RMetS

Royal Meteorological Society

Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach

Arthur F. Lutz,[a,b]* Herbert W. ter Maat,[c] Hester Biemans,[c] Arun B. Shrestha,[d] Philippus Wester[d] and Walter W. Immerzeel[a,b]

[a] *FutureWater, Wageningen, The Netherlands*
[b] *Department of Physical Geography, Utrecht University, The Netherlands*
[c] *Alterra – Wageningen UR, The Netherlands*
[d] *International Centre for Integrated Mountain Development, Kathmandu, Nepal*

ABSTRACT: Climate change impact studies depend on projections of future climate provided by climate models. The number of climate models is large and increasing, yet limitations in computational capacity make it necessary to compromise the number of climate models that can be included in a climate change impact study. The selection of climate models is not straightforward and can be done by following different methods. Usually, the selection is either based on the entire range of changes in climatic variables as projected by the total ensemble of available climate models or on the skill of climate models to simulate past climate. The present study combines these approaches in a three-step sequential climate model selection procedure: (1) initial selection of climate models based on the range of projected changes in climatic means, (2) refined selection based on the range of projected changes in climatic extremes and (3) final selection based on the climate model skill to simulate past climate. This procedure is illustrated for a study area covering the Indus, Ganges and Brahmaputra river basins. Subsequently, the changes in climate between 1971–2000 and 2071–2100 are analysed, showing that the future climate projections in this area are highly uncertain but that changes are imminent.

KEY WORDS   GCM; climate model selection; climate change impact studies; climate model ensemble; Indus; Ganges; Brahmaputra

Received 17 June 2015; Revised 9 October 2015; Accepted 16 November 2015

1. Introduction

Climate change impact studies depend on projections of future climate provided by climate models. Due to their coarse spatial resolution, outputs from general circulation models (GCMs) are usually directly downscaled to higher resolution using empirical–statistical downscaling methods or are used as boundary conditions for regional climate models (RCMs), with their outputs being subsequently downscaled to a higher resolution. The downscaled outputs are then used to assess future climatic changes and to drive other sector-specific models for climate change impact studies. Outcomes from these studies are used by policymakers to support decisions on climate change adaptation measures.

The number of GCMs available for climate change projections is increasing rapidly. For example, the Coupled Model Intercomparison Project Phase 3 (CMIP3) archive (Meehl *et al.*, 2007), which was used for the

Intergovernmental Panel on Climate Change's (IPCC) fourth Assessment Report (IPCC, 2007), contains outputs from 25 different GCMs, whereas the CMIP5 archive (Taylor *et al.*, 2012), which was used for the fifth IPCC Assessment Report (IPCC, 2013), contains outputs from 61 different GCMs. These GCMs often have multiple ensemble members, resulting in an even larger number of available model runs.

Despite improvements in the CMIP5 models compared to CMIP3 in terms of process representation (e.g. Blázquez and Nuñez, 2013; Sperber *et al.*, 2013), uncertainty about the future climate remains large (e.g. Knutti and Sedláček, 2012) and even increases locally with the larger number of models available (e.g. Joetzjer *et al.*, 2013; Lutz *et al.*, 2013). Considering the large number of available climate models and constraints in the available computational and human resources, detailed climate change impact studies cannot include all projections. In practice, one climate model or a small ensemble of climate models is selected for the assessment. Despite the importance of using an ensemble that is representative of the region of interest and shows the full uncertainty range, the selection of models to be included in the ensemble is not straightforward and can be based on multiple criteria.

Climate models are often selected based on their skill to simulate the present and near-past climate (e.g. Pierce *et al.*, 2009; Biemans *et al.*, 2013). This approach is referred to as the past-performance approach. Another approach is the so-called envelope approach, where an ensemble of models covering a wide range of projections for one or more climatological variables of interest is selected from the pool of available models. This approach aims at covering all possible futures as projected by the entire pool of climate models. Some approaches consider only the changes in mean air temperature and total annual precipitation (e.g. Immerzeel *et al.*, 2013; Sorg *et al.*, 2014; Warszawski *et al.*, 2014), whereas other approaches consider more climatological variables using cluster analysis algorithms (e.g. Houle *et al.*, 2012; Cannon, 2014). Another approach uses criteria for model independence to generate a representative selection of models from a larger ensemble, where the ensemble of selected models has characteristics that reflect the larger ensemble (Evans *et al.*, 2013).

The decision on which variables are considered depends on the character and goals of the climate change impact assessment. For example, a study on future hydrological floods will be most interested in changes of extremely high precipitation events, whereas a study on the impacts of climate change for the exploitation of ski slopes in a mountainous area will most likely consider changes in winter temperature and winter precipitation. The major drawback of envelope-based approaches is that the models' skill to simulate climate are not considered as all available climate model runs are considered to have equal plausibility, and only changes in the annual means are criteria for selection. On the other hand, selecting models with only a high skill in simulating present and past climate may lead to the omission of possible futures. These two contrasting methods to select a climate model ensemble will result in different ensembles, with different mean projections and different uncertainties in the climate change projection.

The uncertainty originating from the spread in climate models' projections is considered to be a large source of uncertainty in climate change impact studies; for example, this uncertainty is often larger than model parameter uncertainties, uncertainty stemming from natural variability and structural uncertainties in hydrological models (Minville *et al.*, 2008; Finger *et al.*, 2012). Therefore, the selection of climate models is a crucial step when conducting a climate change impact study.

Here we present an approach to select climate models combining the envelope approach and the past-performance approach. The goal is to select an ensemble consisting of a manageable number of climate model runs, which still represents all possible futures in terms of future mean air temperature and annual precipitation sums as well as future changes in climatic extremes. The climate in the study area is complex with mountainous climate types and monsoon dynamics playing important roles. Because these are often poorly simulated by climate models (Sperber *et al.*, 2013), in addition, we aim to select only models that have sufficient skill in simulating the present day climate in the study area.

2. Study area and data

The approach is illustrated for a study area covering the Indus, Ganges and Brahmaputra river basins, ranging from their sources in the Hindu-Kush Himalayan mountains and Tibetan Plateau to their mouths in the Arabian Sea and Bay of Bengal, respectively (Figure 1). This region is considered a climate change 'hotspot' (De Souza *et al.*, 2015; Nepal and Shrestha, 2015). The ensemble of climate models selected using the approach described in the present study will be downscaled to assess future climate changes at different scales and force hydrological and crop growth models in later stages of the research.

The RCP4.5 and RCP8.5 model runs available in March 2013 are used in the CMIP5 repository (Taylor *et al.*, 2012) as the initial pool of climate models from which the present study's ensemble of models can be selected. For RCP4.5, the total number of model runs available is 94, whereas 69 model runs are available for RCP8.5. For analysis of the projected changes in climatic extremes, the database presented in Sillmann *et al.* (2013a) and Sillmann *et al.* (2013b) is used. This database includes projected changes in the climatic indices as defined by the CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices [ETCCDI, (Peterson, 2005)] for CMIP5 models. To evaluate the model performance of individual GCM runs, the WATCH Forcing Data ERA-Interim (WFDEI) dataset (Weedon *et al.*, 2014) is used, which has been generated using the same methodology as the widely used WATCH Forcing Data (Weedon *et al.*, 2011) by making use of the ERA-interim reanalysis data (Dee *et al.*, 2011). We use the precipitation data in WFDEI, which is bias-corrected to the Global Precipitation Climatology Centre (GPCC) precipitation climatology (Schneider *et al.*, 2013).

3. Methods

For the selection of GCM runs, the envelope-based approach and the past-performance approach are integrated in a three-step selection procedure (Figure 2). The initial selection of climate models is based on the projected average annual changes in the mean temperature and precipitation sum. Subsequently, this selection is refined based on changes in extremes of precipitation and temperature. Ultimately, the final selection is based on validation of the remaining models' past performance to the WFDEI climatic reference product (Weedon *et al.*, 2014).

3.1. Selection of representative concentration pathways

In the climate modelling community, four representative concentration pathways (RCPs) are generally used as a basis for long-term and near-term climate modelling
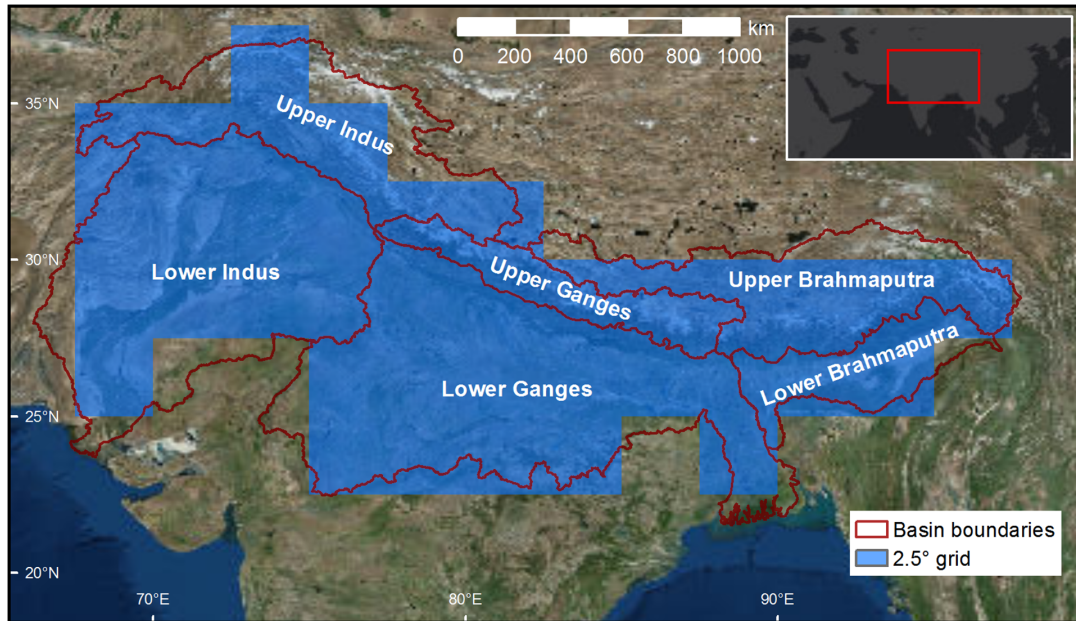
Figure 1. Study area comprising the Indus, Ganges and Brahmaputra river basins. The 2.5° × 2.5° grid used for most steps in the selection procedure is shown in blue.

experiments. The four selected RCPs are considered to be representative of the scientific literature and include one mitigation scenario (RCP2.6), two medium stabilization scenarios (RCP4.5/RCP6) and one very high baseline emission scenario (RCP8.5) (van Vuuren *et al.*, 2011a). RCP2.6 is representative of the low end of the scenario literature in terms of emissions and radiative forcing (van Vuuren *et al.*, 2011b). Often, these scenarios show negative emissions from energy use in the second half of the 21st century. The scenario is shown to be technically feasible, but one of the key assumptions is the full participation of all countries in the world in the short run, including broadening participation beyond Organisation for Economic Co-operation and Development (OECD) countries and commitment of important OECD countries (van Vuuren *et al.*, 2010). As robust, realistic climate change scenarios need to be developed to facilitate the planning of adaptation measures, we choose not to include RCP2.6 for the climate model ensemble. This leaves the choice between two medium stabilization scenarios (RCP4.5 and RCP6) and one very high baseline emission scenario (RCP8.5). The best choice in that case is to include RCP4.5 and RCP8.5, thus including one medium stabilization scenario and the high emission scenario, and covering the entire range of radiative forcing resulting from RCP4.5, RCP6 and RCP8.5. Although we decided to include only RCP4.5 and RCP8.5, the approach presented can evidently be applied to model ensembles for the other RCPs as well.

