Climate change and rapidly rising global water demand are expected to place unprecedented pressures on already strained water resource systems. Successfully planning for these future changes requires a sound scientific understanding of the timing, location, and magnitude of climate change impacts on water needs and availability—not only average trends but also interannual variability and quantified uncertainties. In recent years, two types of large-ensemble runs of climate projections have become available: those from groups of more than 20 different climate models and those from repeated runs of several individual models. These provide the basis for novel probabilistic evaluation of both projected climate change and the resulting effects on water resources. Using a broad range of available ensembles, this research explores the spatial and temporal patterns of high confidence as well as uncertainty in projected river runoff, irrigation water requirements, basin storage yield, and cost estimates of adapting regional water systems to maintain historical supply. Projections of river runoff show robust between-ensemble agreement in regional drying (e.g., southern Africa and southern Europe) and wetting trends (e.g., northeastern United States). By integrating runoff over space and time, the economic effects of adapting supply systems to 2050 water availability show still broader trend agreement across ensembles. That agreement, obtained across such a wide range of multiple-member climate model ensembles in some locations, suggests a high degree of confidence in direction of change in water availability and provides clearer signals for longer-term investment decisions in water infrastructure.

1. Introduction

As a result of rising temperatures and changing and more variable precipitation patterns, climate change will significantly affect the patterns of regional and global water availability and demand. Combined with rapidly rising water demand associated with global economic development, these changes will place unprecedented pressures on some already strained water resource systems.1 Successfully planning for future changes that could exceed past variability and hence impact water availability in unprecedented ways requires a scientific understanding of the timing, location, and magnitude of climate change—not only of average trends but also of interannual variability and associated uncertainties. To develop local and regional adaptation responses to water resource challenges that are robust to this wide range of future conditions, it is essential to characterize the extent of these uncertainties (Lempert and Groves 2010).

1 Note that water demand projections depend on the level of development in the country, among other factors, and as a result, water demand will fall in many developed countries in the future. On the other hand, water demand is likely to rise significantly in many less developed countries as population and per capita consumption increase.
In recent years, two types of large-ensemble runs of climate projections have become available: those from groups of more than 20 different general circulation models (GCMs) that have been publicly distributed via the World Climate Research Programme’s Coupled Model Intercomparison Project phase 3 (CMIP3; see Meehl et al. 2007) and phase 5 (CMIP5) and those from repeated runs of several individual models. Examples of the latter include ensembles of 40 and 17 members that are available from the National Center for Atmospheric Research’s Community Climate System Model, version 3 (CCSM3), and those of the Max Planck Institute’s ECHAM climate models (Deser et al. 2012). We henceforth refer to this type of ensemble as within model, as opposed to between-model ensembles that include runs from different climate models. Additional within-model ensembles are becoming available [e.g., using the Fast Met Office/U.K. Universities Simulator (FAMOUS)] or will become available within the next few years, and many other smaller within-model ensembles are available through the recent release of the CMIP5 archive. In this analysis, we compare the above within-model results from the CCSM, ECHAM, and several CMIP5 ensembles, as well as the between-model results from the full set of 23 different CMIP5 models.

These ensembles have provided new approaches to the probabilistic evaluation of both climate change and the resulting effects on water resources and thus improve our understanding of the timing and location of prudent climate adaptation measures. Using a range of ensembles, this research explores the spatial and temporal patterns of uncertainty in projected river runoff, irrigation water requirements, and basin storage yield. Basin storage yield implications are translated to regional and global adaptation cost estimates for each ensemble, and the resulting implications for water management are discussed. While global-scale climate trends among ensembles are relatively robust, local-scale trends are much less so. Precipitation is extremely variable at local scales, while temperatures are considerably less variable (e.g., from one model grid point to another). Irrigation requirements and basin storage yields involve precipitation averaged over the spatial scale of the basin and are also dependent on temperature through evaporation. A central finding of our work is that projections of these water resource variables tend to have higher levels of within- and cross-ensemble agreement than precipitation trends because of such spatial and temporal averaging of precipitation trends and because of ubiquitous warming. Regarding the relationship between river runoff and precipitation, Roderick et al. (2014) find that the bulk of climate-related changes in annual runoff are driven by changes in precipitation rather than evaporation, owing in part to the role of vegetation and soil characteristics (see also Greve et al. 2014); by extension, one would expect to observe similar levels of agreement between precipitation and runoff.

Several recent studies evaluated patterns of uncertainty in precipitation and temperature trends using recent within-model ensembles, and some considered the timing of signal emergence relative to the noise of climate variability. Deser et al. (2012) show strong similarities in the patterns of variability between the 40- and 17-member ensembles of the NCAR and ECHAM models, respectively. From a signal-to-noise perspective, Mahlstein et al. (2012) evaluate the emergence of projected precipitation signals from noise at a more local level and find few grid cells where emergence occurs. In explaining the sources of uncertainty, Hawkins and Sutton (2009) identify three types—model uncertainty, emissions uncertainty, and internal variability—and emphasize that both model uncertainty and internal variability decline relative to the signal as future emissions increase in the latter part of the twenty-first century. Following on this theme, Hawkins and Sutton (2012) showed that many of the apparent differences in future climate change on shorter time scales (next few decades to midcentury) across the different CMIP3 models result from internal variability rather than modeling uncertainty.

