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, (Penterson, 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 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 ($\Delta T$) and annual precipitation sum ($\Delta P$) between 1971–2000 and 2071–2100, averaged over the $2.5^\circ \times 2.5^\circ$ 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 $\Delta T$ and $\Delta P$ are determined after resampling all GCM data to the same $2.5^\circ \times 2.5^\circ$ grid. These values represent the four corners of the spectrum of projections for temperature and precipitation change. The tenth percentile value for $\Delta T$ and tenth percentile value for $\Delta P$ are in the ‘cold, dry’ corner of the spectrum. The 10th percentile value for $\Delta T$ and 90th percentile value for $\Delta P$ are in the ‘cold, wet’ corner of the spectrum. The 90th percentile value for $\Delta T$ and 10th percentile value for $\Delta P$ are in the ‘warm, dry’ corner of the spectrum. The 90th percentile value for $\Delta T$ and 90th percentile value for $\Delta P$ 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 $\Delta T$ and $\Delta P$ with respect to the entire range of projections in the entire ensemble:

$$D_{P,T}^i = \sqrt{\left(\frac{|P_{T,j} - P_{T,i}|}{P_{T,j}}\right)^2 + \left(\frac{|P_{P,j} - P_{P,i}|}{P_{P,j}}\right)^2}$$

(1)

where $D_{P,T}^i$ is the distance of a model (j)’s $\Delta T$ and $\Delta P$ ($P_{T,j}$ and $P_{P,j}$ respectively) to the corner (i)’s 10th and/or 90th percentile score of $\Delta T$ and $\Delta P$ for the entire ensemble ($P_{T,i}^j$ and $P_{P,i}^j$ 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 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.

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

<table>
<thead>
<tr>
<th>Meteorological variable</th>
<th>ETCCDI index</th>
<th>Index description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>R99pTOT</td>
<td>Precipitation due to extremely wet days (&gt;99th percentile)</td>
</tr>
<tr>
<td>Precipitation</td>
<td>CDD</td>
<td>Consecutive dry days: maximum length of dry spell (P &lt; 1 mm)</td>
</tr>
<tr>
<td>Air temperature</td>
<td>WSDI</td>
<td>Warm spell duration index: count of days in a span of at least 6 days where TX &gt; 90th percentile (TXij is the daily T_max on day i in period j)</td>
</tr>
<tr>
<td>Air temperature</td>
<td>CSDI</td>
<td>Cold spell duration index: count of days in a span of at least 6 days where TN &lt; 10th percentile (TNij is the daily T_min on day i in period j)</td>
</tr>
</tbody>
</table>
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:

\[ S_{\text{score}} = \sum_{i=1}^{n} \min (Z_{GCM} - Z_{OBS}) \]  

(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 (\( S_{\text{score}} \)) 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:

\[ f_1 = 1 - \left( \frac{A_{GCM} - A_{WFDEI}}{2 \cdot A_{WFDEI}} \right)^{0.5} \]  

(3)

\[ f_2 = 1 - \left( \frac{A^+_{GCM} - A^+_{WFDEI}}{2 \cdot A^+_{WFDEI}} \right)^{0.5} \]  

(4)

\[ f_3 = 1 - \left( \frac{A^-_{GCM} - A^-_{WFDEI}}{2 \cdot A^-_{WFDEI}} \right)^{0.5} \]  

(5)

\[ f_4 = 1 - \left( \frac{P_{GCM} - P_{WFDEI}}{2 \cdot P_{WFDEI}} \right)^{0.5} \]  

(6)

\[ f_5 = 1 - \left( \frac{\sigma_{GCM} - \sigma_{WFDEI}}{2 \cdot \sigma_{WFDEI}} \right)^{0.5} \]  

(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 \( \sigma \) 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 for each 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 underestimate the precipitation in the monsoon period.
Selected models warm, dry corner
Selected models warm, wet corner
10th and 90th percentile values $\Delta T$ and $\Delta P$
Selected models cold, dry corner
Selected models cold, wet corner
Other model runs

Figure 3. Projected changes in mean air temperature ($\Delta T$) and annual precipitation sum ($\Delta P$) 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 $\Delta T$ and $\Delta P$. The model runs selected for Step 2 are indicated with colours.