3.2. Initial selection (Step 1): changes in climatic means

The initial selection is based on the range of projections of changes in mean air temperature (ΔT) and annual precipitation sum (ΔP) between 1971–2000 and 2071–2100, averaged over the 2.5° × 2.5° grid cells included in the model domain (Figure 1). This calculation was done using

the Royal Netherlands Meteorological Institute (KNMI) Climate Explorer (http://climexp.knmi.nl). For the model runs included in RCP4.5 and RCP8.5 separately, the 10th and 90th percentile values for ΔT and ΔP are determined after resampling all GCM data to the same 2.5° × 2.5° grid. These values represent the four corners of the spectrum of projections for temperature and precipitation change. The tenth percentile value for ΔT and tenth percentile value for ΔP are in the 'cold, dry' corner of the spectrum. The 10th percentile value for ΔT and 90th percentile value for ΔP are in the 'cold, wet' corner of the spectrum. The 90th percentile value for ΔT and 10th percentile value for ΔP are in the 'warm, dry' corner of the spectrum. The 90th percentile value for ΔT and 90th percentile value for ΔP are in the 'warm, wet' corner of the spectrum. The 10th and 90th percentile values are chosen rather than the minimum and maximum projections to avoid selecting outliers, cf. other studies (e.g. Immerzeel *et al.*, 2013; Sorg *et al.*, 2014). The proximity of the model runs to the 10th and 90th percentile values is derived from the model runs' percentile rank scores corresponding to their projections for ΔT and ΔP with respect to the entire range of projections in the entire ensemble:

$$D_{p_i^T, p_i^P} = \sqrt{\left(\left(\left| P_i^T - P_j^T \right| \right)^2 + \left(\left| P_i^P - P_j^P \right| \right)^2 \right)} \quad (1)$$

where $D_{p_i^T, p_i^P}$ is the distance of a model (*j*)'s ΔT and ΔP (P_j^T and P_j^P, respectively) to the corner (*i*)'s 10th and/or 90th percentile score of ΔT and ΔP for the entire ensemble (P_j^T and P_j^P, respectively). For each corner, the five models with the lowest values for D and data available at a daily time step are selected from the ensemble. Only models that have data available at a daily time step are selected because this is a requirement for an empirical–statistical downscaling
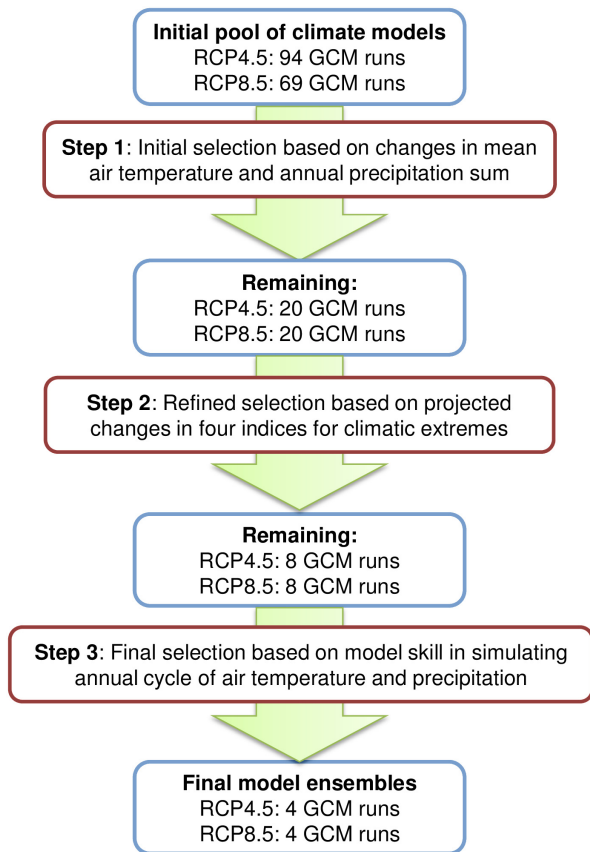
Figure 2. Flow diagram showing the steps taken during the climate model selection procedure.

Table 1. Description of ETCCDI indices leading the refined selection step.

Meteorological variable	ETCCDI index	Index description
Precipitation	R99pTOT	Precipitation due to extremely wet days (>99th percentile)
Precipitation	CDD	Consecutive dry days: maximum length of dry spell ($P < 1$ mm)
Air temperature	WSDI	Warm spell duration index: count of days in a span of at least 6 days where $TX > 90$th percentile (TX_{ij} is the daily T_{max} on day i in period j)
Air temperature	CSDI	Cold spell duration index: count of days in a span of at least 6 days where $TN < 10$th percentile (TN_{ij} is the daily T_{min} on day i in period j)

method to be applied to the GCM runs at a later stage. Nonetheless, model runs with data only available at larger time steps are included in the initial pool of available model runs used to calculate the model runs' percentile scores to have a complete representation of all projected possible futures. The initial selection results in five model runs × four corners = 20 model runs for each RCP.

For GCMs with ensemble members of different initial conditions available (denoted with rxixpx behind the GCM's name), all members are included in the ensemble that is subjected to the initial selection step. The inclusion of all initial condition ensemble members will lead to a different definition of the 10th and 90th percentile values than when only one initial condition ensemble member per GCM is included. We chose to include all initial condition ensemble members because each of them leads to a different future, and there is no way to determine which of the initial condition ensemble members should be preferred over others.

3.3. Refined selection (Step 2): changes in climatic extremes

The number of model runs remaining after the initial selection process is further reduced during the refined selection step. In this step, the model runs are evaluated for their projected changes in climatic extremes. The changes in climatic extremes for air temperature and precipitation are evaluated by considering the changes in

two ETCCDI indices (Peterson, 2005) (Table 1) for both air temperature and precipitation. For characterization of changes in air temperature extremes, changes in the warm spell duration index (WSDI) and the cold spell duration index (CSDI) are analysed. For characterization of changes in precipitation extremes, the precipitation due to extremely wet days (R99pTOT) and the number of consecutive dry days (CDD) are considered. As the climate model ensemble will be used to force hydrological and crop growth models, we have chosen to analyse changes in R99pTOT and CDD as obvious indicators of precipitation extremes leading to associated hydrological extremes. While CDD is an important indicator for dry spells affecting crop growth, WSDI is also an indicator for situations where crops may face water stress due to increased evapotranspiration during warm spells. Changes in WSDI and CSDI both have effects on the cryospheric processes (snow and ice melt/accumulation), which are important in the upstream parts of the study area. The changes in these indices between 2071–2100 and 1971–2000 are calculated from the database constructed by Sillmann *et al.* (2013a, 2013b). Not all GCM runs used for the initial selection are included in this database. For those runs, the ETCCDI indices were calculated using the same procedures as Sillmann *et al.* (2013a, 2013b) used in their study. The indices are calculated from the daily model output for each individual year in the future period (2071–2100) and reference period (1971–2000), for the individual 2.5° × 2.5° grid cells covering the study area (Figure 1). For both periods, the indices are then averaged over a period of 30 years. The changes in the indices are then calculated as a percentual change for the future period with respect to the reference period. Subsequently, these

changes in the indices are averaged over the 2.5° × 2.5° grid cells covering the study area.

For each model selected during the initial selection, the most relevant index for air temperature and the most relevant index for precipitation are considered. For example, for the models in the warm, wet corner, WSDI indicating warm spells and R99pTOT indicating extreme precipitation events are considered. CDD and CSDI are not considered in that case, but they are considered for models in the dry and cold corners, respectively. For the five models initially selected for each corner, the two relevant indices are both ranked and given scores 1–5. For example, in the warm, wet corner, the model with the largest increase in R99pTOT scores five points for that index, whereas the model with the smallest increase in R99pTOT scores one point for that index. Similarly, the model with the largest increase in WSDI scores five points for that index, and the model with the smallest increase in WSDI scores one point for that index. Both scores are then averaged to obtain a final score. Based on that final score, the two models with the highest scores are selected. Thus, for each corner, the number of models is reduced from five to two models. For each RCP, four corners × two models = eight models are selected, which are validated to the climatic reference product in the next step.

3.4. Final selection (Step 3): past performance

The models remaining after the refined selection are subjected to a validation to the WFDEI dataset (Weedon *et al.*, 2014). The selected models are compared to the WFDEI for six subdomains (upstream Indus, upstream Ganges, upstream Brahmaputra, downstream Indus, downstream Ganges, downstream Brahmaputra; Figure 1). The skill assessment is done for the period 1980–2004, and skill scores are calculated for each model, taking into account all monthly values in this dataset spanning 25 years. Criteria to assess each model's ability to simulate the reference climate are comparisons between the model simulation and WFDEI for monthly average mean air temperature and monthly precipitation sums. The validation is done per subdomain (Figure 1) to analyse differences in model performance between the different river basins and between the mountainous upstream parts and downstream parts of the basins and to avoid compensations of overestimations and underestimations in the entire domain.

To assess the performance of the selected GCM runs, skill scores are derived based on earlier work by Perkins *et al.* (2007), Sanchez *et al.* (2009) and Kjellström *et al.* (2010). The calculation of temperature and precipitation skills are different. For the calculation of the skill score of temperature, the approach by Perkins *et al.* (2007) is used. A metric was developed which 'calculates the cumulative minimum value of two distributions of each binned value, thereby measuring the common area between two PDFs':

$$S_{\text{score}} = \sum_{1}^{n} \text{minimum}\left(Z_{\text{GCM}}, Z_{\text{OBS}}\right) \qquad (2)$$

where n is the number of bins used to calculate the probability density function (PDF) for a subdomain, Z_{GCM} is the frequency of values in a given bin from the model and Z_{OBS} is the frequency of values in a given bin from the observed data. The number of bins used is 100. If a model simulates the observed frequencies perfectly, the skill score (S_{score}) will equal one, which is the total sum of the probability at each bin centre in a given PDF.

The skill score of precipitation is based on the work of Sanchez *et al.* (2009), which consists of a collection of five skill score functions, taking into account different aspects of the behaviour of precipitation. These skill score functions are listed here for our case of comparing GCM data to the WFDEI dataset:

$$f_1 = 1 - \left(\frac{\left| A_{GCM} - A_{WFDEI} \right|}{2 \cdot A_{WFDEI}} \right)^{0.5} \qquad (3)$$

$$f_2 = 1 - \left(\frac{\left| A_{GCM}^+ - A_{WFDEI}^+ \right|}{2 \cdot A_{WFDEI}^+} \right)^{0.5} \qquad (4)$$

$$f_3 = 1 - \left(\frac{\left| A_{GCM}^- - A_{WFDEI}^- \right|}{2 \cdot A_{WFDEI}^-} \right)^{0.5} \qquad (5)$$

$$f_4 = 1 - \left(\frac{\left| \overline{P_{GCM}} - \overline{P_{WFDEI}} \right|}{2 \cdot \overline{P_{WFDEI}}} \right)^{0.5} \qquad (6)$$

$$f_5 = 1 - \left(\frac{\left| \sigma_{GCM} - \sigma_{WFDEI} \right|}{2 \cdot \sigma_{WFDEI}} \right)^{0.5} \qquad (7)$$

where A_{GCM} and A_{WFDEI} are the areas below the cumulative distribution functions of the GCMs and WFDEI, respectively. Similarly, A^+ and A^- are the areas right and left of the 50th percentile. P is the mean annual precipitation over the total area, and σ is the standard deviation of the probability distribution function. The distribution as a whole is taken into account through the total area below the density function (f_1, Equation (3)) and the mean (f_4, Equation (6)). High and low precipitation amounts are taken into account by analysing the amounts above the 50th percentile limit (f_2, Equation (4)) and the amounts below the 50th percentile limit (f_3, Equation (5)), respectively. The shape of the distribution is considered through the variance (f_5, Equation (7)). The five skill scores are multiplied to yield a total skill score of precipitation. The skill scores of temperature and precipitation are calculated for the control period for the six subdomains separately. Following Perkins *et al.* (2007), the average is taken from the skill scores for both temperature and precipitation, and these scores are ranked per subdomain. Subsequently, the rankings of the subdomains are summed for each model run, which then results in a ranking for the entire study area, further referred to as general ranking.