Several studies have evaluated the effects of climate change on global water supply and demand. Vörösmarty et al. (2000) use two GCMs to analyze the effect of climate change on global runoff at the 0.5° × 0.5° grid scale and find that the runoff projections developed from these GCMs differ widely. Alcamo et al. (2007) arrived at similar conclusions using the Water—Global Assessment and Prognosis (WaterGAP) model to compute monthly river discharge and worldwide water availability but used only two GCMs under the Special Report on Emissions Scenarios (SRES) A2 and B2 storylines. Arnell (2004) studied the effects of 24 climate scenarios (4 between-model ensembles with 6 models in each) on future runoff in 1300 global basins, finding that regardless of emissions scenario, climate change is projected to increase water stress globally. Milly et al.’s (2005) article considered the outputs of 12 GCMs in an analysis of how climate change will affect runoff in 163 river basins, and although the authors did not consider within-model ensembles, they do find many regions of between-model sign agreement in runoff projections. More recently, Strzepek et al. (2013) evaluated the effects of climate change under 56 different model–SRES scenario combination runs on a set of six hydrological indicators and found that model uncertainties in river
runoff and irrigation water demand tended to be higher in developing than in developed countries. Using the same set of 56 runs in a study of projected U.S. drought patterns, Strzepek et al. (2010) find that measures of drought that incorporate temperature rather than focusing on precipitation only (i.e., Palmer drought severity index vs standardized precipitation index) produce much greater between-model agreement because of the cross-model agreement in temperature trends. Konzmann et al. (2013) also provided a detailed investigation of the impact of climate change on global irrigation water requirements for 19 GCMs, finding strong between-model agreement that requirements will increase in most regions except for southern Asia, where higher precipitation is projected. This result aligns with the conclusions of the current work, which consider a broader range of ensembles.

Research has been conducted on the economic effects of climate change on water resource outcomes, but typically this research has not used the broad range of climate models. Ward et al. (2010) investigated the potential costs of maintaining reservoir supply yield globally but evaluated outcomes using only two climate models. In this paper, we repeat components of this analysis and demonstrate that using different model runs, even from the within-model ensemble used for one of the two scenarios, could generate much different outcomes. The Ward et al. (2010) study emerged from a broader World Bank program called the Economics of Adaptation to Climate Change (EACC), which estimated the costs of adapting to climate change in developing countries at $100 billion yr$^{-1}$ in 2050 (Margulis and Narain 2009). In more recent work, the costs of flooding, droughts, and water quality impairments (respectively, Strzepek et al. 2015; Boelhert et al. 2015a,b) have been estimated using a set of model outputs from the NCAR Community Atmosphere Model, version 3.0 (CAM3; Collins et al. 2004), and the Massachusetts Institute of Technology Integrated Global System Model (IGSM; Monier et al. 2013). However, these studies only evaluated economic effects using two climate models.

In the following sections, we document the methodologies used to investigate patterns of uncertainty across model ensembles and then provide results and conclusions of our analysis.

2. Methods

To translate a suite of climate model outputs into projected effects on water availability and demand, a wide range of data and modeling approaches is required. Broadly, projected monthly temperature and precipitation from numerous climate models were routed through a linked series of models (see diagram in supplemental material Fig. S1) to estimate potential evapotranspiration (PET), irrigation water requirements, river runoff, and basin storage yield. In the remainder of this section, we describe these procedures in greater detail: 1) characterizing the baseline conditions for the analysis, including datasets and issues of scale and resolution; 2) processing of climate model ensembles; 3) runoff modeling; 4) modeling irrigation water requirements; and 5) basin yield modeling.

a. Characterizing baseline conditions

This study employs two river basin resolutions that strike a balance between precision and accuracy and are appropriate for the respective analyses in which they are used. The first is 8951 river basins of the world, which were developed by Strzepek et al. (2011) using the HYDRO1k dataset from the U.S. Geological Survey (USGS) for geographic delineation of basin boundaries. The HYDRO1k dataset is currently the best available for global river basin delineation (Strzepek et al. 2011). The basins in the raw HYDRO1k dataset range significantly in size, from the smallest catchments of less than 1 km$^2$ to drainage areas for rivers such as the Nile or Amazon that are well over the typical grid scale of a GCM (i.e., 2.5° × 2.5°). The 8951 basins were selected to be no smaller than the resolution of available baseline climate data (0.5° × 0.5°) and thus range in size from approximately 2500 km$^2$ (which is similar to a baseline data grid cell of 0.5° × 0.5°) to more than 62 500 km$^2$ (which is similar to a climate model grid cell of 2.5° × 2.5°). For the basin yield analysis, the river runoff projected for these basins is aggregated up to 126 major river basins of the world used in the International Food Policy Research Institute (IFPRI) International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT) model (Rosegrant et al. 2012). Data on reservoir storage are available for these larger basins.

Historical data are needed in this study for two reasons: 1) to model historical PET, runoff, and irrigation water requirements that serve as a basis for calculating changes in those variables and 2) to develop bias-corrected GCM projections so that they are consistent with observed data. Baseline precipitation, temperature, and daily average temperature range data for the 1961–90 period were from the University of East Anglia’s Climatic Research Unit (CRU) Time Series (TS) version 2.1 dataset (Mitchell and Jones 2005). These three variables are needed for estimation of PET using the modified Hargreaves formulation and for runoff using the Climate and Runoff Model II (CLIRUN-II). The CRU TS datasets are the standard reference baselines for the World Meteorological Organization and
provide a monthly time series of these variables on a 0.5° × 0.5° grid (see Fig. S2 in the supplemental material) for mean baseline temperature and precipitation data. To calculate PET, we use the modified Hargreaves approach (Allen et al. 1998; Droogers and Allen 2002), which the Food and Agriculture Organization of the United Nations (FAO) recommends when information is limited on wind speed and other key inputs as opposed to the more data-intensive Penman–Monteith method. A more thorough description of the PET calculation, along with the modified Hargreaves equation, is provided in the supplemental material.

