan Q, Ajami NK, Gao X, Sorooshian S (2007) Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv Water Resour* 30(5):1371–1386
- Eckstein D, Künzel V, Schäfer L, Wings M (2019) Global Climate Risk Index 2020. Berlin. <http://www.germanwatch.org>. Accessed 20 August 2020
- Feldkircher M, Zeugner S (2020) BMS: Bayesian Model Averaging Library. <https://cran.r-project.org/web/packages/BMS/index.html>. Accessed on 27 September 2020
- Fernandez C, Ley E, Steel MFJ (2001) Model uncertainty in cross-country growth regression. *J Appl Econ* 16:563–576. <https://doi.org/10.1002/jae.263>
- Fraley C, Raftery AE, Slougher JM, Gneiting T (2018) ensembleBMA - Probabilistic Forecasting using Ensembles and Bayesian Model Averaging. <https://rdr.io/cran/ensembleBMA/man/>. Accessed on 27 September 2020
- Freer J, Beven K, Ambrose B (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resour Res* 32:2161–2173. <https://doi.org/10.1029/95WR03723>
- Global Climate Risk Index (2019) German Watch, 2018. https://germanwatch.org/sites/germanwatch.org/files/Global%20Climate%20Risk%20Index%202019_2.pdf. Accessed 11 June 2019
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. *Stat Sci* 14(4):382–417
- Huang Y (2014) Comparison of general circulation model outputs and ensemble assessment of climate change using Bayesian approach. *Glob Planet Change* 122:362–370
- Khan F, Pilz J, Ali S (2017) Improved hydrological projections and reservoir management under changing climate in the upper Indus basin. *Pakistan Water and environment journal* 31(2):235–244. <https://doi.org/10.1111/wej.12237>
- Khan F, Pilz J, Amjad M, Wiberg D (2015) Climate variability and its impacts on water resources in the Upper Indus Basin under different climate change scenarios. *Int J Global Warming* 8(1):46–69. <https://doi.org/10.1504/IJGW.2015.071583>
- Kharin VV, Zwiers FW, Gognon N (2001) Skill of seasonal hindcasts as a function of the ensemble size. *Clim Dynamics* 17:835–843. <https://doi.org/10.1007/s003820100149>
- Kumar A, Mitra AK, Bohra AK, Iyengar GR, Durai VR (2012) Multi-model ensemble (MME) prediction of rainfall using neural network s during monsoon season in India. *Meteorol Appl* 19:161–169. <https://doi.org/10.1002/met.254>
- Leedale J, Tompkins AM, Caminade C, Jones AE, Nikulin G, Morse AP (2016) Projecting malaria hazard from climate change in Eastern Africa using large ensembles to estimate uncertainty. *Geospat Health* 11(1):102–114. <https://doi.org/10.4081/gh.2016.393>
- Li G, Shi J (2010) Application of Bayesian model averaging in modeling long-term wind speed distributions. *Renew Energy* 35(6):1192–1202. <https://doi.org/10.1016/j.renene.2009.09.003>
- Lorenz EN (1982) Atmospheric predictability experiments with large numerical model. *Tellus* 34:505–513
- Lutz AF, Herbert WM, Hester B, Arun BS, Philippus W, Immerzeel WW (2016) Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *Int J Climatol* 36:3988–4005. <https://doi.org/10.1002/joc.4608>
- Maraun D (2013) Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J Clim* 26(2):2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1>
- McSweeney CF, Jones RG, Booth BBB (2012) Selecting ensemble members to provide regional climate change information. *J Clim* 25:7100–7120. <https://doi.org/10.1175/JCLI-D-11-00526.1>
- McSweeney CF, Jones RG, Lee RW, Rowel DP (2015) Selecting CMIP GCMs for downscaling over multiple regions. *Clim Dyn* 44:3237–3260. <https://doi.org/10.1007/s00382-014-2418-8>
- Mendlik T, Gobiet A (2016) Selecting climate simulations for impact studies based on multivariate patterns of climate change. *Clim changes* 135:381–393. <https://doi.org/10.1007/s10584-015-1582-0>
- Min SK, Simonis D, Hence A (2007) Probabilistic climate change predictions applying Bayesian model averaging. *Philos Trans R Soc A* 365:2103–2116. <https://doi.org/10.1098/rsta.2007.2070>
- Negggers RAJ (2015) Attributing the behavior of low-level clouds in large-scale models to subgrid-scale parameterizations. *J Adv Model Earth Syst* 7:2029–2043. <https://doi.org/10.1002/2015MS000503>