The analysis until here is based on the general performance of the model. However, one of the main meteorological phenomena in the study area is the Asian monsoon. Most GCMs have difficulty in simulating the right amount of precipitation, and most of them
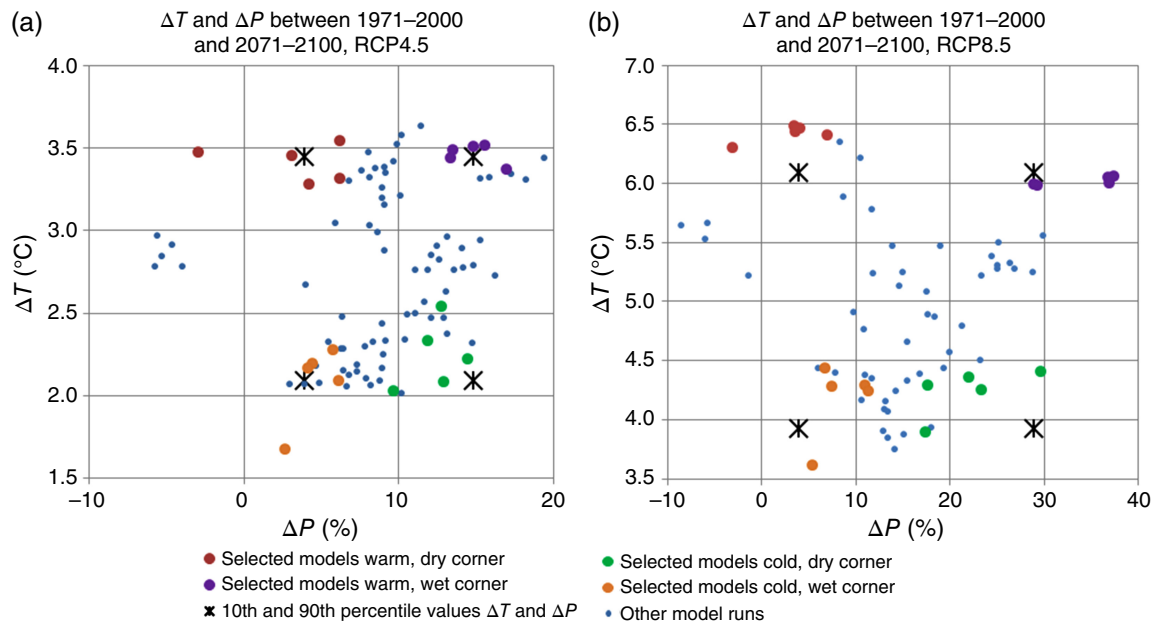
Figure 3. Projected changes in mean air temperature (ΔT) and annual precipitation sum (ΔP) between 2071–2100 and 1971–2000 for all included RCP4.5 (a) and RCP8.5 (b) GCM runs. Black crosses indicate the 10th and 90th percentile values for ΔT and ΔP. The model runs selected for Step 2 are indicated with colours.

underestimate the precipitation in the monsoon period (Sperber *et al.*, 2013). The correct representation of monsoonal precipitation is key for hydrological impact studies in this region, and an additional skill score specifically designed to quantify the capability of the GCM to simulate the monsoon is introduced. This skill score consists of the absolute bias in precipitation of the GCM for the complete monsoon period (June–September). The highest ranked GCM has the smallest absolute bias and the lowest ranked GCM the largest absolute bias. Finally, the two rankings (general ranking and monsoon ranking) are combined and weighted to reach a final ranking. The weight of the general ranking is three, and the weight of the monsoonal ranking is one. Based on this final ranking, the selection from step 2 (two model runs in each corner) is reduced to one model in each corner, forming the final ensemble of climate models.

4. Results

4.1. Selection of models

4.1.1. Initial selection (Step 1): changes in climatic means

The initial selection is made based on the projected changes in mean air temperature (ΔT) and annual precipitation sums (ΔP) between 2071–2100 and 1971–2000, thus indicating the projected change over 100 years (Figure 3, Supplementary Tables 1 and 2). For each GCM run, the distance to the 10th and 90th percentile values in the corners is calculated as described in Section 3.2, and the five models that have daily output available and are located closest to these corners are selected (Figure 3). The range in the projections for ΔT and ΔP is much larger for the RCP8.5 model pool than for the RCP4.5

model pool. ΔT ranges from 1.7 °C to 3.6 °C, and ΔP ranges from −5.7% to +19.4% for RCP4.5, whereas for RCP8.5, these ranges are 3.6–6.5 °C and −8.5% to +37.4%, respectively. This difference in the ranges of projections is also obvious from the differing degree of clustering in the scatter plots (Figure 3). Note that the values are averaged for all 2.5° × 2.5° grid cells in the study area; thus, spatial heterogeneity is not considered. The proximity of selected models to the 10th and 90th percentile values can differ substantially. For example, the selected models in the warm, wet corner for RCP4.5 are all relatively close to the 90th percentile values for ΔT and ΔP, whereas some of the selected models in the cold, wet corner have a considerable distance from the 10th and 90th percentile values for ΔT and ΔP. This is partly due to the relatively small number of models with daily output available in this corner of the model pool. In the cold, wet corner for RCP8.5, all models have considerable distance to the 10th and 90th percentile values of ΔT and ΔP.

4.1.2. Refined selection (Step 2): changes in climatic extremes

For the models remaining after the initial selection, the changes in four ETCCDI indices between 1971–2000 and 2071–2100 are calculated. For each corner, the two models that have the highest combined scores for the changes in the relevant temperature and precipitation indices are selected (models marked yellow in Figure 4). This process of calculating a combined score based on ranking can lead to the situation that models with the largest change in one of the ETCCDI indices are not selected. For example, in the warm, dry corner for RCP4.5, the IPSL-CM5A-MR_r1i1p1 model is not selected although it projects the largest changes in WSDI.

Table 2. Skill scores for GCM runs remaining after Step 2. Skill scores are calculated for precipitation and air temperature for six subdomains separately.

GCM runs	Precipitation skill scores						Mean air temperature skill scores					
	LI	UI	LB	UB	LG	UG	LI	UI	LB	UB	L LG	UG
CMCC-CM_r1i1p1	0.07	0.10	0.24	0.00	0.09	0.14	0.59	0.54	0.65	0.38	0.65	0.64
CMCC-CMS_r1i1p1	0.12	0.06	0.24	0.01	0.15	0.37	0.61	0.53	0.55	0.65	0.57	0.58
CSIRO-Mk3-6-0_r4i1p1	0.02	0.09	0.19	0.41	0.06	0.07	0.50	0.57	0.41	0.62	0.51	0.61
CSIRO-Mk3-6-0_r5i1p1	0.02	0.12	0.17	0.29	0.06	0.06	0.50	0.55	0.40	0.64	0.49	0.56
BNU-ESM_r1i1p1	0.32	0.15	0.10	0.04	0.23	0.29	0.60	0.53	0.40	0.57	0.60	0.43
CCSM4_r2i1p1	0.06	0.07	0.20	0.02	0.27	0.02	0.72	0.49	0.55	0.47	0.65	0.62
CESM1-BGC_r1i1p1	0.07	0.08	0.23	0.02	0.26	0.03	0.75	0.50	0.59	0.50	0.66	0.63
inmcm4_r1i1p1	0.14	0.36	0.15	0.37	0.24	0.21	0.52	0.45	0.27	0.41	0.51	0.59
IPSL-CM5A-LR_r4i1p1	0.06	0.10	0.05	0.29	0.07	0.41	0.61	0.58	0.69	0.43	0.61	0.66
CanESM2_r2i1p1	0.08	0.16	0.13	0.19	0.16	0.11	0.50	0.51	0.33	0.52	0.59	0.54
CanESM2_r3i1p1	0.14	0.16	0.13	0.20	0.11	0.12	0.55	0.46	0.32	0.52	0.63	0.58
bcc-csm1-1_r1i1p1	0.06	0.57	0.12	0.25	0.08	0.09	0.61	0.56	0.43	0.52	0.56	0.52
MRI-CGCM3_r1i1p1	0.04	0.26	0.05	0.10	0.03	0.16	0.48	0.45	0.59	0.41	0.50	0.57
EC-EARTH_r2i1p1	0.01	0.27	0.14	0.20	0.10	0.63	0.51	0.39	0.36	0.47	0.40	0.45

LI, Lower Indus; UI, Upper Indus; LB, Lower Brahmaputra; UB, Upper Brahmaputra; UG, Upper Ganges; LG, Lower Ganges. Note that GCM runs 'CMCC-CMS_r1i1p1' and 'inmcm4_r1i1p1' are selected in Step 2 for both RCP4.5 and RCP8.5 but listed only once in the table.

When looking at the combination of changes in mean air temperature, precipitation sums and changes in the ETC-CDI indices (Figure 5), it is clear that, in general, the models projecting large changes in the means also project large changes in extremes. For example, for RCP4.5, the models projecting the largest increases in mean air temperature also show the largest increases in warm spells and the largest decreases in cold spells (Figure 5, panel a). For RCP8.5, this correlation is less marked (Figure 5(c)). For precipitation in both RCPs, all models project increases in extremely high precipitation events, even the models that project decreasing total precipitation (Figure 5(b) and (c)). Increases in dry spells (CDD) are projected by models that project moderate increases in total precipitation (up to around +10% and +20% for RCP4.5 and RCP8.5, respectively). The generally observed relationship that models projecting a larger increase in total precipitation also project a larger increase in precipitation due to extremely wet days and the strongest decrease in dry periods holds for both RCPs.

The selection of five models in each corner during the initial selection step may lead to the omission of the GCM runs with the largest projected changes in extremes. This is because projected changes in extremes are only considered in the refined selection step; for example, the present study considers the changes in mean air temperature and annual precipitation sum as the leading criteria in its selection approach. The observation that the GCM runs located near the tails of the distribution of projected changes in the mean are also generally located near the tails of the distribution of projected changes in extremes (Figure 5) supports our sequence of selection steps.

4.1.3. Final selection (Step 3): past performance

The final selection of models is based on a validation of model performance to the WFDEI dataset. Table 2 lists the calculated skill scores for the GCM runs remaining from Step 2. The score ranking (Table 3) results in selecting the GCM runs for the final ensemble (Table 4). Note that this way of selecting GCM runs will omit those runs that have good skill for certain subdomains, for example, EC-EARTH_r2i1p1 for the Upper Ganges. The GCM runs that were pre-selected for the warm, wet corner are the models that do not perform well in simulating the present climate. For the warm, wet corner in the RCP4.5 scenario, the selected GCM ranked only 10th in the overall final ranking. This may be indicative that such a future is less probable. Another striking feature is that the same two models (inmcm4_r1i1p1 and CMCC-CMS_r1i1p1) are selected for both RCP4.5 and RCP8.5.

Figures 6 and 7 show the monthly average of mean air temperature and precipitation, respectively, for each subdomain for all models that were selected during Step 2. As could be expected, a large difference between all models is witnessed. Figure 6 shows that all GCMs show a cold bias in wintertime for the Upper Indus. Some models also have difficulties in simulating the annual precipitation cycle (Figure 7). Especially in the Upper Indus, the models diverge in the simulation of the annual cycle. Mountainous climate is less densely monitored, leading to larger errors in mountainous areas in reference climate datasets, such as WFDEI. The larger biases between the simulation models and WFDEI over the mountainous areas may be partly attributed to this. The spread between the climate models is large, but there is reasonable agreement for most of the subdomains. The mean and annual cycle are reasonably captured. Most GCM runs underestimate the precipitation in the lower subdomains. The analysis on the annual cycles indicates that the selected models are correctly selected. The large, observed biases emphasize the necessity of applying downscaling and bias-correction methods before the GCM outputs can be used in an impact study.

RCP	projection	model	ΔR99P (%)	ΔCDD (%)	ΔWSDI (%)	ΔCSDI (%)	ΔT (°C)	ΔP (%)	T_{index} rank	P_{index} rank	Combined score
RCP4.5	Warm, dry	CMCC-CM_r1i1p1	98.3	10.5	1162.5	−97.1	3.3	4.2	4	4	4.0
		CMCC-CMS_r1i1p1	56.9	14.3	938.2	−98.4	3.5	−2.9	2	5	3.5
		IPSL-CM5A-LR_r2i1p1	72.2	4.8	886.1	−96.3	3.4	3.1	1	3	2.0
		IPSL-CM5A-LR_r3i1p1	72.9	3.8	961.0	−96.0	3.3	6.2	3	2	2.5
		IPSL-CM5A-MR_r1i1p1	75.6	1.4	1472.0	−99.7	3.5	6.2	5	1	3.0
	Warm, wet	CanESM2_r1i1p1	107.0	−10.0	633.9	−96.7	3.4	17.0	1	4	2.5
		CSIRO-Mk3-6-0_r3i1p1	57.8	−7.5	1600.8	−94.8	3.4	13.3	4	1	2.5
		CSIRO-Mk3-6-0_r4i1p1	109.3	−9.6	1530.3	−95.4	3.5	15.6	2	5	3.5
		CSIRO-Mk3-6-0_r5i1p1	87.7	−3.8	1634.4	−96.9	3.5	14.8	5	3	4.0
		CSIRO-Mk3-6-0_r6i1p1	69.1	−7.8	1563.8	−96.5	3.5	13.5	3	2	2.5
	Cold, wet	bcc-csm1-1_r1i1p1	104.7	−5.6	490.2	−88.8	2.2	14.5	2	4	3.0
		BNU-ESM_r1i1p1	140.3	−6.7	729.3	−85.9	2.5	12.7	3	5	4.0
		CCSM4_r2i1p1	66.9	−2.9	531.9	−84.8	2.0	9.7	4	3	3.5
		CNRM-CM5_r1i1p1	41.9	−0.7	454.2	−90.6	2.3	11.9	1	2	1.5
		IPSL-CM5B-LR_r1i1p1	35.1	−11.8	424.3	−80.7	2.1	12.9	5	1	3.0
	Cold, dry	bcc-csm1-1-m_r1i1p1	46.1	−2.3	569.3	−87.4	2.3	5.8	2	1	1.5
		CCSM4_r6i1p1	45.8	5.6	809.9	−88.2	2.2	4.5	1	5	3.0
		CESM1-BGC_r1i1p1	57.9	4.3	696.0	−85.6	2.1	6.1	3	4	3.5
		GISS-E2-R_r6i1p1	47.8	−2.2	764.7	−81.0	2.2	4.1	4	2	3.0
		inmcm4_r1i1p1	10.5	2.9	347.5	−63.4	1.7	2.7	5	3	4.0
RCP8.5	Warm, dry	CMCC-CMS_r1i1p1	140.2	24.1	1855.1	−99.4	6.3	−3.1	5	5	5.0
		IPSL-CM5A-LR_r1i1p1	125.4	3.3	1627.4	−99.3	6.5	3.5	2	1	1.5
		IPSL-CM5A-LR_r2i1p1	145.7	9.3	1346.7	−99.7	6.5	4.1	1	3	2.0
		IPSL-CM5A-LR_r3i1p1	143.9	6.4	1679.0	−99.7	6.4	6.9	4	2	3.0
		IPSL-CM5A-LR_r4i1p1	140.8	13.7	1651.8	−99.5	6.4	3.6	3	4	3.5
	Warm, wet	CanESM2_r1i1p1	219.0	−17.7	1378.5	−99.6	6.0	28.9	2	2	2.0
		CanESM2_r2i1p1	227.4	−15.2	1623.6	−99.8	6.1	36.9	5	3	4.0
		CanESM2_r3i1p1	295.5	−11.7	1426.1	−100.0	6.1	37.4	3	5	4.0
		CanESM2_r4i1p1	192.9	−12.6	1460.6	−100.0	6.0	29.4	4	1	2.5
		CanESM2_r5i1p1	261.7	−10.1	1242.2	−99.8	6.0	37.0	1	4	2.5
	Cold, wet	bcc-csm1-1_r1i1p1	220.6	−9.6	1215.6	−97.1	4.4	29.7	5	5	5.0
		CNRM-CM5_r1i1p1	75.6	−3.4	884.7	−99.3	3.9	17.4	2	1	1.5
		IPSL-CM5B-LR_r1i1p1	88.4	−16.7	897.8	−97.1	4.2	23.4	4	2	3.0
		MIROC5_r1i1p1	145.3	−12.4	1337.8	−99.8	4.3	17.7	1	3	2.0
		MRI-CGCM3_r1i1p1	157.0	−12.6	1452.5	−97.4	4.4	22.1	3	4	3.5
	Cold, dry	CCSM4_r1i1p1	112.5	7.9	1451.4	−99.7	4.3	11.0	1	4	2.5
		CCSM4_r6i1p1	102.8	5.0	1858.5	−98.3	4.2	11.4	4	1	2.5
		EC-EARTH_r2i1p1	161.6	9.2	1984.5	−99.3	4.4	6.8	2	5	3.5
		EC-EARTH_r9i1p1	175.7	5.5	1994.2	−98.9	4.3	7.4	3	2	2.5
		inmcm4_r1i1p1	30.3	5.6	905.6	−89.6	3.6	5.4	5	3	4.0

Figure 4. GCM runs analysed during the refined selection step. Models selected for Step 3 are indicated with yellow colour.

4.2. Future climate in the Indus, Ganges and Brahmaputra basins

Averaged over the basins, according to the selected ensembles of GCM runs, mean air temperature increases by 1.7–3.5 °C for RCP4.5 and 3.6–6.3 °C for RCP8.5 between 1971–2000 and 2071–2100 (Table 4). However, there is large spatial variability within the study area (Figure 8). The largest increases in mean air temperature are projected for the upstream mountainous parts of the river basins, whilst the lowest increases are projected for the most downstream parts of the basins. The difference in projected temperature increase between the mountainous upstream and lower downstream parts of the river basin is several degrees. This elevation-dependent warming is in line with what was found in historical temperature records (Rangwala and Miller, 2012; Pepin *et al.*, 2015). Our analysis shows that the observed elevation-dependent warming may become more pronounced in the future. The
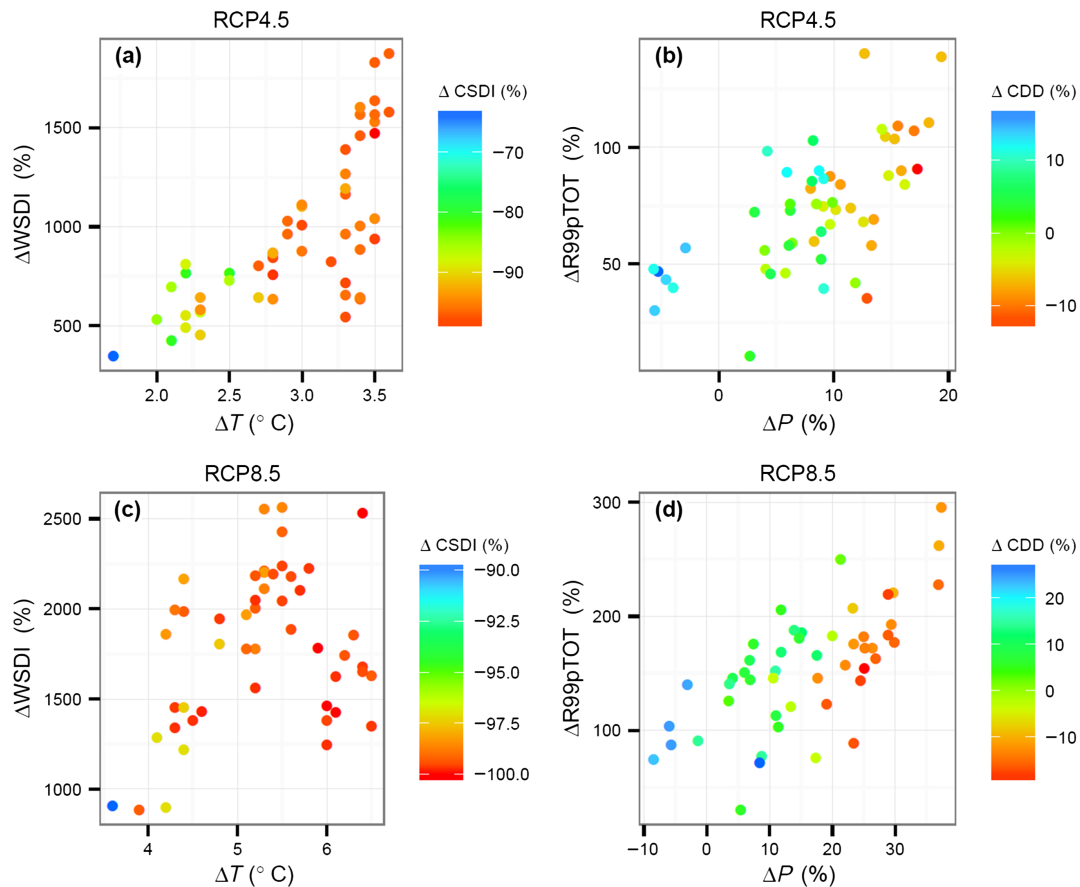
Figure 5. Projected changes in mean air temperature (ΔT), warm spell duration index (ΔWSDI), cold spell duration index (ΔCSDI) between 2071–2100 and 1971–2000 for RCP4.5 (a) and RCP8.5 (c). Projected changes in annual precipitation sum (ΔP), precipitation due to extremely wet days (ΔR99pTOT), consecutive dry days (ΔCDD) between 2071–2100 and 1971–2000 for RCP4.5 (b) and RCP8.5 (d).

uncertainty within the ensemble is, in general, slightly larger for the downstream basins compared to the upstream basins for RCP4.5, although the opposite is observed for the Brahmaputra basin. For RCP8.5, the largest uncertainty in the projections is observed for the upstream basins.

For precipitation in RCP4.5, the largest increases are projected for the eastern part of the Lower Indus and the western part of the Lower Ganges (Figure 8). At the same time, the uncertainty in the projections within the ensemble is also the largest for this region. The RCP8.5 ensemble shows the largest increases in precipitation in the lower Brahmaputra basin, with the largest uncertainty in the projection also in that region. The uncertainty in model projections is spatially more uniform for the RCP8.5 ensemble than for the RCP4.5 ensemble. Note that for both RCPs, the strongest decrease in precipitation is projected for the western part of the Indus basin. Traversing the Indus basin from west to east, the projected precipitation decrease changes to precipitation increase. These differences in projected precipitation changes may well be related to the contrasts in climate between the western Indus basin, with strong climatic influence from westerly systems, and the eastern Indus basin, which has a monsoon-dominated climate (Bookhagen and Burbank, 2010; Maussion *et al.*, 2014).

In addition to large spatial variability, seasonal variability in the climate change projections is large as well (Figure 9). The projected change in temperature shows a significant seasonal pattern for both RCPs. Averaged over the study area, both ensembles project the largest temperature increases for the drier winter months and smallest temperature increase for the wet summer months. For both RCPs, the uncertainty in projections is larger for the winter months compared to the monsoon season (June–September). The uncertainty in model projections for temperature is larger for the RCP8.5 ensemble than for the RCP4.5 ensemble.

The intra-annual projections of precipitation change show some remarkable features (Figure 9). For the RCP4.5 ensemble, in general, precipitation increases are projected for the monsoon season. Especially for July and August, the spread between the projections is extremely large (for July ranging from almost no increase to 225% increase). For January, the mean projection indicates decreasing precipitation, for February, unchanged precipitation and slight precipitation increase for March and April. For May and June, only small changes are projected. The uncertainty in the projections, however, is large here as well. The strong projected precipitation increase in December is also remarkable, with an extremely large uncertainty in the projections. For RCP8.5, the intra-annual patterns

Table 3. Ranking of the GCM runs for the six subdomains.

RCP	Projection	GCM run	Rank per subdomain						Total	General rank	Monsoon rank	Final rank
			LI	UI	LB	UB	LG	UG				
RCP4.5	Warm, dry	CMCC-CM_r1i1p1	8	10	1	14	7	5	45	7	7	6
		CMCC-CMS_r1i1p1	4	12	3	9	8	3	39	5	4	4
	Warm, wet	CSIRO-Mk3-6-0_r4i1p1	14	8	7	1	11	9	50	10	8	10
		CSIRO-Mk3-6-0_r5i1p1	13	7	8	2	12	13	55	14	9	14
	Cold, wet	BNU-ESM_r1i1p1	1	4	10	10	3	7	35	2	3	1
		CCSM4_r2i1p1	3	14	4	13	2	11	47	8	13	9
	Cold, dry	CESM1-BGC_r1i1p1	2	13	2	11	1	10	39	5	10	5
		inmcm4_r1i1p1	9	2	14	3	4	4	36	3	2	2
RCP8.5	Warm, dry	IPSL-CM5A-LR_r4i1p1	6	5	5	6	9	2	33	1	12	3
		CMCC-CMS_r1i1p1	4	12	3	9	8	3	39	5	4	4
	Warm, wet	CanESM2_r2i1p1	10	6	12	7	5	12	52	12	5	12
		CanESM2_r3i1p1	5	11	13	5	6	8	48	9	6	8
	Cold, wet	bcc-csm1-1_r1i1p1	7	1	9	4	10	14	45	7	11	7
		MRI-CGCM3_r1i1p1	12	3	6	12	13	6	52	12	14	13
	Cold, dry	EC-EARTH_r2i1p1	11	9	11	8	14	1	54	13	1	11
		inmcm4_r1i1p1	9	2	14	3	4	4	36	3	2	2

LI, Lower Indus; UI, Upper Indus; LB, Lower Brahmaputra; UB, Upper Brahmaputra; LG, Lower Ganges; UG, Upper Ganges.

Table 4. Final selected ensemble of GCM runs with projected changes in mean air temperature, precipitation and ETCCDI indices between 1971–2000 and 2071–2100 averaged over the study area.

RCP	Projection	GCM run	ΔT (°C)	ΔP (%)	ΔWSDI (%)	ΔCSDI (%)	ΔR99P (%)	ΔCDD (%)
RCP4.5	Cold, wet	BNU-ESM_r1i1p1	2.5	12.7	729.3	−85.9	140.3	−6.7
	Cold, dry	inmcm4_r1i1p1	1.7	2.7	347.5	−63.4	10.5	2.9
	Warm, dry	CMCC-CMS_r1i1p1	3.5	−2.9	938.2	−98.4	56.9	14.3
	Warm, wet	CSIRO-Mk3-6-0_r4i1p1	3.5	15.6	1530.3	−95.4	109.3	−9.6
RCP8.5	Cold, dry	inmcm4_r1i1p1	3.6	5.4	905.6	−89.6	30.3	5.6
	Warm, dry	CMCC-CMS_r1i1p1	6.3	−3.1	1855.1	−99.4	140.2	24.1
	Cold, wet	bcc-csm1-1_r1i1p1	4.4	29.7	1215.6	−97.1	220.6	−9.6
	Warm, wet	CanESM2_r3i1p1	6.1	37.4	1426.1	−100.0	295.5	−11.7

in precipitation change are quite similar. Remarkable is the observation that the extremely large model spread as observed for July, August and December in the RCP4.5 ensemble, does not occur for any month in the RCP8.5 ensemble. Large uncertainty is observed for most months in the RCP8.5 ensemble, with both positive and negative projections of precipitation change for 10 out of 12 months.

Regarding the projected changes in extremes, the wettest, driest, warmest and coldest projections for the two ensembles are analysed (Figure 10). For RCP4.5, the model with the largest projected increase in precipitation shows the largest increase in extreme precipitation events in the Ganges basin and the eastern part of the Lower Indus basin. Decreases in precipitation due to extremely wet days are projected for the western part of the Indus basin. This also holds for the model with the largest projected increase in precipitation for RCP8.5. The western part of the Lower Indus basin is an area where a decreasing precipitation trend is projected (Figure 8). The largest increase in precipitation due to extremely wet days is projected in the Brahmaputra basin according to this GCM run. An increase in dry spells is projected for almost the entire study area for both the RCP4.5 and RCP8.5

model ensemble. The strongest increases in dry spells are projected for the lower Ganges and upper Brahmaputra basins for both RCPs.

The duration of warm spells according to the warmest members in both ensembles increases for the entire study area. For both RCP4.5 and RCP8.5, the increase is the strongest for the Lower Indus basin. The duration of cold spells on the other hand, according to the coldest members in both ensembles, decreases for the entire study area. For RCP4.5, the decrease is the smallest in the Upper Indus and western part of the Lower Indus basin, whereas the strongest decreases are projected for the Upper Brahmaputra basin. For RCP8.5, some grid cells even show a decrease in CSDI of 100%, meaning no period of at least 6 days in a row with the daily minimum temperature lower than the tenth percentile temperature occurs during 2071–2100.

5. Discussion

Although the presented method aims to combine the strengths of envelope-based and past-performance-based selection of GCMs for impact studies, a few limitations remain.
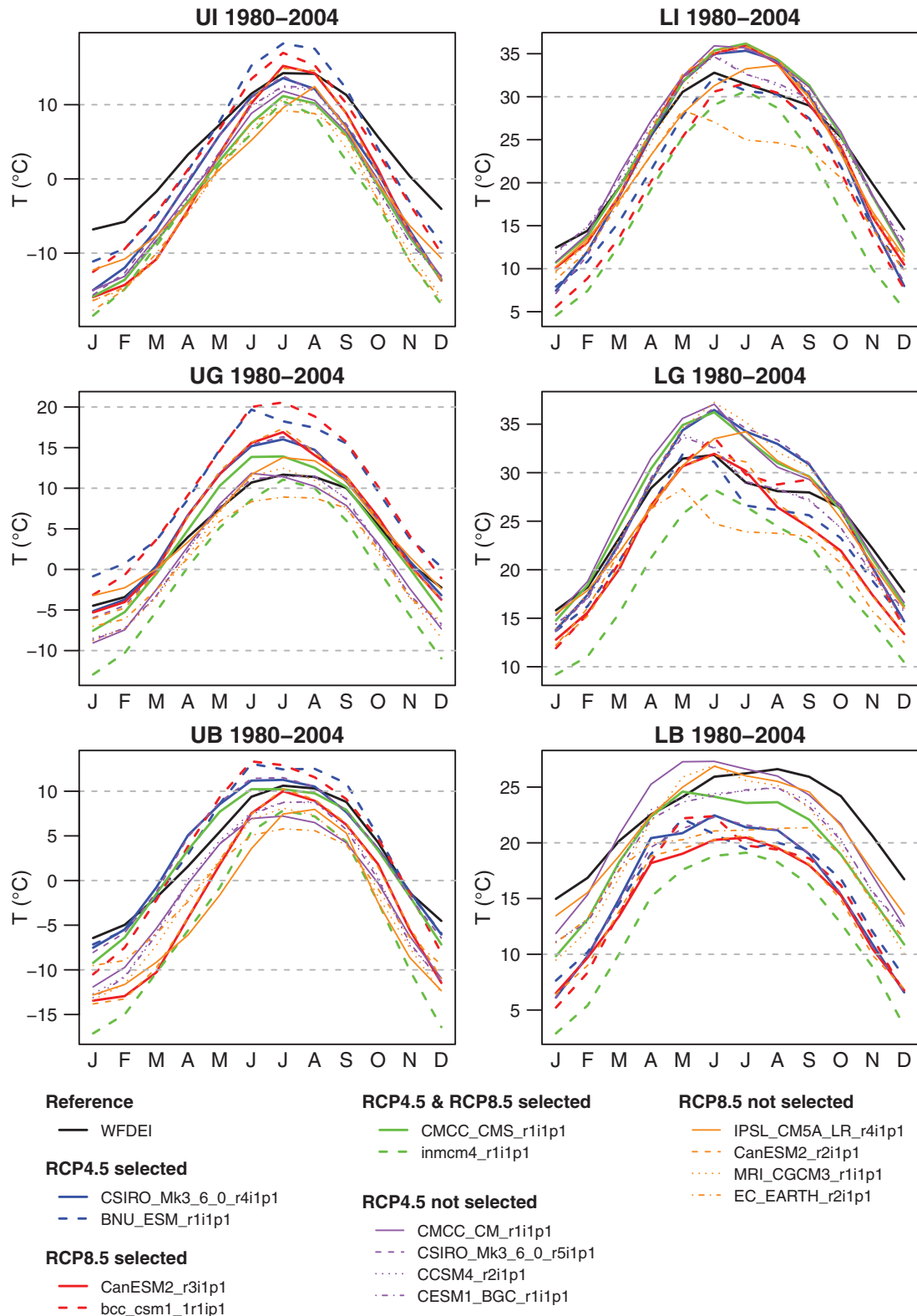
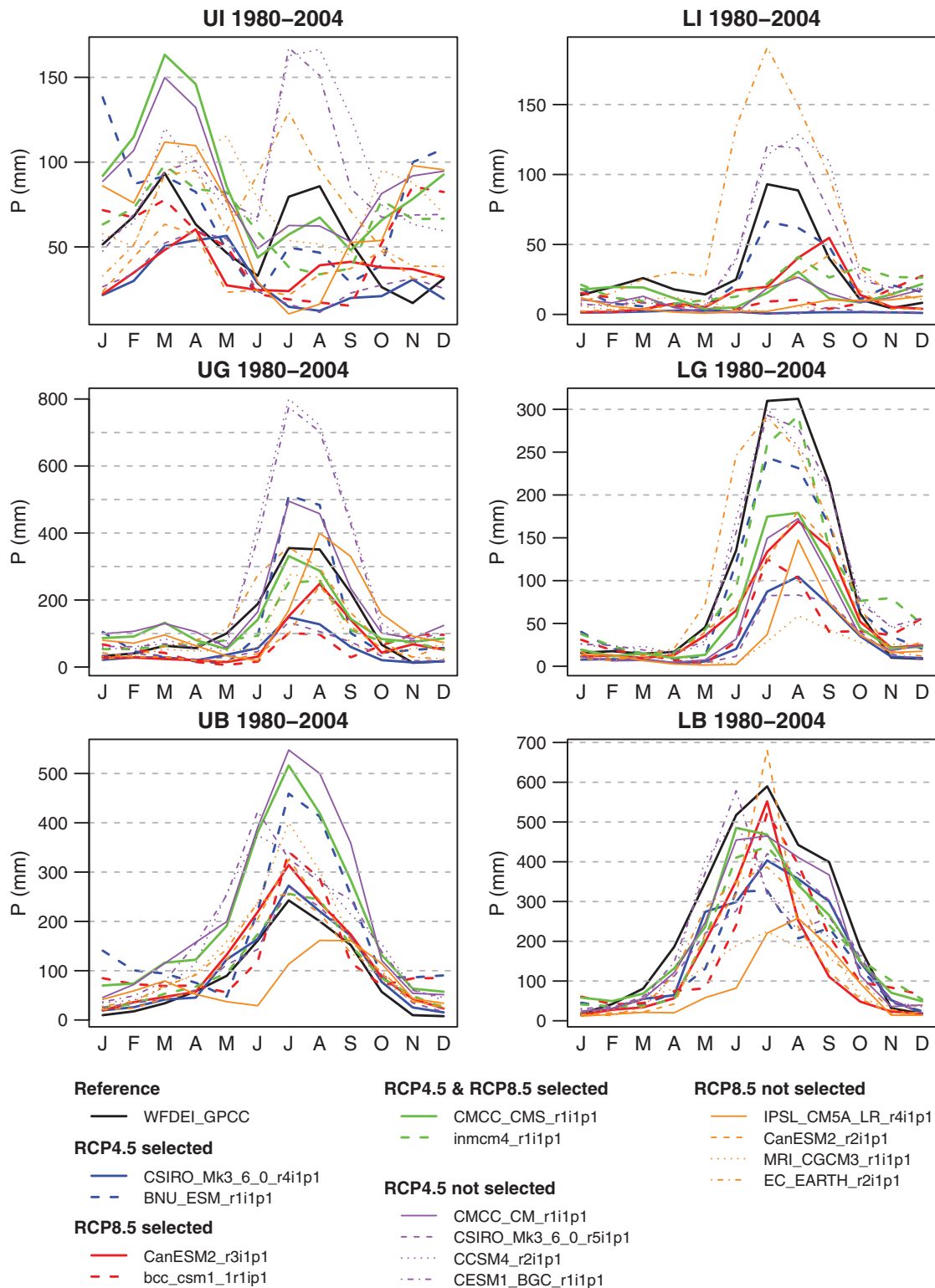Figure 6. Average annual cycles of temperature for WFDEI and GCM runs per subdomain. UI, Upper Indus; LI, Lower Indus; UG, Upper Ganges; LG, Lower Ganges; UB, Upper Brahmaputra; LB, Lower Brahmaputra.