Natural runoff data used to calibrate the simulated runoff outputs rely on two sources. The first is a global gridded dataset of historical average monthly runoff from the Global Runoff Data Centre (GRDC; Global Runoff Data Centre 2007; Fekete et al. 2002). The dataset provides 12 monthly mean values for each grid cell and is currently the best globally available source of terrestrial runoff data. The second source of runoff data is from the IFPRI and provides a time series of 1950–2000 monthly runoff data for the 282 IFPRI food-producing regions (FPUs) of the world, which are intersections between the 126 major global river basins and country boundaries. The resulting global spatial patterns of runoff depth in these two datasets map closely to each other (see Fig. S3 in the supplemental material). Other datasets are being developed (e.g., McMahon et al. 2007) but are not yet available for use in a global runoff study.

b. Climate ensembles: Description and processing approach

The central focus of this study is on the patterns and sources of uncertainty in a broad set of within- and between-model ensembles as shown in Table 1. These 220 GCM runs, which incorporate several greenhouse gas emissions scenarios and dozens of modeling frameworks, contain 5 between-model ensembles and 12 within-model ensembles and thus reflect a wide variability in possible spatial and temporal distribution of precipitation and temperature outcomes. The five emissions scenarios include the B1, A1B, and A2 storylines from the Special Report on Emissions Scenarios (IPCC 2000) employed in the 2007 Fourth Assessment Report (IPCC 2007), as well as two representative concentration pathway (RCP; van Vuuren et al. 2011) scenarios at stabilization levels of 4.5 and 8.5 W m⁻² of forcing employed in the Fifth Assessment Report (IPCC 2013). The individual models and modeling groups used from the CMIP3 and CMIP5 archives are provided in the supplemental material (Tables S1 and S2). The majority of analysis is conducted on three of the larger within-model ensembles (40-member CCSM3, 17-member ECHAM, and 10-member CSIRO RCP4.5) and three of the between-model ensembles (22-member A1B, 23-member RCP4.5, and 20-member RCP8.5). This set of within-model ensembles is selected to ensure that an adequate number of members is available to allow for statistical comparisons, and the between-model ensembles are selected to provide a linkage between the emissions scenario utilized in two of the large within-model ensembles (SRES A1B) and the outcomes projected in the more recent set of CMIP5 models.

We controlled for differing emissions and climate sensitivity by normalizing each GCM run to the mean global temperature and precipitation trajectory of the 22-member SRES A1B CMIP3 ensemble (see supplemental

---

**Table 1. Characteristics of climate model ensembles relied upon in this study.** (Expansions of acronyms are available at [http://www.ametsoc.org/PubsAcronymList](http://www.ametsoc.org/PubsAcronymList).)

<table>
<thead>
<tr>
<th>Model series</th>
<th>Ensemble name and/or emissions scenario</th>
<th>Ensemble type</th>
<th>Members</th>
<th>Time period available</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMIP3</td>
<td>SRES B1</td>
<td>Between</td>
<td>17</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>SRES A1B</td>
<td>Between</td>
<td>22</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>SRES A2</td>
<td>Between</td>
<td>17</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>NCAR CCSM A1B</td>
<td>Within</td>
<td>40</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>NCAR CCSM A1B</td>
<td>Within</td>
<td>7</td>
<td>1900–2061</td>
</tr>
<tr>
<td></td>
<td>MPI ECHAM A1B</td>
<td>Within</td>
<td>17</td>
<td>1950–2099</td>
</tr>
<tr>
<td></td>
<td>MPI ECHAM A1B</td>
<td>Within</td>
<td>4</td>
<td>1900–2099</td>
</tr>
<tr>
<td>CMIP5</td>
<td>RCP4.5</td>
<td>Between</td>
<td>23</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>RCP8.5</td>
<td>Between</td>
<td>20</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>CCSM RCP4.5</td>
<td>Within</td>
<td>6</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>CCSM RCP8.5</td>
<td>Within</td>
<td>6</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>CSIRO RCP4.5</td>
<td>Within</td>
<td>10</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>CSIRO RCP8.5</td>
<td>Within</td>
<td>10</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>CAN RCP4.5</td>
<td>Within</td>
<td>5</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>CAN RCP8.5</td>
<td>Within</td>
<td>5</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>GISS-E2-R RCP4.5</td>
<td>Within</td>
<td>6</td>
<td>1900–2099</td>
</tr>
<tr>
<td></td>
<td>GISS-E2-H RCP4.5</td>
<td>Within</td>
<td>5</td>
<td>1900–2099</td>
</tr>
</tbody>
</table>
material for more detail). The approach is similar to that applied by Hawkins and Sutton (2009) and allows us to identify structural differences between models rather than emissions uncertainties. Changes projected by the models are assessed between a historical baseline period (1961–90) and two “eras”: 2040–59 and 2080–99 (referred to as the 2050 and 2090 eras, respectively). The first era is a relevant time scale for current water infrastructure planning, and the second provides a means to evaluate later signal emergence in many regions where emergence does not occur by midcentury.

