

Invited Review Paper

Review of approaches for selection and ensembling of GCMs

Komaragiri Srinivasa Raju and Dasika Nagesh Kumar

ABSTRACT

Global climate models (GCMs) are developed to simulate past climate and produce projections of climate in future. Their roles in ascertaining regional issues and possible solutions in water resources planning/management are appreciated across the world. However, there is substantial uncertainty in the future projections of GCM(s) for practical and regional implementation which has attracted criticism by the water resources planners. The present paper aims at reviewing the selection of GCMs and focusing on performance indicators, ranking of GCMs and ensembling of GCMs and covering different geographical regions. In addition, this paper also proposes future research directions.

Key words | ensemble, global climate models, indicators, ranking

Komaragiri Srinivasa Raju
Department of Civil Engineering,
BITS Pilani,
Hyderabad Campus, Hyderabad 500 078,
India

Dasika Nagesh Kumar (corresponding author)
Department of Civil Engineering,
Indian Institute of Science,
Bangalore 560 012,
India;
Divecha Centre for Climate Change,
Indian Institute of Science,
Bangalore 560 012,
India
and
Interdisciplinary Centre for Water Research,
Indian Institute of Science,
Bangalore 560 012,
India
E-mail: nagesh@iisc.ac.in

INTRODUCTION

Global climate models (GCMs) are numerical models describing natural mechanisms in the atmosphere, land surface and ocean. GCMs represent the climate system adopting a 3D grid with horizontal coarse resolution of 250–600 km over the world, and 10–20 vertical layers in the atmosphere as well as around 30 layers in the oceans. They are developed to indicate atmospheric physics, dynamics and to simulate past climate for analysing future climate changes. GCMs follow conservation laws (momentum, mass, energy, moisture), fluid dynamics, equation of state and more. Some of the parameters and boundary conditions considered in GCMs are rotation speed of the Earth, thermodynamic and radiation constants of atmospheric gases and clouds, surface elevation, total mass of the atmosphere and its composition, soil type and surface albedo (Schmidt *et al.*

2006). However, lack of complete information about atmospheric processes, approximations during numerical modelling, spatio-temporal scales, coarser or finer resolution, different feedback mechanisms (cloud and solar radiation, greenhouse gases, aerosols, natural and anthropogenic sources, ocean circulation, water vapour and warming, ice and snow albedo), and different perspectives (physical parameterisations, initialisations, and model structures) are the causes of uncertainties that lead to either overestimation or underestimation of values of the considered climate variable, as compared to the observed variables. This inadvertently results in different outcomes for different GCMs for the same forcing (Sood & Smakhtin 2015; Jain *et al.* 2019).

Mandal *et al.* (2019) identified six uncertainty causes, namely, (i) selection process of GCMs, (ii) choice of GCMs, (iii) emission scenarios, (iv) downscaling models, (v) hydrologic model parameters and (vi) model structures. Liu *et al.* (2012), as part of their studies on headwater catchment in China, demonstrated that uncertainty levels

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/wcc.2020.128

associated with GCM outputs were crucial for assessing their impact on climate change. Bosshard *et al.* (2013) found that GCMs were the dominant uncertainty source. Xuan *et al.* (2017), based on their studies in Zhejiang Province, in South-east China, concluded that most GCMs were not able to predict the observed spatial patterns, due to insufficient resolution. Benedict *et al.* (2019) analysed spatial resolution of GCMs and global hydrological models for the Rhine and Mississippi basins. Higher resolution GCMs yielded improved precipitation budget for the Rhine whereas no substantial improvement was found for the Mississippi. Above all, lack of precise observed data to assess the simulating ability of GCMs is another concern (<https://www.tau.ac.il/~colin/courses/CChange/CC5.pdf>; <https://www.climate.gov/maps-data/primer/climate-models>; https://www.ipcc-data.org/guidelines/pages/gcm_guide.html). Keeping this in view, Tian *et al.* (2015) suggested uncertainty assessment along the whole climate modelling chain.