(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 ($\Delta T$) and annual precipitation sums ($\Delta P$) 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 $\Delta T$ and $\Delta P$ is much larger for the RCP8.5 model pool than for the RCP4.5 model pool. $\Delta T$ ranges from 1.7°C to 3.6°C, and $\Delta P$ 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 $2.5^\circ \times 2.5^\circ$ 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, dry corner for RCP4.5 are all relatively close to the 90th percentile values for $\Delta T$ and $\Delta P$, whereas some of the selected models in the cold, wet corner have a considerable distance from the 10th and 90th percentile values for $\Delta T$ and $\Delta P$. 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 $\Delta T$ and $\Delta P$.

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.
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 was 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 GCMs 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.

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

<table>
<thead>
<tr>
<th>GCM runs</th>
<th>Precipitation skill scores</th>
<th>Mean air temperature skill scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LI</td>
<td>UI</td>
</tr>
<tr>
<td>CMCC-CM_r1i1p1</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>CMCC-CMS_r1i1p1</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>CSIRO-Mk3-6-0_r4i1p1</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>CSIRO-Mk3-6-0_r5i1p1</td>
<td>0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>BNU-ESM_r1i1p1</td>
<td>0.32</td>
<td>0.15</td>
</tr>
<tr>
<td>CCSM4_r2i1p1</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>CESM1-BGC_r1i1p1</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>inmcm4_r1i1p1</td>
<td>0.14</td>
<td>0.36</td>
</tr>
<tr>
<td>IPSL-CM5A_LR_r4i1p1</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>CanESM2_r2i1p1</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>CanESM2_r3i1p1</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>bcc-csm1-1_r1i1p1</td>
<td>0.06</td>
<td>0.57</td>
</tr>
<tr>
<td>MRI-CGCM3_r1i1p1</td>
<td>0.04</td>
<td>0.26</td>
</tr>
<tr>
<td>EC-EARTH_r2i1p1</td>
<td>0.01</td>
<td>0.27</td>
</tr>
</tbody>
</table>

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.
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
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 change decreases 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.