Because the spatial resolution of the 8951 basins is generally finer than that of GCMs, it was necessary to match the lower-resolution GCM output with the higher-resolution basin scale. Both statistical and dynamical downscaling have well-studied uncertainties (Kerr 2011), and the only practical approach for a global analysis involving a large number of climate runs is to use projected changes in temperature and precipitation for the GCMs at their native resolutions. The changes in each variable were generated using the delta method, which subtracts the mean monthly modeled baseline from the projected values to produce delta temperature and precipitation. These deltas are then mapped directly onto the 0.5° × 0.5° resolution of the baseline climate data. This approach captures the range of potential climate change impacts at a higher resolution without downscaling the GCMs themselves. We then aggregated these gridded data to the basin scale using GIS software.

c. Runoff modeling

This study employs CLIRUN-II to model changes in runoff under each of the GCM runs. CLIRUN-II (Strzepek and McCluskey 2010; Strzepek et al. 2011) is a one-dimensional infiltration and runoff estimation tool that uses historic runoff as a means to estimate soil characteristics. It is the latest model in a family of hydrologic models developed for the analysis of climate change impacts on runoff. Kaczmarek (1993) presented the theoretical development for CLIRUN—a single-layer, lumped, watershed rainfall runoff model—that he applied to the Yellow River in China (Kaczmarek 1998). A snow-balance model and a suite of PET models were added and the model was renamed WatBal (Yates 1996), which has subsequently been used on a wide variety of spatial scales from small and large watersheds to global (e.g., Huber-Lee et al. 2005; Strzepek et al. 2005). CLIRUN-II builds on the CLIRUN and WatBal frameworks by addressing the issue of modeling extreme events at the monthly and annual level. CLIRUN-II follows the framework of the six-parameter (SIXPAR) hydrologic model (Gupta and Sorooshian 1983, 1985) by adopting a two-layer approach and employs unique conditional parameter estimation procedures. The structure of the model is described in detail in the supplemental material.

CLIRUN-II simulates natural runoff with six calibration parameters and requires natural runoff data to calibrate the model over a historic period. We calibrated the model by minimizing the squared differences between the 12 monthly GRDC runoff values and the 12 monthly averaged CLIRUN-II model outputs from the 1971–80 simulation period; this period was selected to best represent the source period for the 12 months of GRDC runoff data. Note that there are several limitations of using the GRDC dataset for calibration: 1) the dataset provides monthly averages and so yields no information on extremes; 2) gridded data in dry areas with no gauged data are unreliable; and 3) the period of station data availability varies, so there may be temporal inconsistencies in the gridded data. As a result of these issues, calibration performance is closely tied to availability of runoff data (see Fig. S7 in the supplemental material). The choice of CLIRUN-II also introduces uncertainties; prior research suggests that there can be large differences between the results of alternative hydrological models (e.g., Haddeland et al. 2011; Schewe et al. 2014).

d. IWR modeling

Globally, irrigation water requirements (IWR) are the largest consumptive use of water and will be
Fig. 2. Percentage change in ensemble-mean annual precipitation for (top) 23-member CMIP5 RCP4.5 and (bottom) 20-member CMIP5 RCP8.5 from the 1961 to 1990 baseline and the 2010s through the 2090s. In regions shaded white, fewer than two-thirds of the model members agree on the direction of change. Regions shaded gray have at least two-thirds of models agreeing on a small change of between −5% and +5%.
significantly affected by projected rising temperatures and changing and more-variable precipitation. Because of the strong dependence of crop water demands on temperature, climate change will have a more uni-directional (i.e., increasing) effect on crop water demand than on runoff. As a result, we would expect the IWR signal to emerge from the noise sooner in this variable than in water resource variables that depend more directly upon precipitation.

Although detailed crop modeling of irrigation water requirements is far beyond the scope of this work, simplified methods can provide an understanding of water requirements suitable for a global-scale analysis. One such method was developed by FAO (Allen et al. 1998) and employed by IFPRI (Rosegrant et al. 2002) and relies on data that are available at a global scale. These include soil characteristics, temperature, precipitation, and PET, as well as crop-specific information including planting and harvest dates and seasonal timing of water demands. Soil characteristics and timing of water demands are available through FAO (Allen et al. 1998), and a database of global planting and harvest dates for a wide range of crops is available through the University of Wisconsin (Sacks et al. 2010).

In this analysis, we focus on maize and wheat, which are the two largest global crops by growing area and cover 18% and 15% of total global cropland in 2013, respectively (Food and Agriculture Organization 2013). Crop coverage data are provided through the harvested areas and yields dataset, which is available at a $\frac{1}{2}^\circ \times \frac{1}{2}^\circ$ resolution through the University of Wisconsin [described in Monfreda et al. (2008); dataset accessed on 15 Oct 2013 from https://nelson.wisc.edu/sage/data-and-models/datasets.php]. Because this crop coverage dataset also includes rain-fed areas, the area was reduced to irrigated regions only using an FAO dataset of global gridded irrigation data [described in van Velthuizen et al. (2007)]. These data are spatially aggregated up to the level of the 8951 global river basins. Overall, the process outlined by Allen et al. (1998) and Rosegrant et al. (2002) for calculating IWR for a particular crop involves first estimating total monthly crop
water demand [crop evapotranspiration (ETc)], then estimating available monthly supply [effective precipitation (Pe)], and then calculating IWR each month as the difference between these values (ETc \(- Pe\)). These steps are described in more detail in the supplemental material.

**e. Basin yield and adaptation cost estimation**

Basin yield is a measure of the annually reliable water supply from a basin. Much of the water available in a basin during a given year is lost if it is not stored, so storage in a basin can greatly increase its reliable supply or yield. Basin yield is a useful broad indicator of the climate risk to basin-level water resources because it indicates a basin’s ability to absorb the impact of changes in both the mean and variability of flows under climate change. Furthermore, achieving specified basin yield targets has economic implications, allowing the costs of adapting to climate changes to be estimated.

The storage yield curve has been developed as a means of relating basin yield (vertical axis) to basin storage (horizontal axis) and is a measure of the volume of storage needed to achieve a given level of reliable yield. The maximum yield on the curve indicates the mean annual runoff in the basin, and the minimum yield indicates the lowest flow in the runoff time series for the basin in question. In other words, in a basin with no managed storage, the basin yield is assumed to be the lowest recorded annual flow. The shape of the storage yield curve is determined by the historical variability of basin runoff, where a steeper curve indicates a more stable system and a flatter curve a more variable one. So all else being equal, a basin with more variable flows would require more storage to achieve the same level of reliable yield. To construct storage yield curves, we use the sequent peak algorithm (Wiberg and Strzepek 2005, modified from Thomas and Fiering 1962), which is an iterative procedure that identifies the minimum storage volume needed to generate various levels of reliable yield, given a basin inflow time series. In this formulation, the elements include reservoir storage, releases, evaporation and precipitation over the reservoir, and inflow, all at a monthly time step. Climate change can affect all elements of a storage yield curve, including the maximum yield given by the mean annual runoff, the minimum yield assumed to be the lowest annual flow, and the curve’s slope given by the variability of inflows. As a result, basin yield serves as an integrator of the various potential effects of climate change over both time and space. Importantly, climate change will also affect yield reliability (at least

---

3 Supplemental Fig. S11 provides an example of a storage yield curve for the Nile River at Aswan in Egypt and illustrates the information that the curve provides.

4 The current analysis does not include evaporation and precipitation on the surfaces of reservoirs.
via warming and evaporation in any basin, as well as through other climate changes in many), which can be estimated using long-established synthetic flow time series generation techniques that provide confidence intervals on the storage yield curve (Thomas and Fiering 1962). There are two ways that storage yield curves are useful, especially for decision-makers. The first focuses on the impacts of climate change and evaluates the change in yield given fixed basin storage. The second view focuses on adaptation and considers the change in storage that would be required to maintain a fixed yield (see Fig. 1). If yields increase, climate change may provide economic benefits but only if the basin has existing water deficits or growing demands or if markets exist to trade water between basins.

Using baseline and projected annual runoff and low-flow values, baseline and projected storage yield curves were created for each of the 126 basins, for the baseline and across all 220 GCM runs, and for both the 2050 and 2090 eras. Combined with information on reservoir storage in each of the 126 basins from IFPRI (presented in Fig. S12 in the supplemental material), these storage yield curves provide information on changes in basin yields given existing basin storage and, in cases where yields fall, the required increases in storage needed to maintain existing yields. Note that in basins where the ratio of existing basin storage to current mean annual runoff exceeds one, additional storage will provide no additional basin yields. In these basins, falling yields would require alternative, nonstorage adaptation options.

We next estimate the economic impacts of maintaining current basin yields under a changing climate. In basins where yields decline, the adaptation response is assumed to be storage or an alternative to storage (e.g., desalinization, groundwater development, or interbasin transfer) that costs $1 m$^3$ of lost yield, whichever is less expensive. Costs of storage are taken from the International Water Management Institute (Keller et al. 2000) and Wiberg and Strzepek (2005).

![Figure 5. Percent change in precipitation between the 1961–90 baseline and the 2050 era. Box plots are averages of all grid cells in 4° latitude bands of the earth, with the box plots being defined based on ensemble members.](image-url)
which estimates volume–cost relationships for reservoirs. The price of the storage alternative is adopted from Ward et al. (2010), who take a similar approach to the current study in estimating the global costs of maintaining basin yields under climate change.

On the other hand, if basin yield increases under climate change, we assume that surplus water provides economic benefits only if the basin is water stressed. Operationally, if water is below a relative water stress of one on United Nations Educational, Scientific, and Cultural Organization’s (UNESCO) World Water Assessment Programme (2006) water stress index, that basin is assigned no value for surplus water. Other basins are assigned values between $0 and $1 m$^{-3}$, scaled to the basin’s water stress index level (calculated values in each basin are presented in Fig. S13 in the supplemental material).

3. Results: Projected changes

The overall purpose of this section is to analyze spatial patterns of agreement and disagreement between water availability and demand metrics and derive possible
lessons for water management and climate change adaptation. We first present levels of agreement among ensemble patterns of changes in precipitation, which is the primary driver of global patterns of runoff and thus storage yield. These precipitation trend patterns are then compared to patterns of agreement in runoff, IWR, and basin yield trends, as well as the economic implications of maintaining historical yields. As already noted, these variables progressively include to greater degrees the influence of warming as well as precipitation and integrate the effects of climate variability and change over both time and space so that they provide a broader perspective on patterns of agreement that are key for water resources among and across ensembles.

a. Precipitation trend patterns

1) AGREEMENT PATTERNS OVER TIME AND EMISSIONS SCENARIOS

Figure 2 shows that agreement patterns in precipitation trends are spatially consistent over time and with emissions scenarios within each ensemble. As time progresses and emissions rise, the patterns of both agreement and persistent disagreement remain relatively fixed within models, thus delineating where the human-induced signals emerge as emissions increase and where they do not. With regard to emissions, note that the adjustment to the global mean precipitation trend means that the global trend in each decade is consistent across all ensembles. As such, the deeper colors on the RCP8.5 figure (Fig. 2, right) reflect intensification of trends in both directions rather than higher or lower precipitation globally.

The regions of persistent disagreement indicate areas of the globe with meteorological patterns that are difficult to quantify, perhaps because of high levels of natural internal variability or perhaps because of model errors. In contrast, the consistency of agreement patterns suggests emergent behavior characteristic of each model. Not unexpectedly, within-model ensembles have a much greater degree of agreement in direction of change among members (Fig. 3). The 40-member CCSM3 (40NCAR in Fig. 3) and 17-member ECHAM ensembles both show very few regions on the globe where fewer than two-thirds of model runs agree on sign change.