Hughes *et al.* (2014) discussed the cause of inherent errors and uncertainties occurring due to simplification of highly complex atmospheric physics in GCMs. They found multi-model ensemble (MME) was a good fit for the situation in comparison to individual GCMs mainly due to compensation of individual errors. Ahmed *et al.* (2020), as part of their study over Pakistan, presented similar views. Yan *et al.* (2015), as part of their study on Xinjiang Basin, China, suggested an MME to minimise the biases and uncertainties of future climate simulations which is available as supplementary data. Tebaldi & Knutti (2007) observed 'combining models generally increases the skill, reliability and consistency of model forecasts'. Hughes *et al.* (2011) suggested all GCMs in MME for effective representation of climate change whereas Basharin *et al.* (2016) suggested choosing GCMs for MME that described the present climate more precisely, which would facilitate decision-makers in using the predictions effectively. However, Bannister *et al.* (2017) cautioned 'MME does not consider the relative strengths and weaknesses of each model as an ensemble invariably hides the substantial variations between the individual models'. They warned 'If the GCMs collectively misrepresent some component of the forcing or partially cancel each other out, then the future natural variability in an ensemble will be inherently suppressed'. Complimentarily, Raghavan *et al.* (2018), as part of their studies on South East Asia, expressed 'although the ensemble mean of the models

is a better representation of the observed climate, the spread among the individual models is large'. Another problem is that some GCMs partly work with the same code; in other words, similar components of ocean and atmosphere results in identical forecasts (Raju & Nagesh Kumar 2016) and this may mislead the process of MME. More information on uncertainty and MMEs is available in Knutti *et al.* (2010), Najafi *et al.* (2011), Weiland *et al.* (2012), Knutti & Sedláček (2013), Miao *et al.* (2014), Northrop & Chandler (2014), Lutz *et al.* (2016), Song & Chung (2016), You *et al.* (2018), Salman *et al.* (2018) and Jobst *et al.* (2018). Duan *et al.* (2019) carried out an extensive review on multi-model analysis that provides guidelines for robust climate change research. In totality, researchers question the reliability and simulating ability of GCMs due to the uncertainties involved at every level of modelling, as explained earlier (Wilby & Harris 2006; Mujumdar & Nagesh Kumar 2012).

Basharin *et al.* (2016) observed significant improvements in Coupled Model Intercomparison Project 5 (CMIP5)-GCM simulations as compared to its previous generation of Coupled Model Intercomparison Project 3 (CMIP3)-GCMs in terms of 'effect of aerosols, the interaction at the land-ice boundary, stratosphere-troposphere interactions, the carbon cycle, runoff, and biochemical interactions between ecosystems and other processes'. Interestingly, the Coupled Model Intercomparison Project 6 (CMIP6) represents a considerable expansion over CMIP5, in terms of (a) 100 distinct climate models from 40 different modelling groups, (b) eight future scenarios representing Shared Socio-economic Pathways (SSPs) and (c) different experiments conducted. CMIP6 models have a remarkably higher climate sensitivity than those of CMIP5 (Tokarska *et al.* 2020; <https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained>). However, Jain *et al.* (2019) were critical about GCMs' development and suggested focusing on the improvement of GCMs that should perform relatively efficiently over many regions of the world. They discouraged expanding GCMs that may not compete strongly in continental/regional scales. Similar views regarding the number of GCMs are expressed by Allan Hollander (<http://climate.calcommons.org/article/why-so-many-climate-models>). Details about developed centre/institution and name of global climate models are available at: CMIP3 (http://www.ipcc-data.org/sim/gcm_

monthly/SRES_AR4/index.html#Acknowledge), CMIP5 (<https://pcmdi.llnl.gov/mips/cmip5/availability.html>) and CMIP6 (<https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6>).

Pierce *et al.* (2009) raised questions related to the effect of choosing different GCMs for regional climate studies, on downscaled outcomes and the relevance of a blueprint for selecting GCMs (Legates & McCabe 1999; Mujumdar & Nagesh Kumar 2012; Bhattacharjee & Zaitchik 2015; Raju & Nagesh Kumar 2018). Hargreaves & Annan (2014) discussed the origin of GCMs in detail, interpretation and evaluation of ensemble and intricacies in climate modelling. They raised concern about the gap between actual and potential performances and limited benefits irrespective of substantial investment. Sikder *et al.* (2016) opined that GCMs were yet to be improved for prime-time operationalisation of multiscale water management decisions. Knutti (2008) explained the necessity of evaluating GCMs and opined that skilful simulations of the past might not produce skilful predictions of the future. However, Knutti (2008) also cautioned that lack of past skill in simulation might also translate to lack of skill in future predictions. Hence, it is important to analyse each chosen GCM(s) for its efficacy to assess how realistic GCM results are in comparison with observed records of climate (<https://www.climate.gov/maps-data/primer/climate-models>). Otieno & Anyah (2013) suggested a cautious approach while choosing GCMs. Zhang & Yan (2016) recommended assessing the effectiveness of GCM to simulate conservation-based climate zones. Cook *et al.* (2017) opined that selecting an appropriate GCM for the intended application and its performance, ascertaining differences in temporal scale and spatial scale outputs, and inferring results for engineering design are essential. Hence, performance evaluation of GCMs for selection of the best GCM or suitable GCM(s) is essential and leads to confidence of the policymakers and planners for using them for impact assessment studies and other purposes (Perez *et al.* 2014; Aloysius *et al.* 2016).

The present paper is aimed at reviewing the selection of GCMs, focusing on performance indicators, ranking of GCMs, and related ensembling of GCMs covering different geographical regions. In addition, MME of GCMs without explicit evaluation covering different geographical regions is also presented. Even though these topics are separately discussed, they

are interrelated to a considerable extent. The period chosen is from 2006 to 2020 with the selected literature review focusing on CMIP5 and CMIP3 repositories. Efforts are made to discuss only representative studies to effectively focus on the theme.

The present paper describes the general structure for selection of GCMs, literature review on performance measures, ranking and ensembling of GCMs covering different geographical regions, MME without explicit evaluation of GCMs, observation and discussions, followed by summary and conclusions. Acronyms and performance measures used in this paper are presented in Tables A1 and A2 of the Appendix, respectively, which is available as supplementary data.

GENERAL STRUCTURE FOR SELECTION OF THE BEST GCM OR AN ENSEMBLE OF GCMs

After reviewing more than 200 papers on climate modelling, the modelling procedure observed for selection of the best/suitable/ensembling of GCMs is found as follows (irrespective of geographical region):

1. Selection of appropriate climate variables for the chosen region.
2. Selection of appropriate X number of GCMs from either CMIP3 or CMIP5 repository Y [$X \leq Y$] for the chosen region.
3. Collection of observed and simulated values of X GCMs for the chosen climate variable.
4. Identification of evaluation criteria/metrics/indicators to ascertain the simulating ability of X GCMs with observed data.
5. Comparison of outputs of X GCMs with observed data for historic time period in terms of chosen evaluation criteria.
6. Selection of suitable GCM(s) that represent the climate system from X based on (5).
7. Formulate ensemble mechanism if required either with subset of suitable GCMs (from (6)) or with full GCM ensemble.

PERFORMANCE INDICATORS, RANKING OF GCMs AND RELATED ENSEMBLING OF GCMs

Performance indicator is a metric to measure how efficiently a GCM simulates observed data (Gómez-Navarro *et al.* 2012).

Meaningful indicators are required to evaluate GCMs (Tebaldi & Knutti 2007; Knutti *et al.* 2010; Raju & