Targeted model evaluations for climate services: a case study on heat waves in Bangladesh

Article:

Nissan, Hannah, Muñoz, Ángel G. and Mason, Simon J. (2020) Targeted model evaluations for climate services: a case study on heat waves in Bangladesh. Climate Risk Management, 28. ISSN 2212-0963

https://doi.org/10.1016/j.crm.2020.100213

Climate Risk Management 28 (2020) 100213

Contents lists available at ScienceDirect

Climate Risk Management

journal homepage: www.elsevier.com/locate/crm

Targeted model evaluations for climate services: A case study on heat waves in Bangladesh

Hannah Nissan[a,b,*], Ángel G. Muñoz[a], Simon J. Mason[a]

[a] International Research Institute for Climate and Society, Columbia University, 61 Route 9W, Palisades, NY 10964, United States
[b] Centre for Analysis of Time Series, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom

ABSTRACT

Though not a sufficient condition, the ability to reproduce key elements of climate variability over the historical record should be a minimum requirement for placing any confidence in a model's climate forecasts or projections of climate change. When projections are used to guide practical adaptation, model evaluations should focus on the weather and climate events of interest to decision-makers, their physical drivers in the climate system and their variability on decision-relevant timescales. This paper argues for a greater emphasis on such targeted model evaluations to enable useful climate services. We illustrate this approach through a case study on heat waves in Bangladesh, but draw wider conclusions that are applicable to climate services development more broadly.

The simulation of heat waves in Bangladesh is evaluated in several climate models, focusing on timescales relevant to the long-term viability of a heat action plan: the average, interannual variability and seasonality of temperature and heat-wave frequency. Where the physical drivers of variability are broadly captured, a considered interpretation of the models could provide insights into future heat-wave behaviour. However, substantial biases are found in the statistics and in some physical drivers of heat, raising questions about the suitability of some of the models for determining certain aspects of future risk. Specifically, simple bias corrections cannot be used to make inferences about possible future changes in various weather statistics such as timing of heat waves during the year. Results emphasize the potential pitfalls of performing only perfunctory climatological evaluations and highlight areas for model improvement in the simulation of South Asian climate variability.

1. Introduction

"All models are wrong, but some are useful" is an oft-quoted maxim by the statistician George Box (Box, 1979). Although the maxim was originally coined with statistical models in mind, few would doubt that it applies just as validly to dynamical models used in weather and climate research and prediction (Spiegelhalter and Riesch, 2011). The usefulness of numerical weather prediction models is indisputable (Bauer et al., 2015): early warnings provided by weather forecasts have saved many lives, and have the potential to save many more in developing countries (Hallegatte, 2012), while financial and other sectors have taken advantage of forecasts to trade and plan ahead (CME Group, 2007; Ramanathan et al., 1997; Taylor and Buizza, 2003; Roulston et al., 2003). However, that these weather models contain systematic errors is equally indisputable (Bauer et al., 2015). Similarly, both the usefulness and the limitations of *climate* models are indisputable. Considerable progress has been made in seasonal climate forecasting, for example, in recent decades (Goddard et al., 2001; Hansen et al., 2006; Doblas-Reyes et al., 2013), but the probabilistic reliability of seasonal forecasts remains a major challenge over much of the globe (Weisheimer and Palmer, 2014).

Understanding when and where the models are capable of performing well, and why (Tebaldi and Knutti, 2007; Gleckler et al., 2008), is thus essential to ensuring the delivery of useful climate services (Hewitson et al., 2014). While undeniably instrumental in many situations, all weather and climate models are limited in their ability to forecast the future accurately and have the potential to

* Corresponding author at: Centre for Analysis of Time Series, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom.

E-mail address: hannah@iri.columbia.edu (H. Nissan).

https://doi.org/10.1016/j.crm.2020.100213
Received 3 May 2019; Received in revised form 8 November 2019; Accepted 3 February 2020
Available online 08 February 2020

be extremely misleading if used in the wrong way (Knutti et al., 2010; Hewitson et al., 2017; Nissan et al., 2019).

In this paper, we demonstrate the type of model evaluations and interpretations that are needed to underpin tailored climate services for decision-makers, given that the only models we have to work with are flawed. Such evaluations are an essential first step to developing climate services to support decision-making. If the models cannot capture the basics we have no justification for using them to predict the future. Often, however, evaluations will reveal areas of strength and weakness that can inform the way in which climate risks are communicated to decision-makers. A secondary benefit of a decision-tailored evaluation of the models is that a fuller understanding of model errors, focused on the phenomena and timescales of relevance to particular applications, will enable model improvements in the most important areas for practical decision-making. We illustrate these evaluation methods for a case study where long-term information is needed to inform climate change adaptation, but the methods are equally pertinent for climate services across all timescales.

Much of the information generated by climate predictions from seasonal to multi-decadal timescales is of limited use for practical decision-making, in part because of a tendency to focus on average conditions rather than critical thresholds, and because of inadequate consideration of model errors, uncertainties and climate variability (UNDP, 2005; van Aalst et al., 2008; Baethgen, 2010; Hansen et al., 2011; Smith and Stern, 2011; Fischer et al., 2013; Coughlan de Perez et al., 2014; Smith and Petersen, 2014; Coughlan de Perez et al., 2017). Model evaluations should focus on the specific weather and climate events of interest to stakeholders and on the temporal variability of those events on timescales relevant to the decisions being taken (Red Cross Red Crescent Climate Centre, 2017; Nissan et al., 2019). To inform climate *change* adaptation, information is needed about the changing nature of climate variability on decision-relevant timescales (Ranger and Garbett-Shiels, 2012; Baethgen and Goddard, 2013; Vincent et al., 2014; Jones et al., 2015), yet climate change projections suffer from some important limitations, already present at shorter timescales (Muñoz et al., 2017), which seriously compromise their utility for informing practical adaptation. While climate models are capable of reproducing many important aspects of the climate system, including observed temperature trends in many parts of the world, they have deep structural errors (Frigg et al., 2014). Model errors are particularly problematic when projections are desired at any scale smaller than sub-continental, in the extremes of variable distributions and where precipitation is concerned (Stephens et al., 2010; Kharin et al., 2013; Gonzalez et al., 2014); i.e., in virtually all situations in which information might be useful to decision-makers.

Developing services on climate-change timescales poses additional challenges, which make such targeted model evaluations even more critical than on shorter timescales. On weather and, to some extent, seasonal timescales, there exist sufficient past forecasts that can be tested against real outcomes to determine when the model is and is not useful. When projecting decades ahead in a non-stationary climate, there is no such set of previous forecasts. The probabilities of different model outcomes occurring in the real world thus cannot be quantified reliably (Kandlikar et al., 2005; Smith and Stern, 2011; Hallegatte et al., 2012; Smith and Petersen, 2014). Although the failure of climate model ensembles to represent the probability distribution of outcomes has long been recognised, it remains standard practice to interpret the relative frequencies of particular outcomes within a multi-model ensemble as an indication of their likelihood of occurrence, with little consideration of the possibility of outcomes outside the ensemble spread (Smith, 2002; Wood and Moriniere, 2013; Smith and Petersen, 2014; Frigg et al., 2015). Such methods give decision-makers a false impression of confidence and could encourage poor adaptation choices that are not robust to uncertainty (Hall, 2007; Spiegelhalter and Riesch, 2011; Nissan et al., 2019; Thompson and Smith, 2019).

Given that the reliability of probabilistic climate projections cannot be established, thorough model evaluation over the historical record is the only means we have of assessing confidence in model projections (Frigg et al., 2015). Even then, in a non-stationary climate past model performance is a necessary, but not a sufficient, condition of performance in the future (Smith, 2002). Unfortunately, the process of determining the suitability of models to inform practical climate change adaptation decisions cannot be reduced to a one-size-fits all approach (Steynor et al., 2016). Such a model evaluation should not be conducted in order to select the best model, nor to narrow the spread of the future ensemble, but to judge to what extent we can be confident in the model projections from a position of scientific understanding about the situations in which the models perform well and, conversely, where they fail (Knutson et al., 2010; Thompson et al., 2016; Nissan et al., 2019).

The following sub-section outlines the case study used in this paper, which concerns heat waves in Bangladesh. Section 2 explains the data sources and methods. Section 3 presents the results from three categories of model evaluation: basic statistics of temperature, basic statistics of heat-wave frequency and the physical mechanisms of heat-wave formation and variability. In Section 4, we interpret the results of these evaluations to infer where there is potential for the models to be informative about the future behaviour of heat waves in Bangladesh, and where the limitations of the models mean that forecasts and projections should be viewed with particular caution and substantiated by further analysis. This study demonstrates a crucial first step to developing informative projections for decision-makers, namely a practice for targeted model evaluations focused on decision-relevant variables and timescales; an analysis of the projections themselves is beyond the scope of this paper.

1.1. Case study description

We use a hypothetical example in which information is demanded about future heat-wave occurrence in Bangladesh in order to inform strategies for managing heat-health in humans and livestock/poultry and adapting cropping varieties, but the approach could be applied to any decision-making context.

We chose this example for a number of reasons:

- Heat waves are associated with livestock and poultry disease and with crop damage in Bangladesh (Ruane et al., 2013; Amin et al., 2015; Goosen et al., 2018).

- Evidence of the human cost of heat is mounting, and there is momentum among international risk management communities to better manage these health implications, particularly in South Asia (GHHIN 2019; WHO-WMO, 2016). In Bangladesh, mortality rates increase by about 20% during heat waves (Nissan et al. 2017).
- From a practical perspective, temperature is generally better captured by climate models than precipitation, so there is some hope that the models may provide useful information.
- It is commonly assumed that climate change will result in a simple shift of the temperature distribution, and thus that projections of average temperature change can serve as a proxy for temperature extremes (Wade et al., 2015). However, the frequency, location or strength of atmospheric circulation patterns associated with heat waves (Quesada et al., 2012; Miralles et al., 2014; Perkins et al., 2015; Nissan et al., 2017) could also change as emissions increase. An evaluation of these dynamic processes in the models is needed to test this assumption of a simple shift in the temperature distribution.

An effective heat action plan involves managing heat-wave risk across a range of timescales (Knowlton et al., 2014; WHO-WMO, 2016; Tompkins et al., 2018). To support the long-term viability of a heat action plan, information about plausible changes on the following timescales may therefore be useful:

i. Average changes in heat-wave frequency and intensity: to guide long-term policies regarding the development of heat-resistant crop varieties (Hossain and Teixeira da Silva, 2013), health systems and climate services (Nissan and Conway, 2018), and to inform long-term building and infrastructure design (Hacker et al., 2005);
ii. Changes in the timing of heat-waves during the year (the annual cycle or seasonality): seasonal preparedness is the foundation of a heat action plan (e.g. Knowlton et al., 2014; Public Health England et al., 2015) and is vital to the agricultural calendar. Any changes in the timing of heat waves during the year would affect the optimum timing of seasonal measures (Hess and Ebi, 2016). Especially important are changes in the first heat waves of the year, which cause the most mortalities (Baccini et al. 2008; Lee et al., 2014), and the last of the year, to ensure that heat action plans are not terminated prematurely before the season has ended.
iii. Changes in interannual heat-wave variability (Coughlan de Perez et al., 2014): such changes may require strategies to balance resources across years, such as flexible multi-year humanitarian funding mechanisms (Cabot-Venton, 2013), and adaptive communication strategies (Hess and Ebi, 2016). An increase in interannual variability may motivate investments in the generation and uptake of seasonal forecasts to support adaptive agricultural decision-making.

The analyses presented below assess key aspects of model performance that must be understood in order to interpret projected changes in these three timescales of heat-wave variability. We examine the basic statistics of temperature and heat-wave frequency as well as the physical drivers of heat-wave variability on weather-to-interannual timescales. Heat waves in Bangladesh occur during the pre-monsoon season, between April and June (AMJ). This season is characterised by hot weather, thunderstorms and intense but sporadic convective precipitation, which depends on southerly winds from the Bay of Bengal for its supply of moisture (Sanderson and Rafique, 2009). Heat waves are associated with stronger-than-normal westerly winds, which import hot, dry air from northern India, weakening the advection of moisture from the south and suppressing rainfall. The association of heat waves with dry conditions also extends to seasonal and interannual timescales. Anomalously dry soil moisture conditions and low accumulated precipitation occur for up to 60 days in advance of a heat wave. Interannually, variations in seasonal precipitation and soil moisture are related to the frequency of heat-wave days, with drier than normal years associated with higher heat-wave frequencies (Nissan et al., 2017).

2. Data & methods

2.1. Data

Gridded daily data from APHRODITE (Yatagai et al., 2012) were used to evaluate the temperature and precipitation fields simulated by the climate models. APHRODITE is the only long-term data product that provides gridded daily data at a high resolution (0.25°) for South Asia. The product is based on a dense network of many more observing stations than are available to the Global Telecommunications System and thus to most other gridded products, of which there are few available at a daily resolution. There is a lack of publicly available in situ observations for Bangladesh, but APHRODITE has shown closer agreement with station data than a range of other gridded datasets over neighbouring India (Prakash et al., 2015; Bandyopadhyay et al., 2018).

Ideally, direct observations would be used to evaluate all model fields, but examining the physical drivers of heat-wave occurrence requires variables that are unavailable in most datasets: atmospheric circulation, soil moisture and daily minimum and maximum temperature (used to identify heat waves; Nissan et al., 2017). Reanalyses have the advantage of providing gridded data with complete 3D global coverage for a wide variety of fields, and of maintaining physical consistency among variables. The European Centre for Medium-Range Weather Forecasts interim reanalysis product (ERA Interim) was used to evaluate circulation, soil moisture and heat-wave statistics and as an additional evaluation dataset for temperature and precipitation. Analysis fields were obtained for air temperature at 2 m above ground level, and zonal and meridional winds at 500 hPa. Daily maximum and minimum temperatures were taken to be the maximum and minimum of these six-hourly values. Volumetric soil water content was calculated by integrating across all four soil layers in the ERA-Interim land model. Modelled precipitation was used from the reanalysis, since precipitation is not an assimilated variable. Since reanalysis data are themselves the output of a model, albeit one that is constrained by observations, cross-verification of the 500 hPa winds and soil moisture was carried out using the Japanese Meteorological Agency

Table 1
Summary characteristics of climate models evaluated for this study, selected based on availability.

Abbreviation	Modelling centre	Resolution (°E × °N)	Ensemble number	References	Comments
CCSM4	National Center for Atmospheric Research (UCAR, USA)	1.25 × 0.94	6	(Gent et al., 2011)	Precipitation unavailable
CMCC-CESM	Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC, Italy)	3.75 × 3.71	1	(Fogli et al., 2009)	Soil moisture unavailable
CMCC-CM	Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC, Italy)	0.75 × 0.75	1	(Scoccimarro et al. 2011)	
CMCC-CMS	Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC, Italy)	1.88 × 1.87	1	(Scoccimarro et al., 2011)	
CSIRO-Mk3-6–0	CSIRO Climate Science Centre (Australia)	1.88 × 1.87	1	(Gordon et al., 2002)	
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory (GFDL), Princeton University (USA)	2.5 × 2.02	1	(Dunne et al., 2012, 2013)	
HadGEM2-CC	Hadley Centre, UK Met Office	1.88 × 1.25	1	(Martin et al., 2011)	
inmcm4	Russian Institute for Numerical Mathematics (Russia)	2.0 × 1.5	1	(Volodin et al., 2010)	
IPSL-CM5A-LR	Institut Pierre Simon Laplace (France)	3.75 × 1.90	1	(Dufresne et al., 2013)	
MIROC5	University of Tokyo Center for Climate System Research (Japan)	1.41 × 1.40	1	(Watanabe et al., 2010)	
MPI-ESM-MR	Max Planck Institute for Meteorology	1.88 × 1.87	1	(Giorgetta et al., 2013)	Soil moisture unavailable

reanalysis product (JRA55, Kobayashi et al., 2015). The ability of ERA Interim to capture heat-wave statistics was evaluated for the period 1989–2011 for which minimum and maximum temperature data were available from the Bangladesh Meteorological Department (see Supplementary Online Materials).

Data from the historical experiment of the Coupled Model Intercomparison Project-5 (CMIP5) were obtained for 11 climate models (Table 1), selected based on availability of the variables needed for analysis. Since our goal was to evaluate the model processes, rather than the reliability of probabilities derived from the model ensembles, we used one ensemble member for each climate model, selecting the lowest ensemble number for which all variables were available. This approach also accounted for the inconsistent ensemble sizes available for each model.

The models were evaluated over the period 1979–2005, for which CMIP5, ERA Interim and APHRODITE data were available. The APHRODITE data were aggregated to a resolution of $1 \times 1^{\circ}$ and all model data were interpolated onto this same grid before averaging over Bangladesh. This resolution is roughly in the middle of the range of model resolutions and so does not unfairly favour or disadvantage any particular model. The overall conclusions were insensitive to the choice of resolution (not shown).

2.2. Heat-wave definition

Heat-wave days were defined following Nissan et al. (2017), who showed an increase of about 20% in all-cause mortality in Bangladesh when both day- and night-time temperatures remained elevated above the 95th percentile for at least three consecutive days. Percentiles were defined across all days during the study period, 1979–2005. Percentiles were defined separately within each model to correct for mean temperature biases.

2.3. Statistical methods

Heat waves in Bangladesh occur between April and June (AMJ; Nissan et al. 2017), so seasonal averages were calculated over the AMJ heat-wave season. Daily climatologies were computed by fitting a sixth-order harmonic function to the raw daily climatological values from 1979 to 2005. Substantial smoothing was needed to compare the climatological timings of peak precipitation and temperature among the models, which, on comparison with several smoothing methods tested (fourth-, fifth- and sixth-order Fourier series and 30- and 60-day running averages), was best achieved by the sixth-order Fourier series. The climatological peak timings were estimated by finding the maximum of the smoothed climatologies. To calculate the interannual variability in peak timing, the timing of the peak in each year was located from smoothed daily values, and the standard deviation in this timing was computed across all years. To smooth the daily values in each year, a 30-day running average filter was used in place of the Fourier series used for the climatologies because, for individual years, the fit was highly sensitive to the number of harmonics used in the fitted Fourier series.

Linear trends were estimated using two different regression methods. Ordinary least squares (OLS) was used to compute temperature trends, with confidence intervals based on t-statistics under an assumed normal distribution of errors. For annual heat-wave frequency, errors were evidently non-Gaussian, so a different methodology was used, which estimates the trend as the median of ranked pairwise slopes, with errors determined using quantiles of Kendall's t-statistic (Conover, 1999).

Spearman rank correlation coefficients (ρ) were calculated to compare the strengths of relationships between variables in the models and observations/reanalyses. Statistical significance was assessed by comparing the computed value of ρ with the quantiles of the Spearman rank correlation coefficient at the 5% significance level (Conover, 1999).

We do not impose a minimum performance threshold to determine the adequacy of the climate models evaluated, recognising that it is impossible definitively to determine when a model is "good enough" (Knutti et al., 2010; Gleckler et al., 2008). Instead, departures of the models from observations/reanalyses are presented as a proportion of observed/reanalysis interannual variability, since natural year-to-year fluctuations are an important timescale for climate risk management (Ranger and Garbett-Shiels, 2012; Baethgen and Goddard, 2013; Vincent et al., 2014; Jones et al., 2015).

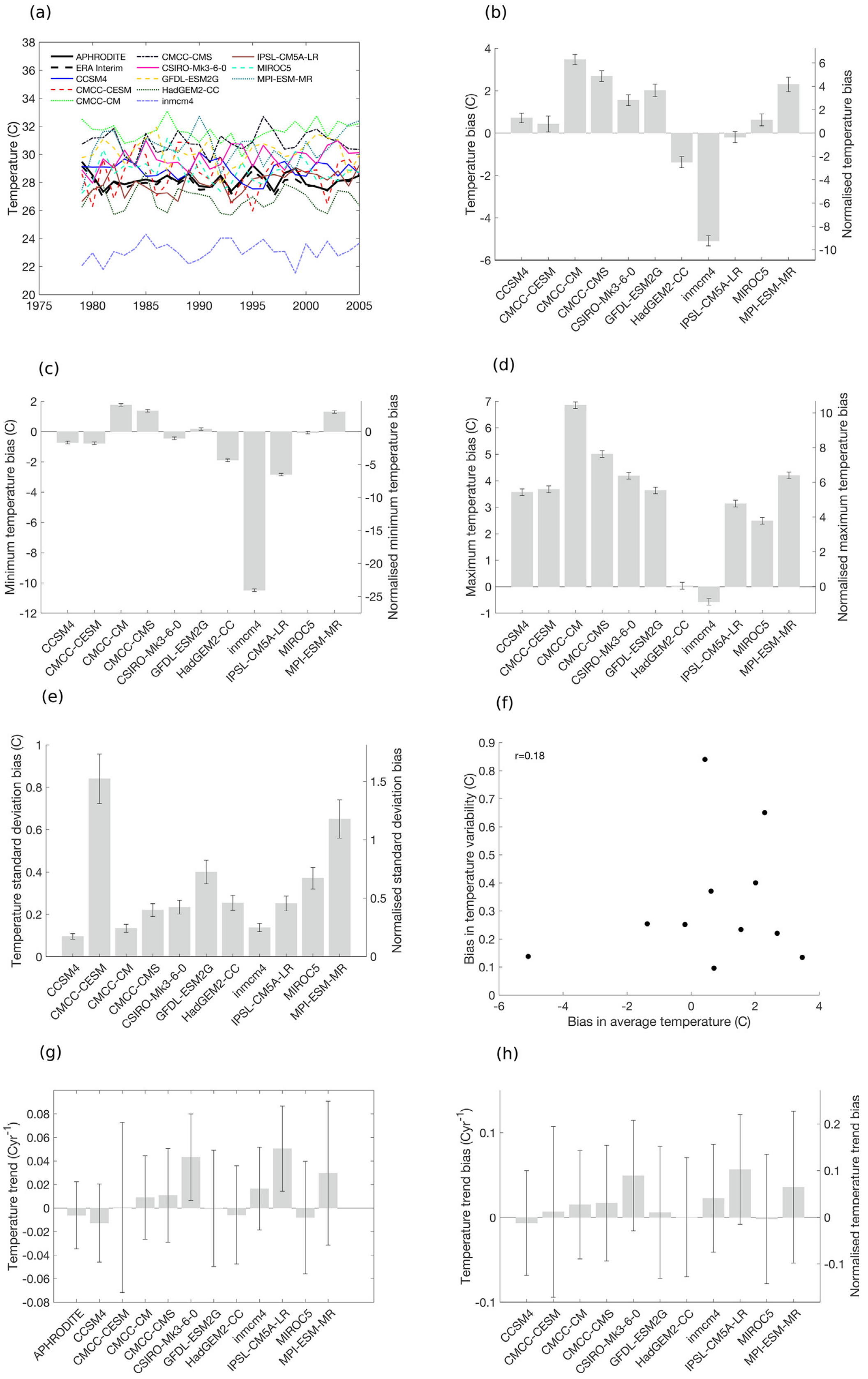
3. Results

3.1. Basic statistics of temperature

3.1.1. Average temperature during the heat-wave season

Fig. 1 compares the average, interannual variability and linear trend in seasonal average (AMJ) temperature over the period 1979–2005, between observations, reanalyses and the eleven CMIP5 models analysed. ERA Interim reanalysis exhibits a slight cool bias compared with gridded observations from APHRODITE (seasonal average temperature of 28.0 ± 0.1 °C vs 28.2 ± 0.1 °C) (Fig. 1a). Interannual variability in AMJ seasonal temperature was similar in both datasets (0.5 ± 0.1 °C). No trend was found in AMJ temperature between 1979 and 2005. Average daily minimum and maximum temperatures during the heat-wave season were 25.3 °C and 31.0 °C, respectively, in the ERA Interim reanalysis (the only data available for minimum/maximum temperatures).