2) AGREEMENT PATTERNS ACROSS MULTIPLE WITHIN-MODEL ENSEMBLES

In certain regions of the globe, agreement remains strong across two or more within-model ensembles regardless of emissions scenario and CMIP series. We average the level of agreement, measured as the fraction of ensemble members that show a positive change in sign, across five ensembles from available sets of CMIP5 RCP4.5 groupings, which is the broadest set of within-model ensembles available in our dataset (Fig. 4). Interestingly, we find that agreement patterns in certain regions are maintained even when patterns from these five within-model ensembles are combined in this way. Consistently drying regions include northern Africa and southern Europe, southern Africa, the southwestern United States and Central America, and Indonesia. Wetting regions include southeastern South America, the northeastern United States, eastern and southern Asia, and the northern latitudes. This suggests that less model uncertainty exists in these regions than in other areas displaying less coherence across runs and models, such as Indonesia or central Africa. Notably, most of the
Fig. 7. Median ensemble SNR of changes in (top) annual runoff and (bottom) dry season runoff across the 126 basins from the 1961–90 baseline to the 2050 era. The signal is the average change between periods, and the noise is the interannual standard deviation of the 2050 era. The dry season is defined as the lowest three months of the baseline period each year for each basin.
land surface areas of large disagreement appear to be in midcontinental (i.e., rather than coastal) regions of North and South America, Australia, Africa, and the Middle East.

3) PATTERNS OF AGREEMENT IN LATITUDE BANDS

Last, patterns of changes in precipitation by latitude band are very similar across both inter- and intra-GCM ensembles, suggesting that the bulk of disagreement between models results from longitudinal differences (Fig. 5).

b. Runoff trend patterns

We next compare the level of agreement between modeled runoff projections at the 8951 and 126 basins to levels of agreement within precipitation projections at the same scales and then evaluate signal-to-noise ratios (SNRs) to identify emergence.

1) AGREEMENT ACROSS ENSEMBLES RELATIVE TO PRECIPITATION

Generally, runoff changes exhibit a spatially similar pattern of agreement to precipitation trends but with a more robust agreement on drying in certain regions such as eastern Australia and northeastern South America (Fig. 6). As these levels of agreement are difficult to clearly identify visually in maps, Table 2 compares the percentage of sign agreement within selected river basins, with wetting or drying agreement of greater than 80% shaded in blue and red. Certain basins, such as the northeastern United States or eastern Mediterranean Sea, have universal agreement across the available models, suggesting robust trends in those regions. The most apparent difference between precipitation and runoff change, however, is a general increase in drying. In the Orinoco (South America), Niger (western Africa), and Yellow (China) Rivers, as well as the Amu Dar’ya (central Asia), this drying moves individual ensembles into the 80% agreement on drying category when precipitation is translated to runoff (basin locations are provided in Fig. S14 in the supplemental material).

2) SNR IN RUNOFF VERSUS PRECIPITATION

The signal-to-noise ratio provides information on when a climate signal is statistically significant and can provide valuable information on trigger points for adaptation. For example, the adaptation strategy of flexible design (De Neufville and Scholtes 2011) involves designing infrastructure systems so that future adjustments can be made once more information becomes available (e.g., building additional hydropower turbine bays and then adding the turbine only if a wet future occurs). As a result, the SNR can provide synthesized information about agreement on trends within an ensemble of model runs. We process the SNR following the approach employed by Deser et al. (2012), who define the signal as the change from a baseline to a given 30-yr projected period and the noise as the interannual standard deviation over those 30 years. In our case, the signal would be the difference between the mean 1961–90 value and the 2050 era (2040–59) value, and the noise would be the interannual standard deviation during the 2050 era. We then report the median SNR over the ensemble for the 126 basins (Fig. 7) and a selection of those basins (Table 3).

By the 2050 era, a drying signal emerges in some arid regions during dry season (lowest three months of runoff

<table>
<thead>
<tr>
<th>Basin</th>
<th>40NCAR</th>
<th>1ECHAM</th>
<th>10R4CSIRO</th>
<th>40NCAR</th>
<th>1ECHAM</th>
<th>10R4CSIRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zambezi</td>
<td>0.03</td>
<td>−0.45</td>
<td>−0.64</td>
<td>−0.06</td>
<td>−0.39</td>
<td>−0.65</td>
</tr>
<tr>
<td>Congo River</td>
<td>1.16</td>
<td>0.34</td>
<td>−1.35</td>
<td>0.44</td>
<td>−0.07</td>
<td>−1.84</td>
</tr>
<tr>
<td>Niger River</td>
<td>0.44</td>
<td>0.41</td>
<td>−0.80</td>
<td>−0.46</td>
<td>0.38</td>
<td>−1.51</td>
</tr>
<tr>
<td>Nile</td>
<td>1.32</td>
<td>0.33</td>
<td>−0.59</td>
<td>0.41</td>
<td>−0.03</td>
<td>−0.33</td>
</tr>
<tr>
<td>Orange River</td>
<td>0.56</td>
<td>−0.27</td>
<td>−0.51</td>
<td>0.09</td>
<td>−0.30</td>
<td>−0.55</td>
</tr>
<tr>
<td>Orinoco</td>
<td>0.77</td>
<td>−0.09</td>
<td>−0.29</td>
<td>0.47</td>
<td>−0.32</td>
<td>−0.66</td>
</tr>
<tr>
<td>Amu Dar’ya</td>
<td>−0.33</td>
<td>−0.12</td>
<td>0.08</td>
<td>−0.33</td>
<td>−0.32</td>
<td>−0.16</td>
</tr>
<tr>
<td>Syr Dar’ya</td>
<td>−0.35</td>
<td>0.19</td>
<td>0.08</td>
<td>−0.29</td>
<td>0.06</td>
<td>−0.05</td>
</tr>
<tr>
<td>Huang Hé (Yellow River)</td>
<td>1.48</td>
<td>0.15</td>
<td>0.16</td>
<td>1.39</td>
<td>−0.34</td>
<td>−0.25</td>
</tr>
<tr>
<td>Black Sea</td>
<td>−0.34</td>
<td>−0.01</td>
<td>−0.10</td>
<td>−0.24</td>
<td>0.35</td>
<td>0.14</td>
</tr>
<tr>
<td>Colorado River</td>
<td>−0.69</td>
<td>−0.06</td>
<td>−0.33</td>
<td>−0.60</td>
<td>−0.09</td>
<td>−0.41</td>
</tr>
<tr>
<td>U.S. Northeast</td>
<td>0.84</td>
<td>0.53</td>
<td>0.59</td>
<td>0.65</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>Eastern Mediterranean Sea</td>
<td>−0.72</td>
<td>−0.64</td>
<td>−0.51</td>
<td>−0.90</td>
<td>−0.69</td>
<td>−0.47</td>
</tr>
<tr>
<td>North African countries</td>
<td>−0.26</td>
<td>−0.47</td>
<td>−0.48</td>
<td>−0.56</td>
<td>−0.82</td>
<td>−0.89</td>
</tr>
</tbody>
</table>

TABLE 3. Median ensemble SNR in annual precipitation and runoff trends for selected basins, where the signal is the trend from the 1961–90 period to the 2050 era, and the noise is the interannual standard deviation of the 2050 era. Cells in italics show SNRs of 0.5 to 1 or between −0.5 and −1, and values in boldface show greater than 1 or less than −1. Basin locations are provided in Fig. S14 in the supplemental material.
for each basin; Fig. 7, right), and a wetting annual signal emerges in others (Fig. 7, left), although the effect is still modest. This absence of apparent emergence in runoff change is similar to the findings of Mahlstein et al. (2012) for annual precipitation trends within the CMIP3 A1B ensemble.\(^5\) As with agreement, comparing precipitation and runoff SNRs across individual basins suggests that runoff tends to move signals to a direction of more drying (Table 3). Some emergent wetting SNRs in precipitation are only mild SNRs in runoff, and many of the drying ratios intensify.\(^6\)

\(^5\) Importantly, note that the high arid basin SNRs during the dry season are most likely attributable to near-zero runoff during those months.

\(^6\) Looking at monthly trends in precipitation and runoff over the 22 CMIP3 A1B runs reveals clearer signs of month-to-month emergence (i.e., whiskers do not overlap zero line) in runoff than in precipitation (see Fig. S15 in the supplemental material).

c. IWR

Irrigation water requirements are the largest global water use and therefore a strong indicator of the effect that climate change will have on global water demand. As a result of the strong dependence on temperature, regional trends are generally positive, and agreement across both between- and within-model ensembles is considerably broader than with precipitation or runoff trends (Fig. 8). Because IWR is the difference between crop water demands and usable rainfall, the strongest positive trends are in regions where rainfall declines are largest. Although one might expect to see clear emergence of the signal from the noise based on these trends, the interannual variation of IWR is driven by both available energy (driven in part by temperature) and precipitation, and this “double noise” causes noise to generally overwhelm the signal, with some notable exceptions in the Middle East and western North America (Fig. 9).
d. Basin storage yield

Sustainable water supply yield from a river basin is influenced by mean annual runoff, interannual flow variability, and available storage infrastructure. Changes in minimum flows, variability, or mean conditions under climate change would affect basin yield. Figure 10 provides an example—for a set of four between-model ensembles for the Zambezi basin in southern Africa—of how climate change is projected to affect storage yield in a basin. Because of its shallow slope, small vertical changes in the storage yield curve can cause large increases in storage requirements to maintain fixed yields.

As described in the methodologies section and suggested in Fig. 10, in basins with falling yields, either additional storage or another source (e.g., desalination) will be needed to meet demands. On the other hand, increases in storage yield may present opportunities for internal basin development or interbasin transfers. Changes in yields closely mirror changes in runoff (Fig. 11, left), whereas resulting changes in storage requirements can be magnified considerably owing to the nonlinear relationship between yield and storage (Fig. 11, right).

As described in the methodologies section, we estimate the economic effects of climate change by assuming that basins with decreasing yields incur costs to maintain historical yields and that basins with increasing yields gain as economic benefits if water is scarce. Costs are cheaper with either increased storage or a backstop of $1 m$^3$ of lost yield, whichever is less expensive. We find that the median annual global net costs of adapting to climate change in the 2050 era is $15 billion yr$^{-1}$ at a 5% discount rate [2010 U.S. dollars (USD); Fig. 12]; this estimate is higher than that of Ward et al. (2010), although that study focused on developing countries only. However, the interquartile ranges of the CMIP3 and CMIP5 ensembles range from $-3$ billion (RCP4.5) to $+35$ billion (A2), and the range of intra-GCM ensembles is from $-10$ billion (NCAR A1B) to $+60$ billion (CSIRO RCP4.5). A lesson to draw from these findings is that it is critically

---

Fig. 9. Median IWR SNR across ensemble members for nine ensembles. The signal is the mean change between 1961–90 and the 2050 era, and the noise is the interannual standard deviation over years in the 2050 era.
important to consider a broad range of climate models when doing adaptation planning.\(^7\)

To evaluate agreement among adaptation cost estimates, it is more relevant to focus on geographic regions rather than the globe. We aggregate the basin-scale costs to the seven World Bank regions, which include East Asia and the Pacific (EAP), Europe and central Asia (ECA), Latin America and the Caribbean (LAC), the Middle East and North Africa (MENA), South Asia (SA), sub-Saharan Africa (SSA), and all other countries (NB). We see an unexpected degree of agreement in direction of economic outcome across these regions (Fig. 13). In the MENA and SSA regions, the interquartile ranges (IQR) of the three between-model and three within-model ensembles all show positive costs. There is a general agreement on benefits in SA and on costs in LAC.

The explanation for the much stronger sign agreement in economic effects has two sources: 1) economic outcomes map to runoff, which includes the stronger sign agreement in temperature trends; and 2) in many basins, changes in precipitation have a magnified effect on runoff. These dynamics can be seen for a selection of basins in Fig. 14, which plots percentage changes in precipitation and runoff versus annual costs for all 220 model runs. The fact that the fitted precipitation line is generally above the runoff line indicates that zero changes in precipitation will still result in costs as a result of the temperature effect. In basins such as the Zambezi, Volta, and Niger Rivers, the precipitation line is steeper than the runoff line such that marginal changes in precipitation have a magnified effect on economic outcomes, leading to greater sign agreement.

\(^{7}\)For example, the World Bank EACC study estimates costs of adaptation in developing countries to be \(\sim \$100\) billion yr\(^{-1}\). The water supply component of this was approximately \$9 billion yr\(^{-1}\). They used two scenarios—NCAR A2 run 1 (dry) and CSIRO A2 run 1 (wet)—to develop these estimates, which result in a range from \(-\$20\) billion to \(+\$95\) billion based on our run 1 of these two ensembles.
4. Conclusions

Climate change and rapidly rising global water demand are expected to place unprecedented pressures on already strained water resource systems. Successfully planning for these future changes requires a sound scientific understanding of the timing, location, and magnitude of climate change impacts on water needs and availability—not only average trends but also interannual variability and associated uncertainties. Here we have explicitly shown the limitations of using single models or even single ensembles for such planning and

![Maps showing median percent change in storage yield across select ensembles.](image-url)
have examined how a broader range of ensembles better informs evaluation of features that are robust to internal variability as well as to multiple model formalisms. Using a range of available within- and between-model ensembles, we have explored the spatial and temporal patterns of high confidence as well as uncertainty in projected river runoff, irrigation water requirements, and basin storage yield. Precipitation is the main

Fig. 12. (top) Median cost across select ensembles (in billions of 2010 USD, discounted at 5%) and (bottom) box plots of global costs for each ensemble (in billions of 2010 USD, discounted at 5%).
climatic driver of uncertainty in projections of water demand and availability, but several water resource variables also depend upon temperature. A central finding of our work is that projections of these water resource variables tend to have higher levels of within- and cross-ensemble agreement than precipitation.

Between-model patterns of high confidence and uncertainty in precipitation trends tend to remain fixed over time and with increased forcing (Fig. 2), suggesting emergent behavior in each model that has been observed in prior research (e.g., Strzepek and Schlosser 2010). Projected changes in precipitation and temperature drive modeled changes in river runoff, and we observe strong spatial patterns of multiple-ensemble agreement and disagreement in both precipitation and runoff trends (Fig. 6). Regions with robust cross-ensemble drying trends include southern Europe, northern Africa, western Australia, southern Africa, eastern Brazil, and northern Mexico; wetting trends occur in the northeastern United States, Canada, northern regions of the globe, and parts of Southeast Asia.

Basin yield has the advantage of integrating changes in both the mean and variability of projected runoff over time. We find that relative to changes in precipitation, patterns of changes in basin yield are both magnified and systematically drier as a result of the dependence on land surface dynamics and temperature. Because of the temporally integrating effects of basin yield and monetary discounting, the costs of maintaining historical yields show still stronger patterns of agreement across GCM ensembles, particularly when focusing on agreement within broad geographic regions (Fig. 13). If the robust patterns of projected increases in irrigation water requirements (Fig. 8) were incorporated into basin supply needs to more fully capture the effects of climate change, the model agreement observed here would be broader still. The fact that agreement is obtained across such a broad range of multiple-member GCM ensembles in some locations suggests a high degree of confidence in direction of change in water availability in these regions and provides clearer signals for longer-term investment decisions in water infrastructure. For example, river runoff projections show robust between-ensemble agreement in regional drying (e.g., southern Africa, southern Europe, and northern South America) and wetting trends (e.g., northeastern United States and northern latitudes). Aggregating results to seven global regions, the impacts of adapting basin-scale supply systems to 2050 water availability show remarkable trend agreement across ensembles, with costs occurring in sub-Saharan Africa, Latin America and the Caribbean, and the Middle East and northern Africa and benefits occurring in central and South Asia. The variability in estimated adaptation costs across within-model ensembles illustrates the importance of considering a broad range of climate models...
when doing adaptation planning. Our results illustrate how reliance on the outputs of too few models can generate highly misleading results in regions where consideration of multiple models shows that the direction and magnitude of projected changes are highly uncertain.

Among important directions for future work is an assessment of whether some ensembles are more appropriate for certain regions based, for example, on their statistical performance relative to the observed climate of that region (i.e., the skill of the ensemble). Although our screening assessment comparing basin yield outputs using modeled and observed climate inputs suggests that model skill is generally too poor for clear region–ensemble couplings to emerge (see supplemental material), a more thorough investigation is needed. In addition, further research is needed to evaluate implications for basin yield and adaptation costs incorporating a wider set of projected global changes, including rising food demands, population increases, and so forth. Here we have provided an indicative analysis of how historical basin yields and irrigation water requirements would be affected under projected climate change but assuming current population, per capita food consumption, and per capita water use; further work is required to consider other projected global changes.

Acknowledgments. We are grateful for technical contributions provided by Diane Ivy and Yohannes Gebretsadik. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP. We also thank the climate modeling groups (listed in Tables S1 and S2 of the supplemental material of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for
Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

REFERENCES