- Patricola CM, Cook KH (2012) Mid-twenty-first century warm season climate change in the Central United States Part I: regional and global model predictions. *Clim Dynamics* 40:551–568. <https://doi.org/10.1007/s00382-012-1605-8>
- Qu B, Zhang X, Pappenberger F, Zhang T, Fang Y (2017) Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian model averaging. *Water* 9(2):74. <https://doi.org/10.3390/w9020074>
- Raftery AE (1993) Bayesian model selection in structural equation models. In: Bolen KA, Long JS (eds) *Testing structural equation models*. Sage, Boca Raton, pp 163–180
- Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev* 133:1155–1174. <https://doi.org/10.1175/MWR2906.1>
- Raftery AE, Zheng Y (2003) Long-run performance of Bayesian model averaging. *J Am Stat Assoc* 98(464):931–938. <https://doi.org/10.1198/016214503000000891>
- Raftery A, Hoeting J, Volinsky C, Painter I, Yeung KY (2020) Bayesian Model Averaging (BMA). <https://cran.r-project.org/web/packages/BMA/index.html>. Accessed on 27 September 2020
- Reum JCP, Blanchard JL, Holsman KK, Aydin K, Hollowed AB, Hermann AJ, Cheng W, Faig A, Haynie AC, Punt AE (2020) Ensemble projections of future climate change impacts on the eastern bering sea food web using a multispecies size spectrum model. *Front Mar Sci* 7:124. <https://doi.org/10.3389/fmars.2020.00124>
- Rivington M, Matthews KB, Bellocchi G, Buchan K (2006) Evaluating uncertainty introduced to process-based simulation model estimates by alternative sources of meteorological data. *Agric Syst* 88:451–471. <https://doi.org/10.1016/j.agsy.2005.07.004>
- Rojas R, Feyen L, Dassargues A (2008) Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resour Res*. <https://doi.org/10.1029/2008WR006908>
- Räty O, Räisänen J, Ylhäisi JS (2014) Evaluation of delta change and bias correction methods for future daily precipitation: intermodel cross-validation using ENSEMBLES simulations. *Clim Dyn* 42:2287–2303. <https://doi.org/10.1007/s00382-014-2130-8>
- Salathe EP, Mote PW, Wiley MW (2007) Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States pacific northwest. *Int J Climatol* 27:1611–1621. <https://doi.org/10.1002/joc.1540>
- Sloughter MJ, Raftery AE, Gneiting T, Fraley C (2007) Probabilistic quantitative precipitation forecasting using Bayesian Model Averaging. *Am Meteorol Soc* 135:3209–3220. <https://doi.org/10.1175/MWR3441.1>
- Smith J, Hulme M (1998) Climate change scenarios. In: Feenstra J, Burton I, Smith J, Tol R (eds) *UNEP Handbook on methods for climate change impact assessment and adaptation studies*. United Nations Environmental Programme, Nairobi, Kenya and Institute for Environmental Studies, Amsterdam
- Stedinger JR, Vogel RM, Lee SU, Batchelder R (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour Res*. <https://doi.org/10.1029/2008WR006822>
- Intergovernmental Panel on Climate Change (IPCC). Climate change (2013) The physical science basis. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research Atmospheres* 106(D7):7183–7192
- Themeßl M, Gobiet A, Leuprecht A (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int J Climatol* 31(10):1530–1544. <https://doi.org/10.1002/joc.2168>
- Users customized prior: <http://bms.zeugner.eu/custompriors.php>. Accessed on 30 May 2019
- Uusitalo L, Lehtikoinen A, Helle I, Myrberg K (2015) An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environ Model Softw* 63:24–31. <https://doi.org/10.1016/j.envsoft.2014.09.017>
- Vruntg JA, Robinson BA (2007) Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resour Res* 43:01411. <https://doi.org/10.1029/2005WR004838>
- Wood AW, Leung LR, Sridhar V, Lettenmaier DP (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim Change* 62:189–216

- Yang H, Wang B, Wang B (2011) Reducing biases in regional climate downscaling by applying Bayesian model averaging on large-scale forcing. *Clim Dynamics* 39(9–10):2523–2532. <https://doi.org/10.1007/s00382-011-1260-5>
- Zeugner S (2011) Bayesian model averaging with BMS. Technical report. <http://bms.zeugner.eu/tutorials/bms.pdf>
- Zhang X, Srinivasan R, Bosch D (1999) Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. *J Hydrol* 374(2–3):307–317
- Zheng Y, Keller AA (2007) Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach. *Water Resour Res* 43:08407. <https://doi.org/10.1029/2006WR005345>
- Zhang X, Srinivasan R, Bosch D (2009) Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. *J Hydrol* 374(3–4):307–317. <https://doi.org/10.1016/j.jhydrol.2009.06.023>

Firdos Khan holds a PhD degree from University of Klagenfurt, Austria and is currently working as assistant professor at National University of Sciences and Technology (NUST) in Islamabad, Pakistan. He obtained his MPhil in Statistics from Quaid-i-Azam University Islamabad in 2008. From 2008 to 2013, he worked as Scientist at Global Change Impact Studies Centre, where he was involved in research in dynamical modelling and statistical downscaling in the area of climate change and its impacts. As a young scientist from Pakistan, he was selected by the International Institute for Applied Systems Analysis (IIASA) to participate in the Young Scientists Summer Program (YSSP) at Laxenburg, Austria, in 2012. Dr. Firdos is an active member of the European Geosciences Union (EGU) and the American Geophysical Union (AGU) since 2014 and 2017, respectively.

Authors and Affiliations

Firdos Khan¹  · Jürgen Pilz² · Shaukat Ali³

✉ Firdos Khan
fkyousafzai@gmail.com

¹ School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

² Institut Für Statistik, Alpen-Adria Universität, Universitätsstraße 65-67, 9020 Klagenfurt, Austria

³ Global Change Impact Studies Centre (GCISC), Ministry of Climate Change, Islamabad 44000, Pakistan