Average temperature during the heat-wave season is higher than observed in most models, with absolute biases ranging from −5.1 °C (INMCM4) to +3.5 °C (CMCC-CM) (Fig. 1a and b, left axis). These biases are outside the range of observed interannual variability for all models except CMCC-CESM and IPSL (Fig. 1b, right axis). Most of the warm bias can be attributed to daily maximum (day-time) temperatures being too high; daily minimum (night-time) temperatures were closer to observed (in the reanalysis) for most models (Fig. 1c & d). Thus, the diurnal range is too large in all models except HadGEM2 and INMCM4. For IPSL, the



(caption on next page)

Fig. 1. Evaluation of seasonal average temperature between April and June (AMJ) over Bangladesh in 11 CMIP5 models compared with APHRODITE and ERA Interim. (a) annual AMJ-averaged temperature (°C) in all models (colours) vs APHRODITE (bold black line) and ERA Interim (dashed bold black line); (b) bias in AMJ-averaged daily mean temperature compared with APHRODITE (left axis, °C) and the bias normalised by the observed interannual standard deviation (right axis, no units); (c) as in (b) but for daily minimum temperature, compared with ERA Interim; (d) as in (c) but for daily maximum temperature; (e) bias in interannual standard deviation of AMJ-averaged daily mean temperature compared with APHRODITE (left axis, °C) and the bias normalised by the observed standard deviation (right axis, no units); (f) biases in average (b) and standard deviation (e) of AMJ temperature in the models ($r = 0.19$), relative to APHRODITE; (g) OLS linear trend in AMJ-averaged daily mean temperature (°Cyr^{-1}); (h) bias in the OLS linear trend in AMJ-averaged daily mean temperature compared with APHRODITE (left axis, °Cyr^{-1}), and the bias normalised by the observed standard deviation (right axis, yr^{-1}). Error bars in (b), (c), (d) and (e) represent the standard error in the mean/standard deviation, as appropriate.

warm bias in day-time temperature (Fig. 1d) is offset by a cold bias in night-time temperature (Fig. 1c), resulting in no overall bias in mean-temperature (Fig. 1b). INMCM4 shows a substantial cold bias (−5.1 °C) in seasonal mean temperature, but in this case the model error mainly relates to night-time temperatures, consistent with Volodin et al. (2017); AMJ-averaged daily minimum temperatures in this model are over 10 °C lower than observed in the reanalysis.

All models exaggerate the interannual variability in temperature during the heat-wave season, with two models (CMCC-CESM and MPI) fluctuating over twice as strongly as observed from year-to-year (Fig. 1e). All but two models simulated insignificant trends, in agreement with observations, while the IPSL and CSIRO models produced significant warming (Fig. 1g). However, none of the models had a significant bias in the simulated trend (Fig. 1h).

3.1.2. Annual cycle of temperature

Temperature in Bangladesh is strongly seasonal, with the hottest conditions experienced between April and May and the coldest conditions in January (Fig. 2a). The annual cycle of maximum (day-time) temperature peaks close to 30 °C (averaged across the country), followed by slightly cooler temperatures during the monsoon season. Minimum (night-time) temperatures exhibit a less well-defined peak around June, with hot nights persisting for several months. These cycles are consistent with those computed using station data from 1989 to 2011 by Nissan et al. (2017, Fig. 2). The timings of the hottest day- and night-time temperatures vary substantially from year to year. The hottest days can arrive as early as March or as late as the end of May, while the warmest nights occur between mid-May and mid-July (Fig. 2a).

Substantial variation in the shape of the seasonal temperature cycles can be seen among the climate models (Fig. 2). Most models simulate a peak in maximum temperature during the pre-monsoon AMJ season, as observed in the reanalysis. However, some fail to capture this characteristic cycle (e.g., CMCC-CESM and IPSL). All eleven CMIP5 models analysed have a delayed peak in maximum temperature, with peaks occurring later during the year than observed (Fig. 3). For minimum temperature, the peaks in the models are distributed more evenly around the observed peak in June. All models fall within the range of observed interannual variability for minimum temperature, but for three models (CSIRO, HadGEM2 and IPSL) maximum temperature peaks so late in the year that it is outside the range of observed interannual variability.

3.2. Basic statistics of heat waves

3.2.1. Annual heat-wave frequency

Fig. 4a shows the climate model biases in average annual frequency of heat-wave days. Between 1979 and 2005, an average of 2.3 heat-wave days were observed per year according to ERA-Interim. Biases are not significant for two models (CMCC-CESM and MIROC5). CCSM4 is the only model that simulates too few heat waves, except INMCM4, which fails to produce any heat waves at all. Since the heat waves were defined relative to the temperature distribution in each model, these biases are in addition to any mean temperature bias and therefore cannot be accounted for through a bias correction of average temperature. Fig. 5 confirms that the observed distributions of daily minimum and maximum temperatures are not simply shifted by the models: differences are also evident in the shapes of the distributions. CMCC-CM, CMCC-CMS, GFDL, CSIRO, HadGEM2, IPSL and MPI have significantly more heat waves than the reanalysis. Biases range from −56% to +75% of the observed interannual variability in the number of heat-wave days (Fig. 4a, right axis).

Fig. 4b shows the interannual variability in the frequency of heat-wave days, measured as the standard deviation across all years. Between 1979 and 2005, the number of heat-wave days per year varied interannually with a standard deviation of 3.3 days (ERA Interim). In keeping with the large modelled interannual temperature variability during the heat-wave season (Fig. 1e), the interannual heat-wave variability is also too high in most models, reaching close to twice the reanalysis value in several models (Fig. 4b, right axis). CCSM4 exhibits a different behaviour to the other models: the underestimation of interannual heat-wave variability in CCSM4 appears to contradict the positive bias in interannual temperature variability shown in Fig. 1e. A bias in temperature variability cannot therefore be used generically to infer biases in the variability of extremes.

Fig. 4c and d compare two methods of estimating the trend and error bars in annual frequency of heat-wave days over the study period. Trends, though not significant for all but two models (IPSL and MPI) according to the parametric method, are substantially larger than those estimated using the non-parametric method. The non-parametric method indicates no trend in the reanalysis, nor in any of the models, save two. These two models, GFDL and IPSL, showed large and significant positive trends when estimated parametrically, but have strong negative trends using the non-parametric approach, though the true trend could still be zero at the uppermost limit of the uncertainty range. The MPI model has a significant positive trend when estimated using OLS, reducing to zero

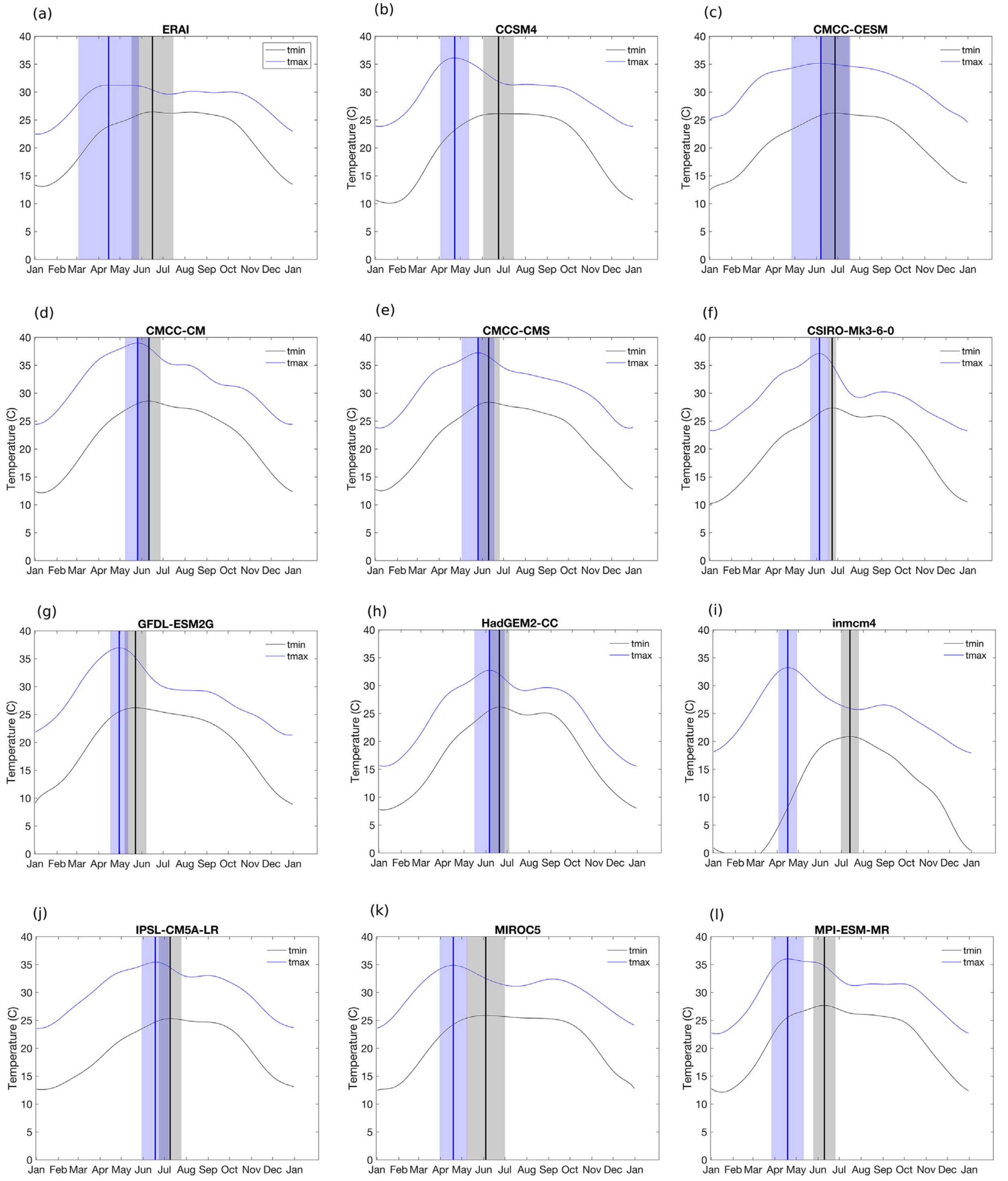


Fig. 2. Daily climatologies of minimum (black) and maximum (blue) temperatures over Bangladesh in ERA Interim and in 11 CMIP5 models. The vertical lines indicate the climatological timing of peak minimum (blue) and maximum (black) temperature. The wide vertical bands represent the interannual variability in the timing of peak minimum (grey) and maximum (blue) temperatures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

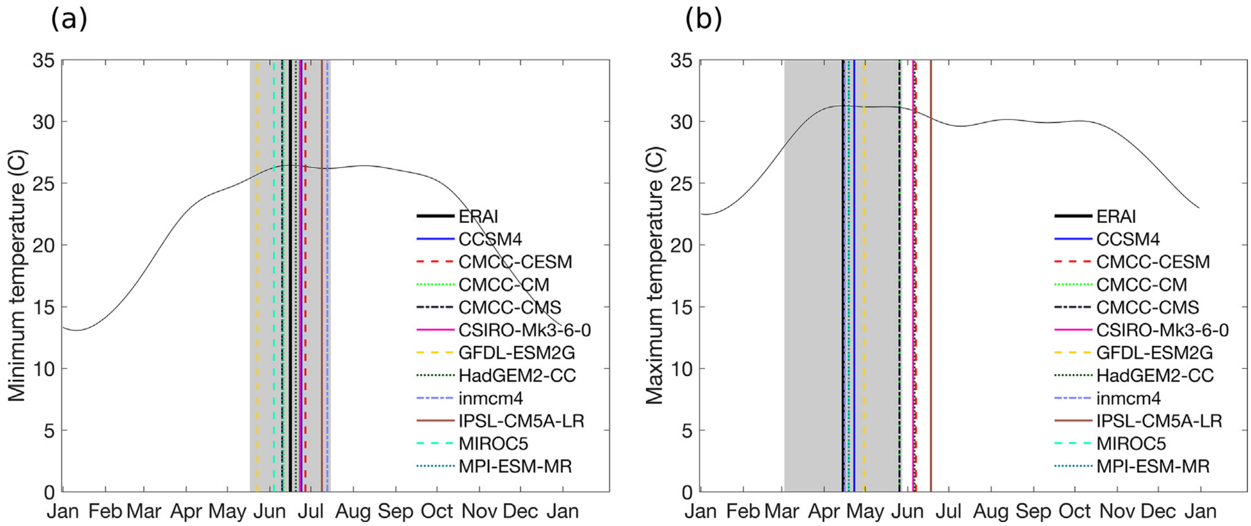


Fig. 3. Daily climatology of ERA Interim (a) minimum and (b) maximum temperature (black curves). Vertical lines represent the timing of the peak climatological temperature during the year in ERA Interim (thick black line) and in the CMIP5 models (coloured lines). The grey band represents the interannual variability in the timing of peak temperature in ERA Interim.

using non-parametric linear regression. Error ranges are asymmetric in Fig. 4d and include zero trend for all models.