The first issue is related to the scale of the application. During the first selection step, projected changes are averaged over the entire area, and this may dilute the spatial variation in projected changes. The same scaling issues apply to selection based on the changes in the extremes

in the second step of our approach. This is not unique to this approach but holds for most studies that cover a large spatial domain. A potential solution is to divide the study area into multiple parts and apply the selection approach to each part independently. However, the drawback of this

Figure 7. Average annual cycles of precipitation for WFDEI and GCM runs per subdomain. UI, Upper Indus; LI, Lower Indus; UG, Upper Ganges; LG, Lower Ganges; UB, Upper Brahmaputra; LB, Lower Brahmaputra.

approach is the introduction of physical inconsistencies and erroneous boundary effects in the climate forcing at the transition from one GCM to another.

The second limitation relates to the fact that the envelope of changes in means is leading in the selection approach. This may result in a reduction in the range of projections

of changes in climatic extremes in the ensemble. The same holds for the final step when model runs are selected based on their skill in simulating past climate. Because the selection based on skill is introduced in the last step, with less GCM runs available to select from than before Steps 1 and 2, the selected models do not necessarily have the
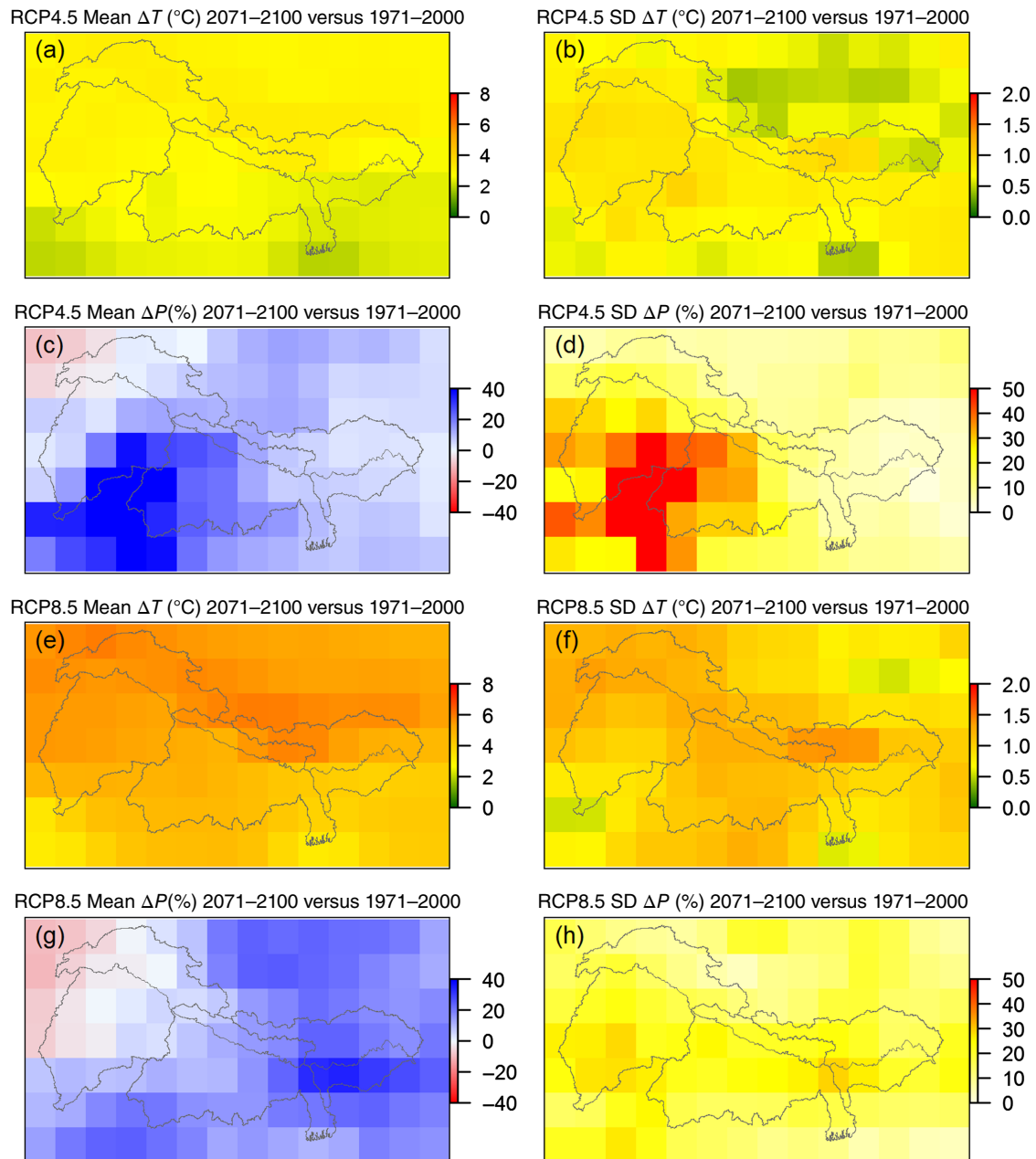
Figure 8. Projected changes between 1971–2000 and 2071–2100 in mean air temperature and annual precipitation sum in the Indus, Ganges and Brahmaputra basins. Panels a–d show results for RCP4.5 model ensemble and panels e–h show results for RCP8.5 model ensemble. Left panels show the mean of the model ensemble, and right panels show the standard deviation of the model ensemble. Grey lines indicate upper and lower basin boundaries.

best skill in simulating past climate. This indicates that the sequence of selection, or weighting of different criteria, is not fixed and may be subject to changes, depending on the type of climate change impact study. Another approach combining skill and envelope is introduced by McSweeney *et al.* (2012). For a study area in Vietnam, the researchers first eliminate model runs that do not have sufficient skill in simulating the monsoon, given its importance for their study area, before assessing the range of projections provided by the remaining models. In our approach we choose the range in projected climate change to be leading over the historical performance to ensure including all possible futures as projected by the CMIP5

multi-model ensemble, which is desirable for planning of adaptation strategies. Previous analyses of CMIP5 GCMs' skills to simulate the precipitation patterns in the study area, which are largely dominated by monsoon dynamics, showed that CMIP5 GCMs have poor skills in simulating the important features of the precipitation dynamics in our study area (Sperber *et al.*, 2013; Palazzi *et al.*, 2014; Sperber and Annamalai, 2014). As no single model or group of models clearly stands out in performance for the study area, this constitutes another reason to prefer to include all possible futures. To show how the chosen ordering of selection steps affects the model skill of the selected ensembles, we test how the model skill
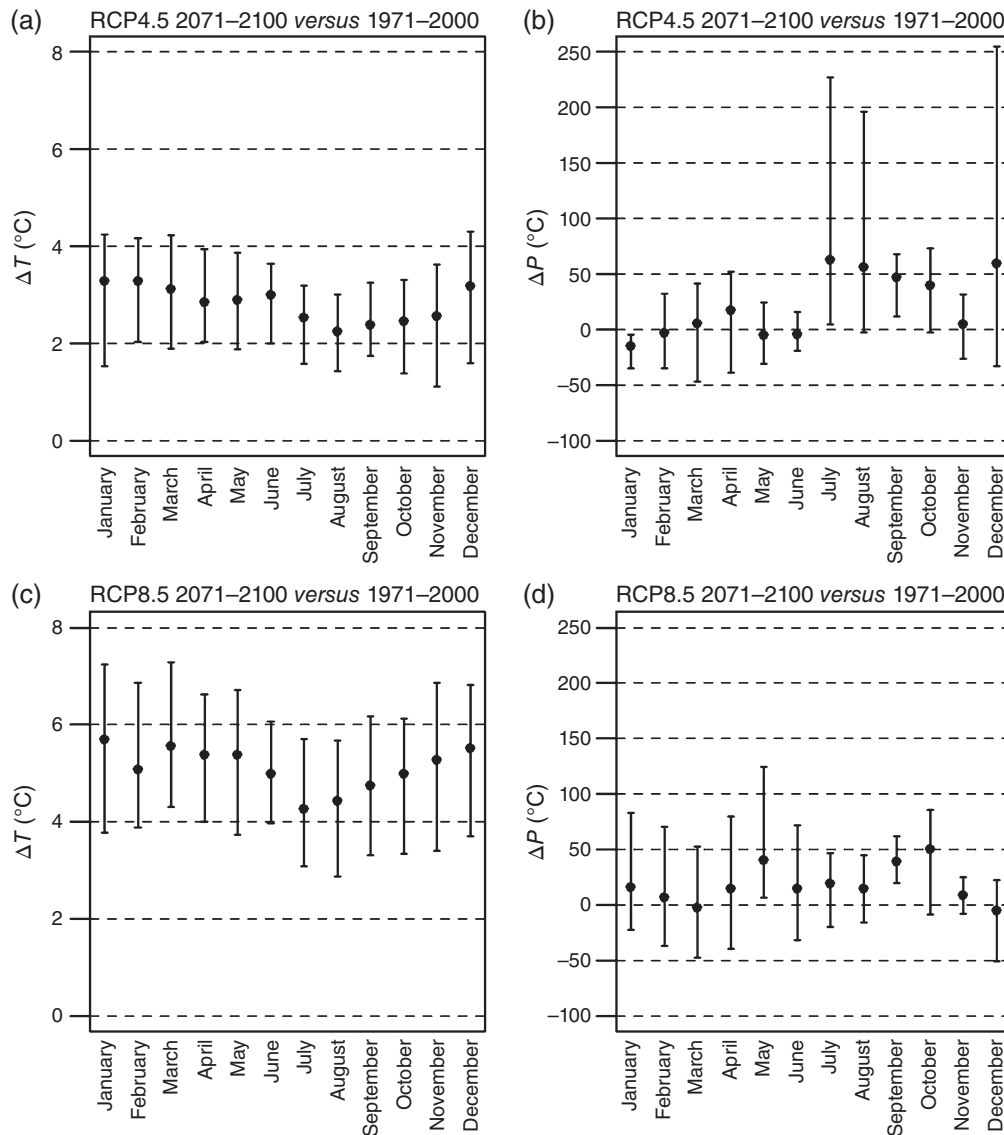
Figure 9. Study-area averaged monthly projected changes in temperature and precipitation between 1971–2000 and 2071–2100. Dots indicate model ensemble mean projection, and whiskers indicate the range of projections according to the model ensemble.

of selected models compares to model skill in the total model ensembles. Figure 11 shows how the biases in air temperature and precipitation in the ensembles of model runs remaining after selection Step 2 correspond to the biases in the total model ensembles for winter season, monsoon season and on an annual scale. Although the distributions differ slightly and especially for the winter season, the distributions of the selected models and total model ensembles are generally in the same magnitude, indicating that the skills of the selected models are comparable to the skills of the models in the entire ensemble. Thus, we conclude that there is no significant bias towards models with poorer skills in the selected model ensembles.

Third, the weighting of different skill scores introduces a degree of subjectivity. In our approach of testing a model's past-performance, more weight is assigned to the model's skill in simulating the entire annual cycle of precipitation and temperature than to its skill in simulating the monsoon.

Furthermore, equal weight is assigned to the indicators for precipitation extremes and temperature extremes in calculating model scores in Step 2 of the presented approach. Larger weight has been assigned to the model's skill to simulate the entire annual cycle of precipitation and temperature and equal weight to air temperature and precipitation extremes because cryospheric processes in the upstream parts of the basins are important for the hydrological regimes of these basins (Schaner *et al.*, 2012; Lutz *et al.*, 2014). Changes in these cryospheric processes are driven by the combined effect of changes in precipitation and temperature. In each of these issues, the weighting could be adjusted for other studies, depending on which climatic variables are more important in each particular case.

Finally, our approach assumes independency of all model runs, whereas some models share identical model code or use the same forcing and validation data, leading to model interdependency (Jun *et al.*, 2008; Masson and
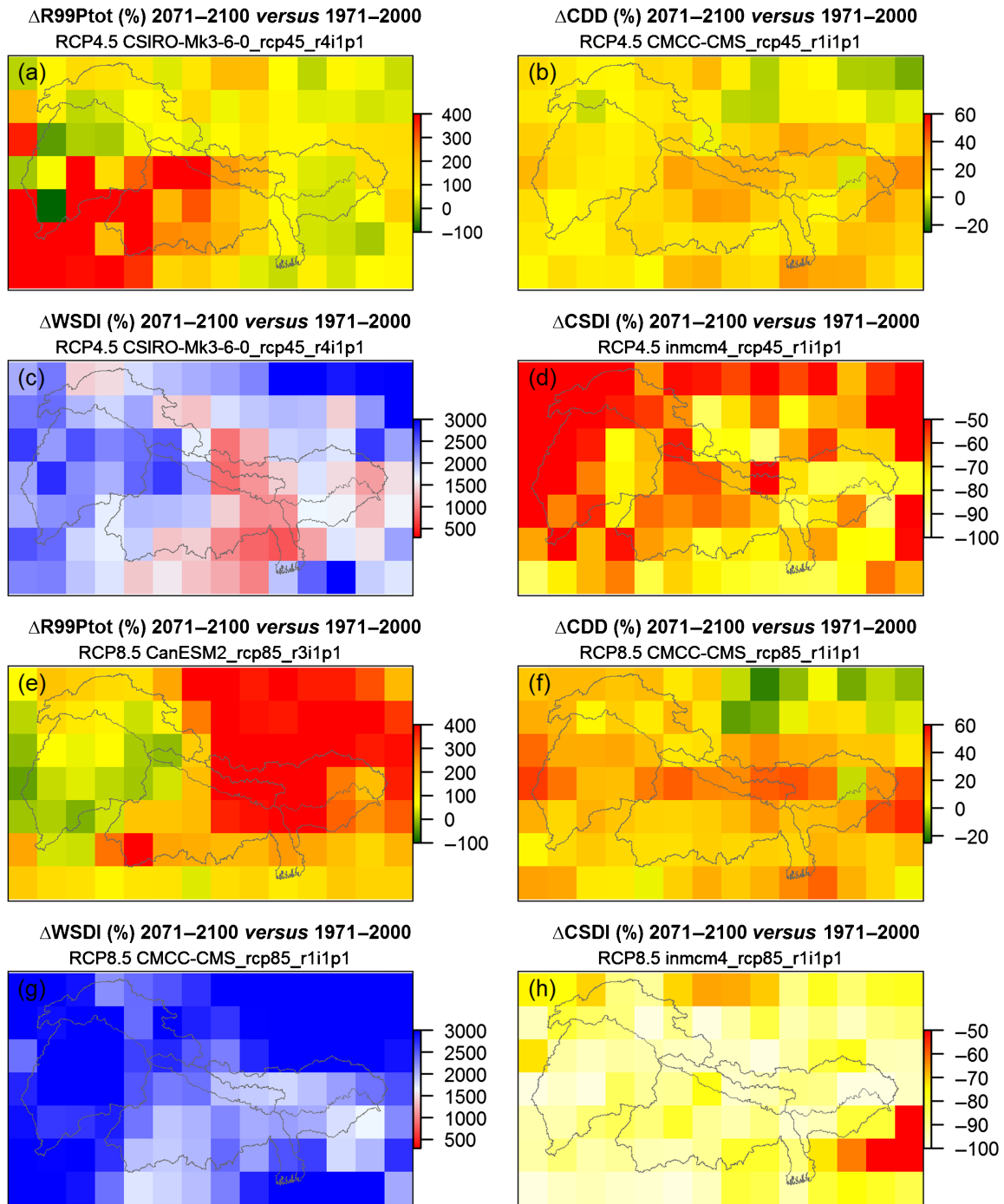
Figure 10. Projected changes in ETCCDI indices between 1971–2000 and 2071–2100 for RCP4.5 ((a)–(d)) and RCP 8.5 ((e)–(h)). Changes in the corresponding index are shown for the wettest ((a) and (e)), driest ((b) and (f)), warmest ((c) and (g)) and coldest ((d) and (h)) members in the four-model ensembles.

Knutti, 2011; Knutti *et al.*, 2013). In our case where multiple initial condition ensemble members of the same GCM are included, this interdependency is particularly large. The approach as described here could be expanded by considering weighting metrics for the degree of CMIP5 model interdependencies (e.g. Sanderson *et al.*, 2015a, 2015b).

The use of GCMs for climate change impact studies is common practice, although the usefulness of this approach has recently been questioned (Kundzewicz and Stakhiv, 2010; Bakker, 2015). As also demonstrated in the present study, the biases between GCM simulations and observations are large, and many GCMs lack the skill to simulate important regional climatic features, such as the Asian monsoon in our case (Sperber *et al.*, 2013). Dynamic and empirical–statistical downscaling and bias-correction methods may help to reduce the biases but cannot completely bridge them (Pielke and Wilby, 2012). This stresses that the currently available climate change projections must be used with caution when defining climate change adaptation policies.
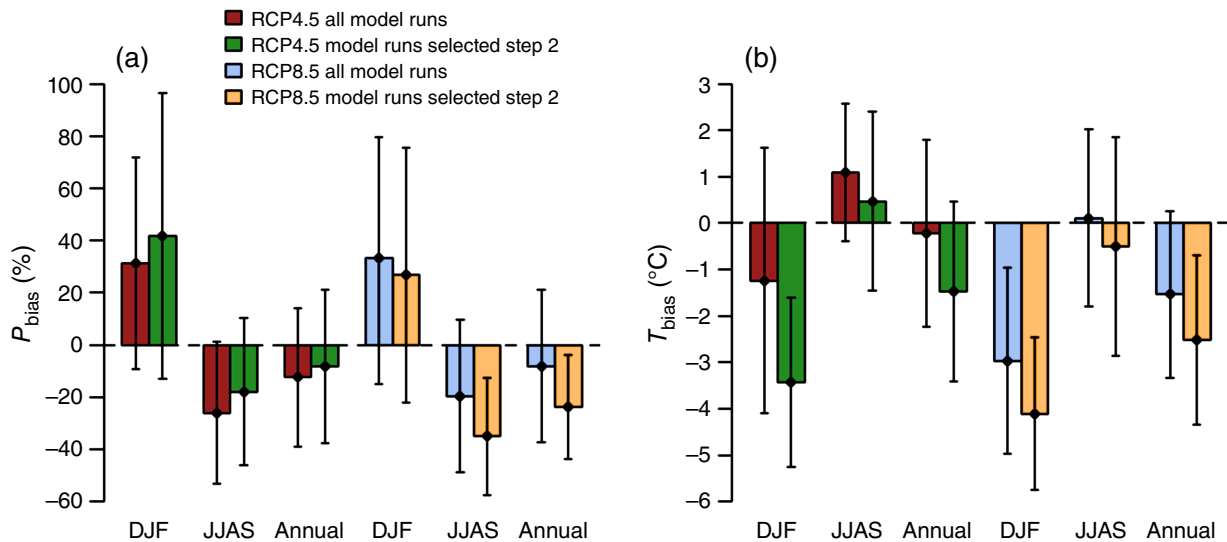
Figure 11. Comparison of GCM biases in all model runs and model runs remaining after selection step 2. Biases are calculated using monthly data for 1980–2004 with respect to WFDEI data (Weedon *et al.*, 2014) for precipitation (a) and mean air temperature (b). Biases are calculated for the winter season (DJF), monsoon season (JJAS) and annually. Bars indicate the ensemble mean biases and whiskers indicate the standard deviations of the ensemble's biases.

6. Conclusions

The selection of climate models for climate change impact studies is not straightforward, while at the same time, it is a crucial step when conducting such a study. The approach presented here seeks the optimal balance between ensuring that the selected GCMs represent changes in average and extreme climatic conditions well but at the same time have reasonable skill in simulating the past climate, with a particular focus on the monsoon dynamics.

The ensembles of selected GCM runs for RCP4.5 and RCP8.5 show that the uncertainty of future climate in this region is very large. Projections of mean air temperature indicate an increase ranging from 1.7 °C to 6.3 °C between 1971–2000 and 2071–2100, averaged over the three river basins, with stronger warming at higher altitudes. The uncertainty in future precipitation is larger, with area-averaged projections ranging from −3.1 to +37.4%. All GCM runs included in the selected ensembles project increases in warm spells and decreases in cold spells. Besides, all selected GCM runs project increases in extremely high precipitation events, even the GCM runs that project decreases in annual precipitation sum. Model runs projecting decreases in precipitation project increasing dry spells, and model runs projecting strong increases in precipitation project a reduction of dry spells. However, an increase in dry spells is projected by the model runs that project a minor increase in precipitation. These numbers are averages over the entire Indus, Ganges and Brahmaputra river basins. Spatial variability as well as seasonal variability, in terms of the mean projections and the uncertainty in the projections, are very large. This means that the future climate of the region remains very uncertain, which may compromise defining adequate adaptation policies.