<table>
<thead>
<tr>
<th>RCP</th>
<th>Projection</th>
<th>GCM run</th>
<th>Rank per subdomain</th>
<th>Total General rank</th>
<th>Monsoon rank</th>
<th>Final rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LI</td>
<td>UI</td>
<td>LB</td>
<td>UB</td>
<td>LG</td>
</tr>
<tr>
<td>RCP4.5</td>
<td>Warm, dry</td>
<td>CMCC-CM_r1i1p1</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CMCC-CMS_r1i1p1</td>
<td>4</td>
<td>12</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Warm, wet</td>
<td>CSIRO-Mk3-6-0_r4i1p1</td>
<td>14</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CSIRO-Mk3-6-0_r5i1p1</td>
<td>13</td>
<td>7</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Cold, wet</td>
<td>BNU-ESM_r1i1p1</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCSM4_r2i1p1</td>
<td>3</td>
<td>14</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Cold, dry</td>
<td>CESM1-BGC_r1i1p1</td>
<td>2</td>
<td>13</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inmcm4_r1i1p1</td>
<td>9</td>
<td>2</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>RCP8.5</td>
<td>Warm, dry</td>
<td>IPSL-CM5A-LR_r4i1p1</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CMCC-CMS_r1i1p1</td>
<td>4</td>
<td>12</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Warm, wet</td>
<td>CanESM2_r2i1p1</td>
<td>10</td>
<td>6</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CanESM2_r3i1p1</td>
<td>5</td>
<td>11</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Cold, wet</td>
<td>bcc-csm1-1_r1i1p1</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MRI-CGCM3_r3i1p1</td>
<td>12</td>
<td>3</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Cold, dry</td>
<td>EC-EARTH_r2i1p1</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inmcm4_r1i1p1</td>
<td>9</td>
<td>2</td>
<td>14</td>
<td>3</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>RCP</th>
<th>Projection</th>
<th>GCM run</th>
<th>(T) (°C)</th>
<th>(P) (%)</th>
<th>(WSDI) (%)</th>
<th>(CSDI) (%)</th>
<th>(R99P) (%)</th>
<th>(CDD) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCP4.5</td>
<td>Cold, wet</td>
<td>BNU-ESM_r1i1p1</td>
<td>2.5</td>
<td>12.7</td>
<td>729.3</td>
<td>-85.9</td>
<td>140.3</td>
<td>-6.7</td>
</tr>
<tr>
<td></td>
<td>Cold, dry</td>
<td>inmcm4_r1i1p1</td>
<td>1.7</td>
<td>2.7</td>
<td>347.5</td>
<td>-63.4</td>
<td>10.5</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>Warm, cold</td>
<td>CSIRO-Mk3-6-0_r4i1p1</td>
<td>3.5</td>
<td>-2.9</td>
<td>938.2</td>
<td>-98.4</td>
<td>56.9</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>Warm, wet</td>
<td>CSIRO-Mk3-6-0_r5i1p1</td>
<td>3.5</td>
<td>15.6</td>
<td>1530.3</td>
<td>-95.4</td>
<td>109.3</td>
<td>-9.6</td>
</tr>
<tr>
<td>RCP8.5</td>
<td>Cold, dry</td>
<td>inmcm4_r1i1p1</td>
<td>3.6</td>
<td>5.4</td>
<td>905.6</td>
<td>-89.6</td>
<td>30.3</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>Warm, cold</td>
<td>CSIRO-Mk3-6-0_r5i1p1</td>
<td>6.3</td>
<td>-3.1</td>
<td>1855.1</td>
<td>-99.4</td>
<td>140.2</td>
<td>24.1</td>
</tr>
<tr>
<td></td>
<td>Cold, wet</td>
<td>bcc-csm1-1_r1i1p1</td>
<td>4.4</td>
<td>29.7</td>
<td>1215.6</td>
<td>-97.1</td>
<td>220.6</td>
<td>-9.6</td>
</tr>
<tr>
<td></td>
<td>Warm, cold</td>
<td>CanESM2_r3i1p1</td>
<td>6.1</td>
<td>37.4</td>
<td>1426.1</td>
<td>-100.0</td>
<td>295.5</td>
<td>-11.7</td>
</tr>
</tbody>
</table>

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.
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
The 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
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.

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.

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 increases in 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.

Acknowledgements

This work was carried out as part of the Himalayan Adaptation, Water and Resilience (HI-AWARE) consortium under the Collaborative Adaptation Research Initiative in Africa and Asia (CARIAA) with financial support from the UK Government’s Department for International Development and the International Development Research Centre, Ottawa, Canada. This work was also partially supported by core funds of ICIMOD contributed by the governments of Afghanistan, Australia, Austria, Bangladesh, Bhutan, China, India, Myanmar, Nepal, Norway, Pakistan, Switzerland, and the United Kingdom.

We acknowledge the World Climate Research Program’s Working Group on Coupled Modeling, which is responsible for CMIP5, and we thank the climate modelling groups for producing and making available their model output. We thank G.J. van Oldenborgh (Royal Netherlands Meteorological Institute) for making available and assistance with the KNMI Climate Explorer tool. Furthermore we thank J. Sillmann and colleagues for making available the output of their analysis on changes in climate change indices in the CMIP5 multi-model ensemble. We thank two anonymous reviewers for their constructive comments leading to significant improvement of the manuscript. The authors declare that they have no conflict of interest.

The views expressed in this work are those of the creators and are not necessarily attributable to their organizations and do not necessarily represent those of the UK Government’s Department for International Development, the International Development Research Centre, Canada or its Board of Governors.

Supporting Information

The following supporting information is available as part of the online article:
on behalf of the Royal Meteorological Society.

References


© 2016 The Authors. International Journal of Climatology published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.