3.2.2. *Timing of heat waves during the year*

Heat waves occur between April and June in Bangladesh, with most in May (Fig. 6a), consistent with results using station observations in Nissan et al. (2017, Fig. 3a). The climate models analysed consistently show a delay in the timing of the heat-wave season (Fig. 6). April heat waves are only captured by GFDL, MIROC5 and MPI, with all other models simulating a later start to the heat-wave season, in May. Heat waves also persist later in the year in eight of the models (all three CMCC models, CSIRO, HadGEM2, IPSL, MIROC5 and MPI). In several models, the maximum interannual range during the peak heat-wave month is much higher than observed in the reanalysis. These differences in magnitude and phase of the average heat-wave cycle in each model are summarised using the root mean squared error (RMSE) in Fig. 6l.

3.3. *Mechanisms of heat-wave formation and variability*

In this section, we investigate possible underlying reasons for the systematic errors indicated in the Sections 3.1 and 3.2. Circulation and precipitation composites are analysed to assess whether the physical mechanisms found to drive heat-wave formation in Bangladesh (Nissan et al., 2017), are well-captured by the models.

3.3.1. *Weather timescale*

In the mid-atmosphere at 500 hPa, heat waves are associated with an anticyclonic circulation pattern positioned over central India. This pattern brings strong westerly winds over northern India and over Bangladesh, which turn northerly towards the coast of Bangladesh and over the Bay of Bengal (Fig. 7a and 6b). Eight of the models associate heat waves in Bangladesh with north/north-westerly circulation from northern India, though there are circulation differences in other parts of the domain, especially to the East of Bangladesh (Fig. 7). However, it is clear from the composites that the CMCC-CESM and IPSL models generate heat waves through an entirely different circulation pattern from that observed in ERA Interim. Pattern correlations between the heat-wave circulation composite observed in ERA Interim and each model quantify these differences, and confirm the conclusions drawn from visual comparison (Fig. 7m). The MPI, CMCC-CM, CMCC-CMS and CCSM4 models perform best (Spearman's correlations > 0.8), while IPSL and CMCC-CESM have the least satisfactory performance.

All the models analysed associated heat waves with below-normal precipitation, in agreement with observations (Fig. 8). However, in all models except HadGEM2 and MIROC5, the strength of this association is underrepresented and in the GFDL and IPSL models it is particularly weak.

3.3.2. *Seasonal timescale*

The persistent negative accumulated precipitation (Fig. 8) and soil moisture (Nissan et al., 2017) anomalies, seen for several weeks before a heat wave, provide an opportunity to investigate plausible reasons for the delayed annual cycle of maximum temperature and heat-wave frequency seen in Fig. 3 and Fig. 6. An absence of surface water is a precondition for intense heat waves, enabling temperatures to rise to extreme levels by shifting the partition of surface heat fluxes from latent to sensible heating. Any biases in the annual cycle of soil moisture in the models will therefore be reflected in the temperature cycle, particularly for

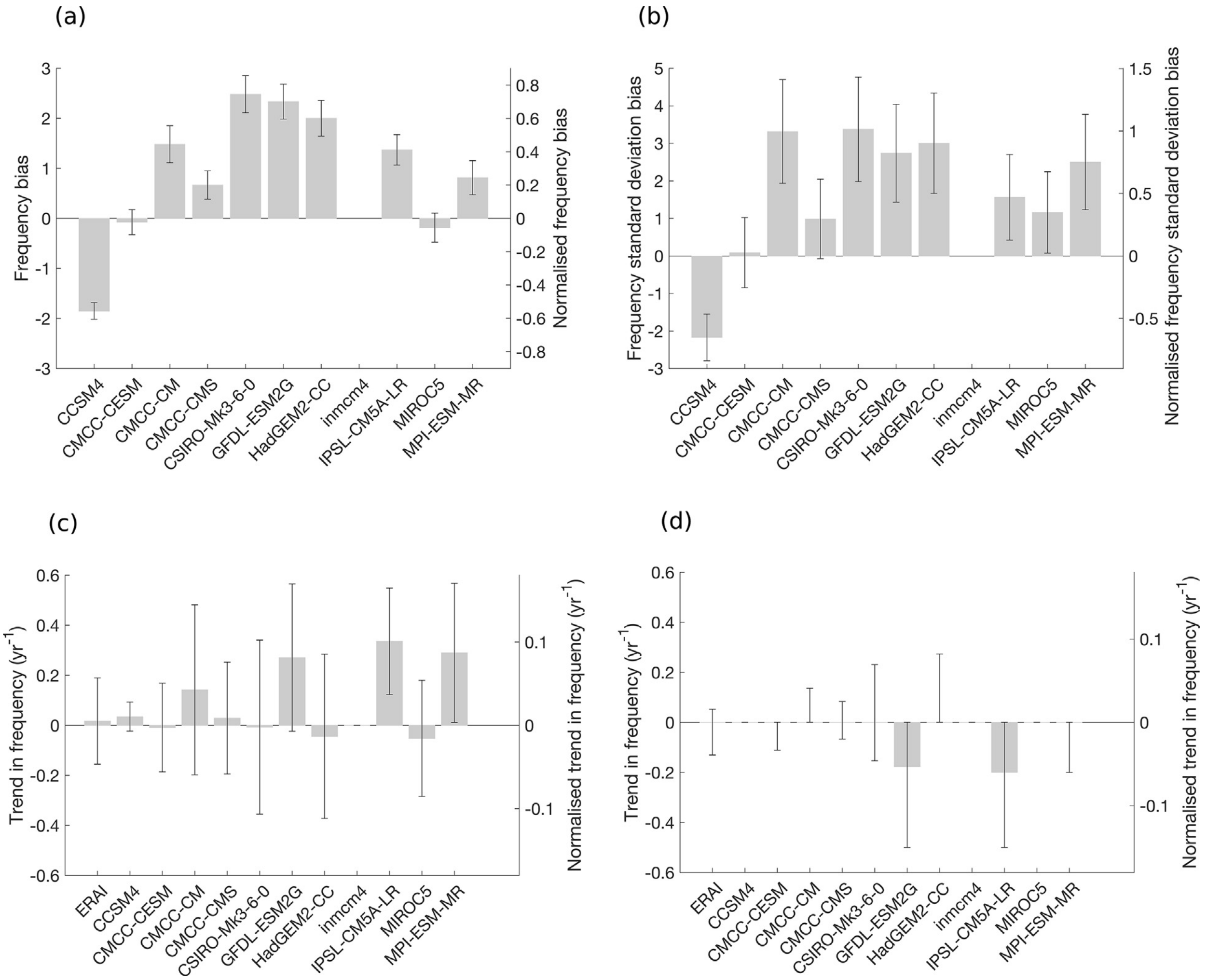


Fig. 4. Evaluation of annual frequency of heat-wave days in Bangladesh in 11 CMIP5 models compared with ERA Interim. (a) Bias in annual average frequency of heat-wave days (left axis) and the bias normalised by the observed (ERA Interim) interannual standard deviation (right axis); (b) bias in the interannual standard deviation of frequencies of heat-wave days (left axis) and the bias normalised by the observed interannual standard deviation (right axis); (c) & (d): trend in the frequency of heat-wave days (left axis), and the trend normalised by the observed standard deviation in frequency of heat-wave days (right axis). Trends and errors were estimated using (c) the OLS method and (d) ranked pairwise slope estimation with errors calculated using Kendall's t-statistic (see Section 2.3).

maximum temperatures.

Modelled soil moisture content in those CMIP5 models for which this variable was available, differs markedly from ERA Interim, both in the average (Fig. 9) and annual cycle (Fig. 10). Soil moisture during the AMJ heat-wave season averaged 0.88 m^3m^{-2} in the ERA Interim reanalysis, with an interannual standard deviation of 0.04 m^3m^{-2} (Fig. 9a). Comparison with JRA55 reveals some uncertainty in the quality of the reanalysis products: seasonal average soil moisture in JRA55 was 1.10 m^3m^{-2}, with a year-to-year standard deviation of 0.07 m^3m^{-2}. Model biases relative to ERA Interim ranged from -0.79 to $+4.74$ m^3m^{-2} (Fig. 9b), or -1.02 to $+4.51$ relative to JRA55. Regardless of the choice of reanalysis, these biases are at least an order of magnitude greater than the observed year-to-year variability in soil moisture for all models except CSIRO and HadGEM2 (Fig. 9b). In MIROC5, the bias is over 100 times greater than ERA Interim interannual variability, or 60 times greater than the variability in JRA55.

Ignoring, for the moment, the large biases in soil moisture, Fig. 10 compares the patterns of soil moisture seasonality in the climate models with ERA Interim and JRA55. Although the amplitude of the annual soil moisture cycle is notably larger in JRA55 than in ERA Interim, the reanalyses do agree on the timing of minimum and maximum soil moisture during the year. However, in the models, the delayed seasonality of maximum temperature and heat-wave frequency is mirrored by a delayed cycle of soil moisture. All the models show a clear lag of at least one month in the timing of the driest month. The same is true for the timing of the month with the most soil moisture during the year, except for the case of MIROC5, which simulates a peak in soil moisture in the same month as ERA Interim and JRA55, August. Fig. 11 explores whether deficiencies in the simulation of precipitation might explain the late timing of heat waves in most of the models. With the exception of MIROC5 and CMCC-CESM, peak precipitation in all the models

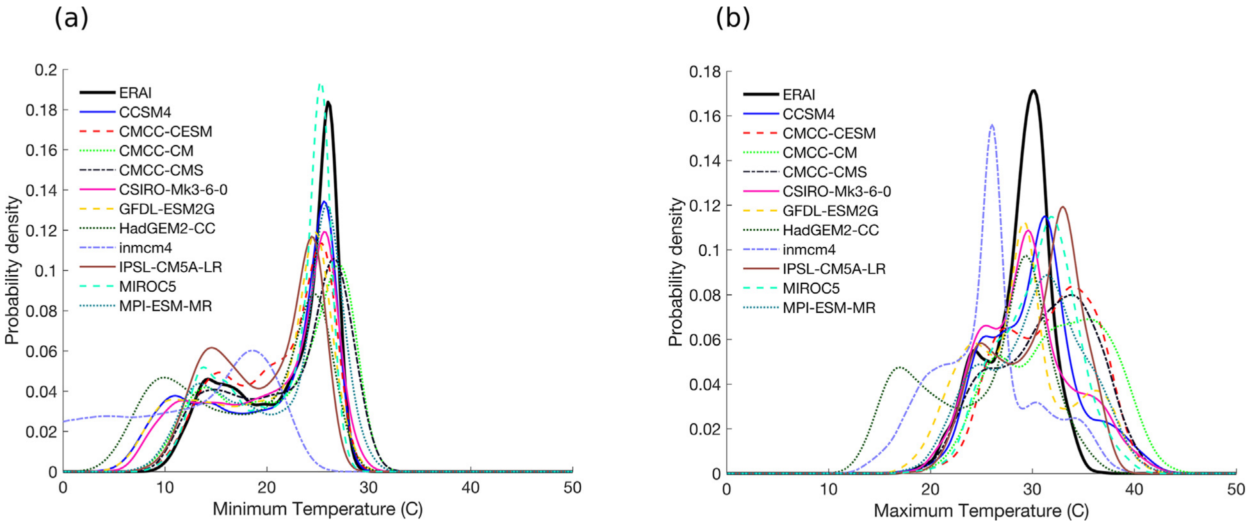


Fig. 5. Distributions of daily minimum and maximum temperatures in ERA Interim reanalysis (bold black curve) and in the CMIP5 models (coloured curves).

occurs later than observed. This delay is greater than the observed standard deviation in the timing of peak precipitation in four of the models (CMCC-CM, CMCC-CMS, CSIRO and IPSL (Fig. 11k)).

3.3.3. Interannual timescale

Years with more rainfall and soil moisture between April and May tend to have fewer heat-wave days (Nissan et al., 2017). Correlations between the number of AMJ heat-wave days and total seasonal precipitation and soil moisture were −0.55 and −0.58, respectively, between 1979 and 2005, in ERA Interim. For comparison, correlations for 1989–2011 were −0.3 and −0.6, respectively, using heat-wave frequencies calculated from meteorological station data (Nissan et al., 2017).

On weather timescales, the strength of the association between dry conditions and heat waves is underrepresented in all but one model (Fig. 8). This relationship is similarly weak on interannual timescales in several models (Fig. 12a): negative Spearman's correlations were found between AMJ total precipitation and heat-wave frequency for all models, but most underestimated the strength of the association and two were statistically insignificant. The observed negative correlation between AMJ soil moisture and heat-wave frequency (in ERA Interim) was captured by CMCC-CMS and HadGEM2, but was substantially underestimated by five of the models.

4. Discussion

In section 1, we identified three timescales as being particularly important for heat-risk management: i) annual average, ii) timing during the year (the annual cycle or seasonality) and iii) interannual variability. In this section, we ask to what extent the climate models analysed here could be used to infer information about future changes in heat-wave behaviour on these timescales. The selection of any minimum performance threshold for the models would be arbitrary, and would depend on the application and the decision-maker's appetite for risk. Instead, we aim to understand the biases and limitations of the models relating to these decision-relevant timescales in order to enable a critical interpretation of future projections.

(i) Changes in average heat-wave frequency

With the exception of the IPSL model, biases in seasonal average temperature exceed the range of observed interannual variability for all models. Since heat waves are defined relative to the climatology of daily temperature, some of the effect of the bias in mean temperature should be removed by the calculation. However, significant biases in annual heat-wave frequency still exist in eight of the models (Fig. 4a), suggesting that the models do not simply shift the observed daily temperature distribution. Rather, differences in performance for minimum and maximum temperature (Fig. 1c and d) and in the shapes of the minimum and maximum temperature distributions (Fig. 5) suggest that a simple evaluation or bias correction of climatological average temperature would not be sufficient to account for errors in the diurnal temperature range and in heat-wave occurrence.

The average heat-wave biases in the models could be attributable to persistence problems, as heat waves are defined only when hot conditions last for at least three days. The annual cycles of minimum and maximum temperature (Fig. 2) also provide some insight. Heat waves are defined when high day- and night-time temperatures occur at the same time. Observed maximum temperature peaks earlier in the year than minimum temperature (mid-April vs mid-June), but the timing of these peaks is sufficiently close to enable some of the hottest days and nights to occur simultaneously, at least during some years. The shapes of the minimum

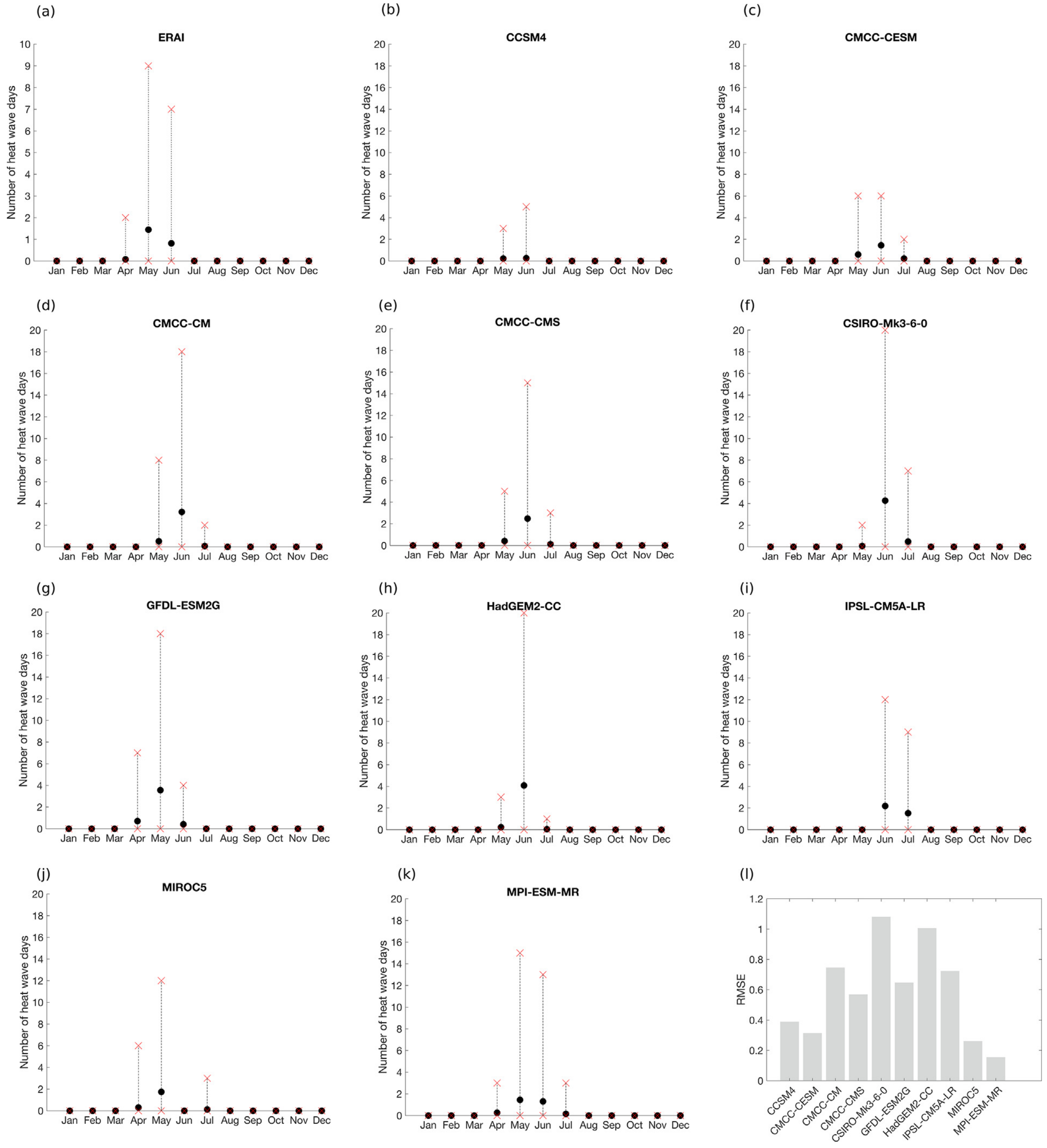
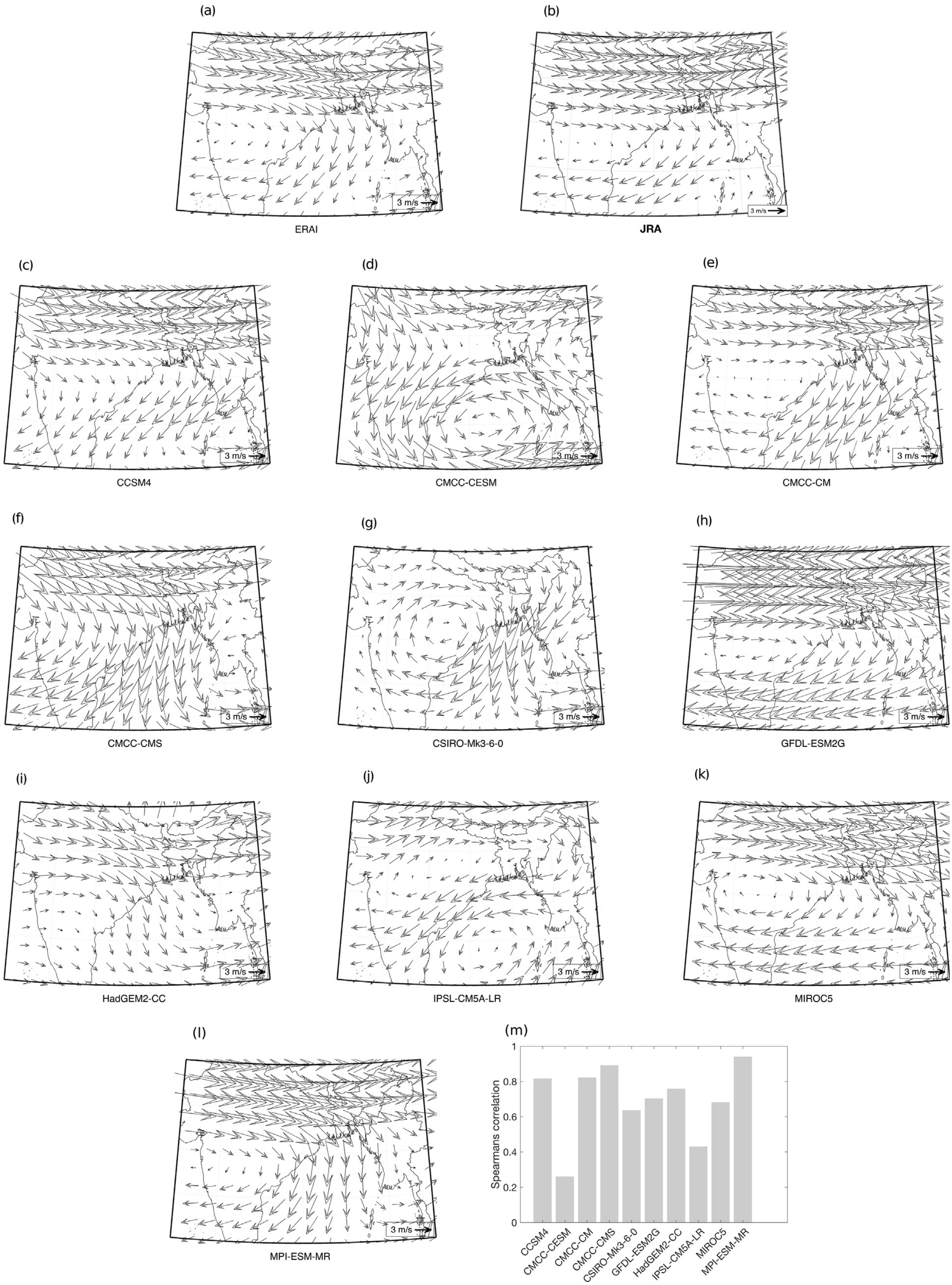


Fig. 6. Average monthly frequency (black dots) and interannual range (red crosses) of heat-wave days in ERA Interim and in 10 CMIP5 models over the period 1979–2005. The INMCM4 model was excluded because it did not simulate any heat waves. The last plot shows the root mean squared error in the average monthly number of heat-wave days in the CMIP5 models compared with ERA Interim. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and maximum temperature cycles are similarly important. Less pronounced seasonality indicates greater interannual variability in the timing of the hottest days and nights, enabling them to occur simultaneously even when there is a large difference in the average timing of the minimum and maximum temperature peaks.