Supporting Information

The following supporting information is available as part of the online article:

Table S1. Complete pool of RCP4.5 model runs considered in this study.

Table S2. Complete pool of RCP8.5 model runs considered in this study.

References

Bakker AMR. 2015. *The Robustness of the Climate Modelling Paradigm.* PhD thesis, Vrije Universiteit, Amsterdam.

Biemans H, Speelman LH, Ludwig F, Moors EJ, Wiltshire AJ, Kumar P, Gerten D, Kabat P. 2013. Future water resources for food production in five South Asian river basins and potential for adaptation – a modeling study. *Sci. Total Environ.* **468–469**: S117–S131, doi: 10.1016/j.scitotenv.2013.05.092.

Blázquez J, Nuñez MN. 2013. Analysis of uncertainties in future climate projections for South America: comparison of WCRP-CMIP3 and WCRP-CMIP5 models. *Clim. Dyn.* **41**(3–4): 1039–1056, doi: 10.1007/s00382-012-1489-7.

Bookhagen B, Burbank DW. 2010. Toward a complete Himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *J. Geophys. Res.* **115**(F3): 1–25, doi: 10.1029/2009JF001426.

Cannon AJ. 2014. Selecting GCM scenarios that span the range of changes in a multimodel ensemble: application to CMIP5 climate extremes indices. *J. Clim.* **28**: 1260–1267, doi: 10.1175/JCLI-D-14-00636.1.

De Souza K, Kituyi E, Harvey B, Leone M, Murali KS, Ford JD. 2015. Vulnerability to climate change in three hot spots in Africa and Asia: key issues for policy-relevant adaptation and resilience-building research. *Reg. Environ. Change* **15**: 747–753, doi: 10.1007/s10113-015-0755-8.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**(656): 553–597, doi: 10.1002/qj.828.

Evans JP, Ji F, Abramowitz G, Ekström M. 2013. Optimally choosing small ensemble members to produce robust climate simulations. *Environ. Res. Lett.* **8**(4): 1–4, doi: 10.1088/1748-9326/8/4/044050.

Finger D, Heinrich G, Gobiet A, Bauder A. 2012. Projections of future water resources and their uncertainty in a glacierized catchment in the Swiss Alps and the subsequent effects on hydropower production during the 21st century. *Water Resour. Res.* **48**(2): 1–20, doi: 10.1029/2011WR010733.

Houle D, Bouffard A, Duchesne L, Logan T, Harvey R. 2012. Projections of future soil temperature and water content for three Southern Quebec forested sites. *J. Clim.* **25**(21): 7690–7701, doi: 10.1175/JCLI-D-11-00440.1.

Immerzeel WW, Pellicciotti F, Bierkens MFP. 2013. Rising river flows throughout the twenty-first century in two Himalayan glacierized watersheds. *Nat. Geosci.* **6**(8): 1–4, doi: 10.1038/ngeo1896.

IPCC. 2007. *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Mitigation of Climate Change.* IPCC: Geneva, Switzerland.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* IPCC: Cambridge, UK.

Joetzjer E, Douville H, Delire C, Ciais P. 2013. Present-day and future Amazonian precipitation in global climate models: CMIP5 versus CMIP3. *Clim. Dyn.* **41**(11–12): 2921–2936, doi: 10.1007/s00382-012-1644-1.

Jun M, Knutti R, Nychka DW. 2008. Spatial analysis to quantify numerical model bias and dependence. *J. Am. Stat. Assoc.* **103**(483): 934–947, doi: 10.1198/016214507000001265.

Kjellström E, Boberg F, Castro M, Christensen JH, Nikulin G, Sánchez E. 2010. Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Clim. Res.* **44**: 135–150, doi: 10.3354/cr00932.

Knutti R, Sedláček J. 2012. Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Change* **3**: 369–373, doi: 10.1038/nclimate1716.

Knutti R, Masson D, Gettelman A. 2013. Climate model genealogy: generation CMIP5 and how we got there. *Geophys. Res. Lett.* **40**(6): 1194–1199, doi: 10.1002/grl.50256.

Kundzewicz ZW, Stakhiv EZ. 2010. Are climate models "ready for prime time" in water resources management applications, or is more research needed? *Hydrol. Sci. J.* **55**(7): 1085–1089, doi: 10.1080/02626667.2010.513211.

Lutz AF, Immerzeel WW, Gobiet A, Pellicciotti F, Bierkens MFP. 2013. Comparison of climate change signals in CMIP3 and CMIP5 multi-model ensembles and implications for Central Asian glaciers. *Hydrol. Earth Syst. Sci.* **17**(9): 3661–3677, doi: 10.5194/hess-17-3661-2013.

Lutz AF, Immerzeel WW, Shrestha AB, Bierkens MFP. 2014. Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Clim. Change* **4**: 587–592, doi: 10.1038/NCLIMATE2237.

Masson D, Knutti R. 2011. Climate model genealogy. *Geophys. Res. Lett.* **38**(8): 1–4, doi: 10.1029/2011GL046864.

Maussion F, Scherer D, Mölg T, Collier E, Curio J, Finkelnburg R. 2014. Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia reanalysis. *J. Clim.* **27**(5): 1910–1927, doi: 10.1175/JCLI-D-13-00282.1.

McSweeney CF, Jones RG, Booth BBB. 2012. Selecting ensemble members to provide regional climate change information. *J. Clim.* **25**(20): 7100–7121, doi: 10.1175/JCLI-D-11-00526.1.

Meehl GA, Covey C, Delworth T, Latif M, Mcavaney B, Mitchell JFB, Stouffer RJ, Taylor KE. 2007. The WCRP CMIP3 multimodel dataset: a new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**(9): 1382–1394.

Minville M, Brissette F, Leconte R. 2008. Uncertainty of the impact of climate change on the hydrology of a Nordic watershed. *J. Hydrol.* **358**(1–2): 70–83, doi: 10.1016/j.jhydrol.2008.05.033.

Nepal S, Shrestha AB. 2015. Impact of climate change on the hydrological regime of the Indus, Ganges and Brahmaputra river basins: a review of the literature. *Int. J. Water Resour. Dev.* **31**(2): 201–218, doi: 10.1080/07900627.2015.1030494.

Palazzi E, von Hardenberg J, Terzago S, Provenzale A. 2014. Precipitation in the Karakoram-Himalaya: a CMIP5 view. *Clim. Dyn.* **45**: 21–45, doi: 10.1007/s00382-014-2341-z.

Pepin N, Bradley RS, Diaz HF, Baraer M, Caceres EB, Forsythe N, Fowler H, Greenwood G, Hashmi MZ, Liu XD, Miller JR, Ning L, Ohmura A, Palazzi E, Rangwala I, Schöner W, Severskiy I, Shahgedanova M, Wang MB, Williamson SN, Yang DQ. 2015. Elevation-dependent warming in mountain regions of the world. *Nat. Clim. Change* **5**: 424–430, doi: 10.1038/nclimate2563.

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J. 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Clim.* **20**: 4356–4376, doi: 10.1175/JCLI4253.1.

Peterson TC. 2005. Climate change indices. *WMO Bull.* **54**(2): 83–86.

Pielke RA, Wilby RL. 2012. Regional climate downscaling: what's the point? *Eos* **93**(5): 52–53, doi: 10.1002/wea543.Wilby.

Pierce DW, Barnett TP, Santer BD, Gleckler PJ. 2009. Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. U. S. A.* **106**(21): 8441–8446, doi: 10.1073/pnas.0900094106.

Rangwala I, Miller JR. 2012. Climate change in mountains: a review of elevation-dependent warming and its possible causes. *Clim. Change* **114**: 527–547, doi: 10.1007/s10584-012-0419-3.

Sanchez E, Romera R, Gaertner MA, Gallardo C, Castro M. 2009. A weighting proposal for an ensemble of regional climate models over Europe driven by 1961–2000 ERA40 based on monthly precipitation probability density functions. *Atmos. Sci. Lett.* **10**: 249–254, doi: 10.1002/asl.

Sanderson BM, Knutti R, Caldwell P. 2015a. A representative democracy to reduce interdependency in a multimodel ensemble. *J. Clim.* **28**(13): 5171–5194, doi: 10.1175/JCLI-D-14-00362.1.

Sanderson BM, Knutti R, Caldwell P. 2015b. Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J. Clim.* **28**(13): 5150–5170, doi: 10.1175/JCLI-D-14-00361.1.

Schaner N, Voisin N, Nijssen B, Lettenmaier DP. 2012. The contribution of glacier melt to streamflow. *Environ. Res. Lett.* **7**: 034029, doi: 10.1088/1748-9326/7/3/034029.

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B. 2013. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* **115**: 15–40, doi: 10.1007/s00704-013-0860-x.

Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D. 2013a. Climate extremes indices in the CMIP5 multimodel ensemble: part

1. Model evaluation in the present climate. *J. Geophys. Res.-Atmos.* **118**(4): 1716–1733, doi: 10.1002/jgrd.50203.

Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D. 2013b. Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. *J. Geophys. Res.-Atmos.* **118**(6): 2473–2493, doi: 10.1002/jgrd.50188.

Sorg A, Huss M, Rohrer M, Stoffel M. 2014. The days of plenty might soon be over in glacierized Central Asian catchments. *Environ. Res. Lett.* **9**(10), doi: 10.1088/1748-9326/9/10/104018.

Sperber KR, Annamalai H. 2014. The use of fractional accumulated precipitation for the evaluation of the annual cycle of monsoons. *Clim. Dyn.* **43**(12): 3219–3244, doi: 10.1007/s00382-014-2099-3.

Sperber KR, Annamalai H, Kang IS, Kitoh A, Moise A, Turner A, Wang B, Zhou T. 2013. The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.* **41**(9–10): 2711–2744, doi: 10.1007/s00382-012-1607-6.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**(4): 485–498, doi: 10.1175/BAMS-D-11-00094.1.

van Vuuren DP, Stehfest E, den Elzen MGJ, van Vliet J, Isaac M. 2010. Exploring IMAGE model scenarios that keep greenhouse gas radiative forcing below 3W/m^2 in 2100. *Energy Econ.* **32**(5): 1105–1120, doi: 10.1016/j.eneco.2010.03.001.

van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK. 2011a. The representative concentration pathways: an overview. *Clim. Change* **109**(1–2): 5–31, doi: 10.1007/s10584-011-0148-z.

van Vuuren DP, Stehfest E, den Elzen MGJ, Kram T, van Vliet J, Deetman S, Isaac M, Klein Goldewijk K, Hof A, Mendoza Beltran A, Oostenrijk R, van Ruijven B. 2011b. RCP2.6: exploring the possibility to keep global mean temperature increase below 2°C. *Clim. Change* **109**: 95–116, doi: 10.1007/s10584-011-0152-3.

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O, Schewe J. 2014. The inter-sectoral impact model intercomparison project (ISI-MIP): project framework. *Proc. Natl. Acad. Sci. U. S. A.* **111**(9): 3228–3232, doi: 10.1073/pnas.1312330110.

Weedon GP, Gomes S, Viterbo P, Shuttleworth WJ, Blyth E, Österle H, Adam JC, Bellouin N, Boucher O, Best M. 2011. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.* **12**(5): 823–848, doi: 10.1175/2011 JHM1369.1.

Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P. 2014. The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-interim reanalysis data. *Water Resour. Res.* **50**(9): 7505–7514, doi: 10.1002/2014WR015638.