In the INMCM4 model, the highest maximum temperatures occur in mid-April, while the highest minimum temperatures do not arrive until mid-July. This lag is considerably longer than observed (Fig. 2a) and is longer than in any of the other models analysed. Compounding the effect of this lag, the cycles of minimum and maximum temperature peak sharply, confining the occurrence of hot days and nights to a small window each year and reducing the opportunity for overlap. Consequently, INMCM4 fails to simulate any



(caption on next page)

Fig. 7. Wind composites on heat-wave days between 1979 and 2005. Wind vectors are shown at 500 hPa and composites are weighted by heat-wave duration. The last plot shows pattern correlations between the CMIP5 models and ERA Interim composites; 1 indicates a perfect match between the wind patterns.

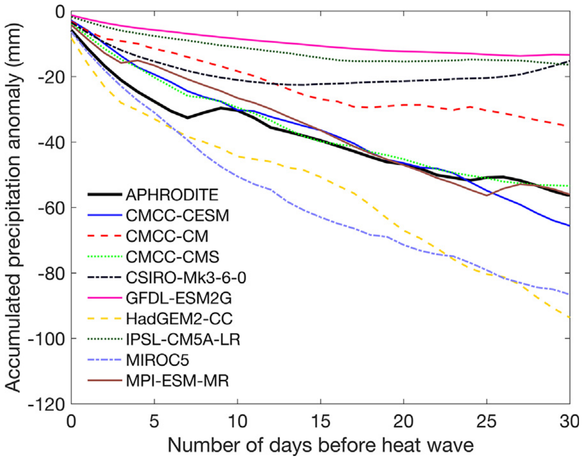


Fig. 8. Average accumulated precipitation anomaly (mm) on the days preceding a heat wave, weighted by heat-wave duration, in APHRODITE and the CMIP5 models. Anomalies were calculated relative to the smoothed daily climatology (see Section 2.3) and accumulated backwards in time from the first day of a heat wave. Confidence intervals were computed using bootstrap sampling of heat-wave days, with 1000 samples.

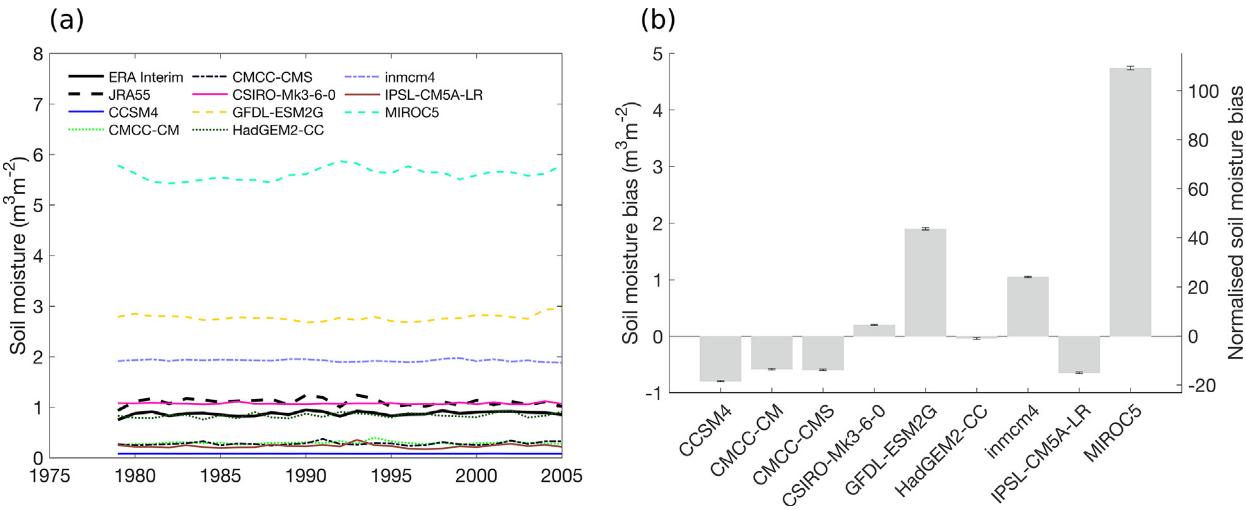


Fig. 9. (a) Time series of annual AMJ-averaged soil moisture (m^3m^{-2}), averaged over Bangladesh, in the CMIP5 models (colours) vs ERA Interim (bold black line) and JRA55 (dashed bold black line); (b) bias in AMJ-averaged soil moisture (left axis, m^3m^{-2}) and the bias normalised by the ERA Interim interannual standard deviation (right axis, no units). Error bars represent the standard error in the mean.

heat waves at all during the study period (including one- and two-day heat waves that were not included in our definition). CCSM4 is the only other model that does not produce enough heat waves each year (Fig. 4a) and, once again, this bias can be understood in the context of the annual cycle of temperature. In CCSM4, the lag between the average occurrence of the hottest days and nights exceeds the interannual variability in the timings of these peaks, making it less likely that hot days and nights will occur at the same time.

It is entirely possible that models that accurately reproduce the statistics of a given phenomenon could do so for the wrong reasons, leading to a cancelation of errors that masks underlying problems in the model. In such cases it does not make sense to look at a models' future projections. In broad agreement with observations (Fig. 7), heat waves are associated with north/north-westerly circulation over Bangladesh in all models except CMCC-CESM and IPSL (and INMCM4, which does not produce any heat waves). The anticyclonic circulation in the CSIRO model is constrained to central India in this model but extends further west in the reanalysis. In the HadGEM2 model, the eastward circulation over northern India and over Bangladesh on heat-wave days does not return over central and southern India; an analysis over a larger domain would be required to discern whether the centre of the anticyclone has been displaced further south in the HadGEM2 model. Two models (CMCC-CESM and IPSL) associate heat waves with profoundly different atmospheric circulation patterns than observed and so should not be relied upon for projections of a change in heat-wave

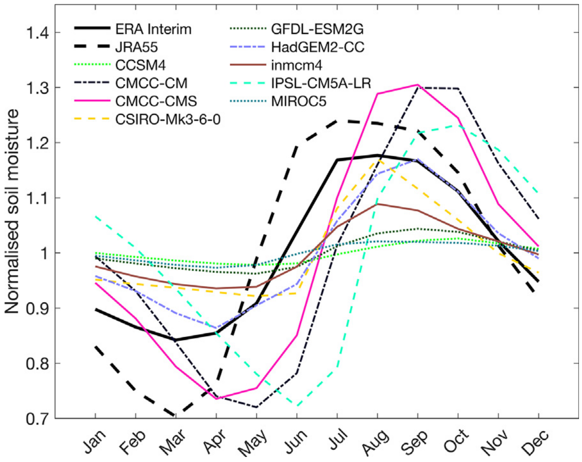


Fig. 10. Evaluation of the annual cycle of monthly soil moisture (m^3m^{-2}), averaged over Bangladesh, in the CMIP5 models compared with ERA Interim and JRA55 reanalyses. Monthly data were normalised by the corresponding annual average for comparison on a common scale.

frequency in the future. As observed, all the models associated heat waves with below-normal precipitation for this time of year, indicated by a decline in accumulated rainfall anomalies over the 30 days preceding a heat wave (Fig. 8). However, the strength of this relationship is underestimated in most models, and is particularly weak in the GFDL and IPSL models.

The further into the future projections are made, the greater the importance of model errors relating to any trends, and in this area the models perform well. Average temperature during the heat-wave season in Bangladesh did not change between 1979 and 2005 (-0.006 ± 0.029 °C yr^{-1} according to APHRODITE data) (Fig. 1g), a result confirmed using ERA Interim and JRA55 data (not shown). Seven of the eleven models analysed simulated some warming from 1979 to 2005, while three indicated slight cooling. Consistent with observations, however, none of these trends were statistically significant. The observed trend in annual frequency of heat-wave days was similarly insignificant over the study period, and is captured by all models (Fig. 4d).

The difference between Fig. 4c and d emphasizes the importance of selecting the appropriate statistical method for evaluating the models, particularly when calculating trends in extreme events. The overall conclusion, that trends in heat-wave frequency are insignificant for most models, remains unchanged between the most commonly-used parametric ordinary least squares (OLS) methodology and the more appropriate rank methodology with errors calculated non-parametrically using Kendall's *t*-statistic (see Section 2.3). However, both the magnitudes of the trends and the error bars differ substantially, particularly for a few models (GFDL, IPSL and MPI). Evidently, failing to select the appropriate statistical methodology could lead to unwarranted over- or under-confidence in the models.

(ii) Changes in the timing of heat waves during the year

The potential for the models to be informative about future changes in heat-wave seasonality in Bangladesh is compromised by problems regarding the modelled seasonality of temperature and heat-wave frequency in the recent past. While the annual cycle of minimum temperature is relatively well represented, all the models have a delay in the timing of the hottest maximum temperatures in the year (Fig. 2 and Fig. 3). Consequently, most models exhibit a delayed end to the heat-wave season. This lagged annual cycle appears to be caused by a late onset of the monsoon in the models (Fig. 10 and Fig. 11).

Given the failure of the models to capture the seasonality of temperature and heat-wave frequency, as well as its underlying physical drivers, it is unlikely that they could be informative about future changes in heat-wave seasonality without substantial further analysis. Such an analysis should focus on understanding the reasons for the delayed onset of the monsoon in Bangladesh in the models, and assessing how the timing of monsoon onset could be affected by climate change based on physics (Dessai et al., 2018).

(iii) Changes in interannual heat-wave variability

Despite the importance of interannual variability in climate risk management, this timescale has largely been neglected in studies evaluating models on the local scales that are relevant to practical decision-making (Laprise et al. 2008). The poor correlation found between biases in average temperature and interannual temperature variability ($r = 0.19$, Fig. 1f) suggests that a simple evaluation of climatological temperature is insufficient to account for errors in modelled interannual variability. Biases in average seasonal mean temperature ranged from negative to positive (Fig. 1b), but all models simulated higher interannual standard deviation than observed (Fig. 1e). This result is consistent with the weak relationship between model performance regarding the climatology and interannual variability over the tropics (Gleckler et al., 2008).

Projections of a change in interannual heat-wave variability should be interpreted with caution, because Fig. 12 suggests that some of its driving physical mechanisms are too weakly represented. Observed seasonal total precipitation is negatively correlated

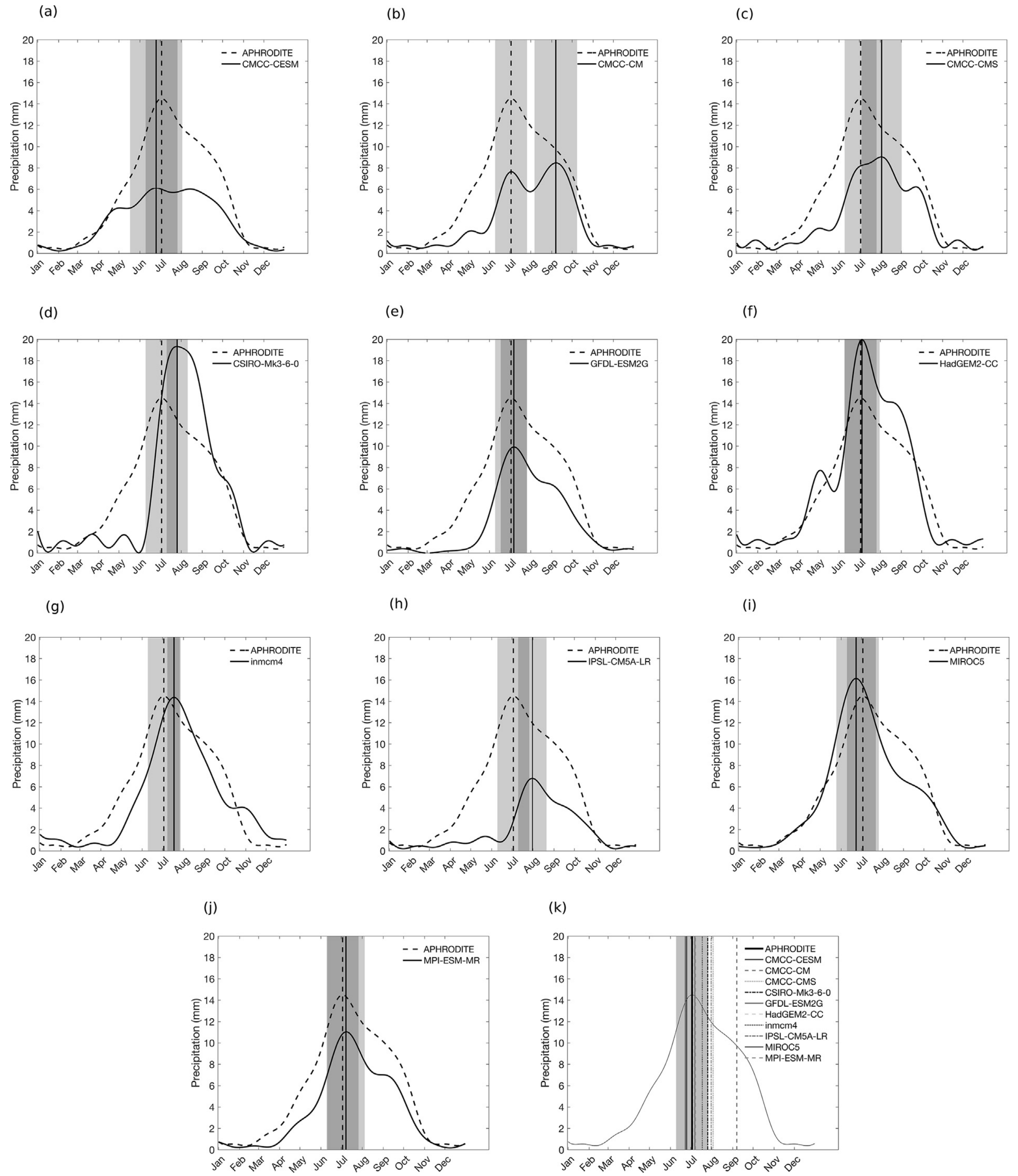


Fig. 11. (a)-(j) Smoothed daily climatologies of Bangladesh precipitation in APHRODITE (dashed curves) and in the CMIP5 models (bold curves), for the period 1979–2005. The vertical bold (dashed) line indicates the model (observed) timing of climatological peak precipitation during the year. The grey vertical bands represent the interannual variability in the timing of peak precipitation. (k) is a summary of panels (a)-(j): daily climatology of observed precipitation (thin black curve) and its peak timing (thick black vertical line), interannual variability of observed peak rainfall timing (grey vertical band) and timing of climatological peak precipitation in all the models (coloured vertical lines).

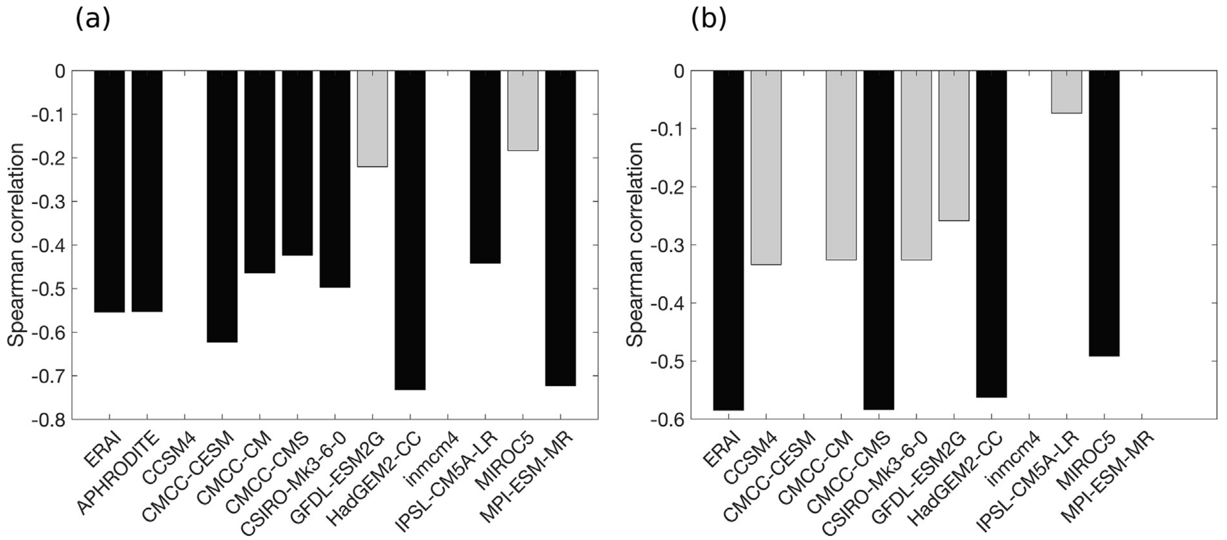


Fig. 12. Spearman correlation between the number of heat-wave days occurring each year between April and June, and total (a) precipitation and (b) soil moisture in the same season, over the period 1979–2005. Black (grey) bars indicate statistically significant (insignificant) correlations at the 5% significance level.

with heat-wave frequency ($r = -0.55$), but the strength of the correlation is underestimated by most models and overestimated by three (Fig. 12a). Similarly, low seasonal soil moisture is associated with a higher frequency of heat-wave days in ERA Interim ($r = -0.58$), but this relationship is underrepresented in all models, with only CMCC-CMS, HadGEM2 and MIROC5 coming close to the observed correlation. Further work would be needed to investigate the causes of these biases to enable a critical interpretation of future projections.

5. Conclusions

Targeted model evaluations highlight where the models should not be relied upon, but they can also point to some decisions that could be informed by a considered interpretation of the model projections. The lack of trend in temperature and heat-wave frequency in Bangladesh is well-captured by the models and therefore, for those models which capture the circulation pattern associated with heat waves (all but two), projections for average heat-wave frequency could potentially be informative. However, the weak connection between heat waves and reduced rainfall means that changes in pre-monsoon season precipitation (and the reasons for that change) should also be considered. Model biases may not remain constant over time; in particular, although the models have captured the lack of trend observed in recent decades, plans should remain flexible to respond to trends that may emerge in the future. Changes in the timing of heat waves during the year cannot be inferred from the model projections without a fuller understanding of the reasons for the delayed annual rainfall cycle in the models. Instead, periodic analyses of seasonality (for instance every decade), combined with investments in seasonal forecasts and early warning systems, could be used to ensure that heat action plans continue to provide protection throughout the hot season. Significant biases in heat-wave variability were noted and, though the physical drivers of this variability were broadly captured, the strength of these relationships were underestimated in most models. Inferences about future changes in heat-wave variability could be made by exploring modelled changes in interannual rainfall and soil moisture variability, and the physical mechanisms behind them.

However, these tailored evaluations can raise more questions about the interpretation of future projections than they address. For example, do the best-performing models project similar changes, or is the range of projections unaltered by eliminating the worst models? How does one incorporate insights from model evaluations into tailored climate services, e.g. through better-communicated statements about confidence in projected changes, or by highlighting the range of physically-plausible changes based on an understanding of how climate change could affect the processes driving important climate events and their variability on decision-relevant timescales? These questions are beyond the scope of this study. However, we note that one purpose of conducting targeted model evaluations is not to narrow the range of future projections, but to inform judgments about the appropriate level of confidence to place in projected changes from a position of scientific understanding, given that no verifiable measure of uncertainty in future projections is possible. Introducing greater subjectivity may well lead to lower, though (we argue) more realistic, levels of confidence in projected changes. Sound decision-making under uncertainty requires an understanding of the range of plausible outcomes, not just the best-guess scenario. Diversity in projections can thus be a benefit rather than a hindrance to effective adaptation, by encouraging more robust adaptation strategies in the face of greater uncertainty (Hallegatte, 2009).

The wider message of this study, which applies beyond the particular case of heat waves in Bangladesh, is that the appropriate use of climate models to guide climate adaptation necessitates a more nuanced evaluation of their performance over the historical record than is often undertaken. Restricting model evaluations to the most basic of variables (most commonly climatological average temperature or precipitation) sacrifices the opportunity to discern where models may provide valuable insights and where they could be misleading when applied to practical situations. However, the question of how to interpret the results looms large. In some cases, it is obvious: where the physical drivers of important climatic events on the timescale of interest are not captured, it makes no sense to look at a model's future projections or to apply bias correction procedures to account for climatological errors. More often, the path forward is ambiguous: models deviate from observations in ways that limit their utility to inform practical decision-making, but not by so much that they necessarily have no useful insights. In these cases, models need not be rejected completely, but nor should they be followed blindly. Where the physical drivers are broadly captured but biases exist, it could be informative to examine projected changes alongside a consideration of how the drivers themselves may change and the physical reasoning behind these changes. Thus, the utility of models for guiding climate adaptation rests on a nuanced understanding of their strengths and weaknesses, supported by knowledge about the scientific drivers of projected changes. This approach necessitates an element of subjectivity and expert judgment, since a thorough evaluation cannot be simplified to a one-size-fits all approach.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the Earth Institute at Columbia University, the UK Fulbright Commission and by ACToday, the first of Columbia University's World Projects. ÁGM was also partially supported by the NOAA award NA18OAR4310275.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crm.2020.100213.

References

Amin, R., Zhang, J., Yang, M., 2015. Effects of climate change on the yield and cropping area of major food crops: A case of Bangladesh. Sustainability 7 (1), 898–915. https://doi.org/10.3390/su7010898.
Baccini, M., et al., 2008. Heat effects on mortality in 15 European Cities. Epidemiology 19 (5), 711–719.
Baethgen, W.E., 2010. Climate risk management for adaptation to climate variability and change. Crop Sci. 50, p.S-70.
Baethgen, W.E., Goddard, L., 2013. Latin American perspectives on adaptation of agricultural systems to climate variability and change. In: In: Hillel, D., Rosenzweig, C. (Eds.), Handbook of Climate Change and Agroecosystems: Global and Regional Aspects and Implications. ICP Series on Climate Change Impacts, Adaptation, and Mitigation, vol. 2 Imperial College Press.
Bandyopadhyay, A., et al., 2018. Comparison of various re-analyses gridded data with observed data from meteorological stations over India. In: 13th International Conference on Hydroinformatics, pp. 190–198.
Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature 525 (7567), 47–55.
Box, G.E.P., 1979. Robustness in the Strategy of Scientific Model Building. *Robustness in Statistics*. Elsevier 201–236.
Cabot-Venton, C., 2013. Value for Money of Multi-year Approaches. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/204946/Cost_Effectiveness_of_Multi-year_Humanitarian_Funding_-_Final_Report.pdf [accessed 29/10/2019].
CME Group, 2007. Alternative Investments: Weather Futures and Options. Retrieved from http://www.cmegroup.com/trading/weather/files/Weather_fact_card.pdf [accessed 29/10/2019].
Conover, W.J., 1999. Practical Nonparametric Statistics, 3rd ed. John Wiley Sons Inc.
Coughlan de Perez, E., et al., 2014. Science to prevent disasters. Nat. Geosci. 7 (2), 78–79. https://doi.org/10.1038/ngeo2081.
Coughlan de Perez, E., et al., 2017. Should seasonal rainfall forecasts be used for flood preparedness? Hydrol. Earth Syst. Sci. 21, 4517–4524.
Dessai, S., et al., 2018. Building narratives to characterise uncertainty in regional climate change through expert elicitation. Environ. Res. Lett. 13 (7), 074005.
Doblas-Reyes, F.J., et al., 2013. Seasonal climate predictability and forecasting: status and prospects. Wiley Interdiscip. Rev. Clim. Change 4 (4), 245–268.
Dufresne, J.-L., et al., 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Clim. Dyn. 40 (9–10), 2123–2165.
Dunne, J.P., et al., 2013. GFDL's ESM2 global coupled climate-carbon earth system models. Part II: carbon system formulation and baseline simulation characteristics. J. Clim. 26 (7), 2247–2267.
Fischer, E.M., Beyerle, U., Knutti, R., 2013. Robust spatially aggregated projections of climate extremes. Nat. Clim. Change 3 (12), 1033–1038.
Fogli, P.G. et al., 2009. INGV-CMCC Carbon (ICC): A Carbon Cycle Earth System Model. CMCC Research Paper No. 61.
Frigg, R., et al., 2014. Laplace's Demon and the adventures of his apprentices. Philos. Sci. 81, 31–59.
Frigg, R., Smith, L.A., Stainforth, D.A., 2015. An assessment of the foundational assumptions in high-resolution climate projections: the case of UKCP09. Synthese.
Gent, P.R., et al., 2011. The community climate system Model Version 4. J. Clim. 24.
GHHIN: Global Heat Health Information Network, 2019. First Global Forum on Heat and Health. 17-20 Dec 2018, Hong Kong. https://ghhin.org [accessed April 30 2019].
Giorgetta, M.A., et al., 2013. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. J. Adv. Model. Earth Syst. 5 (3), 572–597.
Gleckler, P.J., Taylor, K.E., Doutriaux, C., 2008. Performance metrics for climate models. J. Geophys. Res. 113 (D6), D06104.
Goddard, L., et al., 2001. Current approaches to seasonal to interannual climate predictions. Int. J. Climatol. 21 (9), 1111–1152.
Gonzalez, P.L.M., et al., 2014. Stratospheric ozone depletion: a key driver of recent precipitation trends in South Eastern South America. Clim. Dyn. 42 (7–8), 1775–1792.
Goosen, H., Hasan, T., Saha, S.K., Rezwana, N., Rahman, R., Assaduzzaman, M., Kabir, A., Dubois, G., Terwisscha van Scheltinga, C. 2018. Nationwide Climate Vulnerability Assessment in Bangladesh.
Gordon, H. et al., 2002. The CSIRO Mk3 Climate System Model. CSIRO Atmospheric Research Technical Paper No. 60.
Hacker, J.N., Belcher, S.E., Connell, R.K., 2005. Beating the heat: keeping UK building cool in a warming climate. UKCIP, Oxford. https://doi.org/10.1119/1.2339869.
Hall, J., 2007. Probabilistic climate scenarios may misrepresent uncertainty and lead to bad adaptation decisions. Hydrol. Process. 21 (8), 1127–1129.
Hallegatte, S. et al., 2012. Investment Decision Making Under Deep Uncertainty: Application to Climate Change. Policy Research Working Paper, (6193), p.41.
Hallegatte, S., 2009. Strategies to adapt to an uncertain climate change. Global Environ. Change 19 (2), 240–247.

Hansen, J., et al., 2006. Translating climate forecasts into agricultural terms: advances and challenges. Clim. Res. 33 (1), 27–41.
Hansen, J.W., et al., 2011. 2 In: Review of seasonal climate forecasting for agriculture in Sub-Saharan Africa Experimental Agriculture. Cambridge University Press, pp. 205–240. https://doi.org/10.1017/S0014479710000876.
Hess, J.J., Ebi, K.L., 2016. Iterative management of heat early warning systems in a changing climate. Ann. N. Y. Acad. Sci. 1382 (1), 21–30.
Hewitson, B., et al., 2017. Climate information websites: an evolving landscape. Wiley Interdiscip. Rev. Clim. Change 8 (5), e470.
Hewitson, B.C., et al., 2014. Interrogating empirical-statistical downscaling. Clim. Change 122 (4), 539–554.
Hossain, A., Teixeira da Silva, J.A., 2013. Wheat production in Bangladesh: its future in the light of global warming. AoB Plants 5https://doi.org/10.1093/aobpla/pls042. pls042–pls042.
Jones, L., et al., 2015. Ensuring climate information guides long-term development A shortfall in knowledge and data. Nat. Clim. Change 5, 812–814.
Kandlikar, M., Risbey, J., Dessai, S., 2005. Representing and communicating deep uncertainty in climate-change assessments. C.R. Geosci. 337 (4), 443–455.
Kharin, V.V., et al., 2013. Changes in temperature and precipitation extremes in the CMIP5 ensemble. Clim. Change 119 (2), 345–357.
Knowlton, K., et al., 2014. Development and implementation of South Asia's first heat-health action plan in Ahmedabad (Gujarat, India). Int. J. Environ. Res. Public Health 11 (4), 3473–3492.
Knutson, T.R., et al., 2010. Tropical cyclones and climate change. Nat. Geosci. 3.
Knutti, R., et al., 2010. Challenges in combining projections from multiple climate models. J. Clim. 23 (10), 2739–2758.
Kobayashi, S., et al., 2015. The JRA-55 reanalysis: general specifications and basic characteristics. J. Meteorol. Soc. Jpn. Ser. II 93 (1), 5–48.
Laprise, R., et al., 2008. Challenging some tenets of Regional Climate Modelling. Meteorol. Atmos. Phys. 100 (1–4), 3–22.
Lee, M., Nordio, F., Zanobetti, A., Kinney, P., Vautard, R., Schwartz, J., 2014. Acclimatization across space and time in the effects of temperature on mortality: a time-series analysis. Environ. Health 13 (1), 89.
Martin, M., et al., 2011. The HadGEM2 family of Met Office Unified Model climate configurations. Geosci. Model Dev. 4, 723–757.
Miralles, D.G., et al., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7 (5), 345–349.
Muñoz, Á.G., et al., 2017. A weather-type-based cross-time-scale diagnostic framework for coupled circulation models. J. Clim. 30.
Nissan, H., et al., 2017. Defining and predicting heat waves in Bangladesh. J. Appl. Meteorol. Climatol. 56 (10), 2653–2670.
Nissan, H., et al., 2019. On the use and misuse of climate change projections in international development. WIREs Clim. Change 10 (e579).
Nissan, H., Conway, D., 2018. From advocacy to action: projecting the health impacts of climate change. PLoS Med. 15 (7).
Perkins, S.E., Argüeso, D., White, C.J., 2015. Relationships between climate variability, soil moisture, and Australian heatwaves. J. Geophys. Res.: Atmos. 120 (16), 8144–8164.
Prakash, S., et al., 2015. Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season. Int. J. Climatol. John Wiley & Sons Ltd 35 (9), 2326–2338. https://doi.org/10.1002/joc.4129.
Public Health England, NHS England, Local Government Association, UK Met Office, 2015. Heatwave plan for England.
Quesada, B., et al., 2012. Asymmetric European summer heat predictability from wet and dry southern winters and springs. Nature Climate Change.
Ramanathan, R., et al., 1997. Short-run forecasts of electricity loads and peaks. Int. J. Forecast. 13, 161–174.
Ranger, N., Garbett-Shiels, S.-L., 2012. Accounting for a changing and uncertain climate in planning and policymaking today: lessons for developing countries. Clim. Develop. 4 (February 2013), 288–300.
Red Cross Red Crescent Climate Centre, 2017. Bridging Science , Policy and Practice. Report of the International Conference on Climate Risk Management. Pre-Scoping Meeting for the IPCC Sixth Assessment Report. 5-7 April 2017. Nairobi. Retrieved from https://www.climatecentre.org/downloads/files/RCCC%20IPCC%20Nairobi%20Report%202018-4%20V5.pdf [accessed 29/10/2019].
Roulston, M.S., et al., 2003. Using medium-range weather forecasts to improve the value of wind energy production. Renew. Energy 28 (4), 585–602.
Ruane, A.C., Major, D.C., Yu, W.H., Alam, M., Hussain, S.G., Khan, A.S., Hassan, A., Hossain, B.M.T. Al, R., Goldberg, ... Rosenzweig, C., 2013. Multi-factor impact analysis of agricultural production in Bangladesh with climate change. Glob. Environ. Change 23 (1), 338–350. https://doi.org/10.1016/j.gloenvcha.2012.09.001.
Sanderson, M., Rafique, A., 2009. Pre-monsoon rainfall and its variability in Bangladesh: a trend surface analysis. Hydrol. Sci. Bull. 24 (3), 277–287.
Scoccimarro, E., et al., 2011. Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model. J. Clim. 24 (16), 4368–4384.
Smith, L.A., 2002. What might we learn from climate forecasts? Proc. Natl. Acad. Sci. U.S.A. 99 (Suppl. 1), 2487–2492.
Smith, L.A., Petersen, A.C., 2014. Variations on reliability: connecting climate predictions to climate policy. In: Boumans, M., Hon, G., Petersen, A.C. (Eds.), Error and Uncertainty in Scientific Practice. Pickering & Chatto, London, pp. 137–156.
Smith, L.A., Stern, N., 2011. Uncertainty in science and its role in climate policy. Philos. Trans. A, Math. Phys. Eng. Sci. 369 (1956), 4818–4841.
Spiegelhalter, D.J., Riesch, H., 2011. Don't know, can't know: embracing deeper uncertainties when analysing risks. Philos. Trans. R. Soc. A 369, 4730–4750.
Stephens, G.L., et al., 2010. Dreary state of precipitation in global models. J. Geophys. Res.: Atmos. 115 (D24).
Steynor, A., et al., 2016. Co-exploratory climate risk workshops: experiences from urban Africa. Clim. Risk Manage. 13, 95–102.
Taylor, J.W., Buizza, R., 2003. Using weather ensemble predictions in electricity demand forecasting. Int. J. Forecast. 19 (1), 57–70.
Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philos. Trans. A, Math. Phys. Eng. Sci. 365 (1857), 2053–2075.
Thompson, E., Frigg, R., Helgeson, C., 2016. Expert judgment for climate change adaptation. Philos. Sci. 83 (5), 1110–1121. https://doi.org/10.1086/687942.
Thompson, E.L., Smith, L.A., 2019. Escape from model-land. Econ.: The Open-Access, Open-Assessment J. 13 (2019–40), 1–15.
Tompkins, A., et al., 2018. Predicting climate impacts on health at sub-seasonal to seasonal timescales. In: Robertson, A., Vitart, F. (Eds.), The Gap Between Weather and Climate Forecasting: Sub-seasonal to Seasonal Prediction. Elsevier.
UNDP, 2005. Adptation policy frameworks for climate change. Developing Strategies, Policies and Measures. Cambridge University Press, Cambridge, United, Kingdom and New York, NY, USA.
van Aalst, M.K., Cannon, T., Burton, I., 2008. Community level adaptation to climate change: the potential role of participatory community risk assessment. Global Environ. Change 18 (1), 165–179.
Vincent, K. et al., 2014. Actual and Potential Weather and Climate Information Needs for Development Planning in Malawi: Results of a Future Climate for Africa Pilot Case Study. Retrieved from https://www.weadapt.org/sites/weadapt.org/files/legacy-new/placemarks/files/54cf919a07063malawi-report.pdf [accessed 29/10/2019].
Volodin, E.M., Dianskii, N.A., Gusev, A.V., 2010. Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. Izvestiya Atmos. Ocean. Phys. 46 (4), 414–431.
Volodin, E.M., et al., 2017. Simulation of the present-day climate with the climate model INMCM5. Clim. Dyn. 49, 3715–3734. https://doi.org/10.1007/s00382-017-3539-7.
Wade, S. et al., 2015. Developing H + + climate change scenarios for heat waves, droughts, floods, windstorms and cold snaps. Retrieved from https://www.theccc.org.uk/wp-content/uploads/2015/10/Met-Office-for-the-ASC-Developing-H-climate-change-scenarios-for-heatwaves-droughts-floods-windstorms-and-cold-snaps3.pdf [accessed 29/10/2019].
Watanabe, M., et al., 2010. Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. J. Clim. 23 (23), 6312–6335.
Weisheimer, A., Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. J. R. Soc. Interface 11 (96).
WHO-WMO Joint Office for Climate and Health, 2016. Improving health preparedness for extreme heat events in South Asia, Colombo, Sri Lanka. Retrieved from http://www.wmo.int/pages/prog/wcp/wcsp/rcofs/sascof/health/documents/1st-South-Asia-Climate-Services-Forum-for-Health.pdf [accessed. 29/10/2019].
Wood, L., Moriniere, L., 2013. Malawi climate change vulnerability assessment. USAID. https://doi.org/10.13140/RG.2.1.2965.6487.
Yatagai, A., et al., 2012. APHRODITE: constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges. Bull. Am. Meteorol. Soc. 93 (9), 1401–1415.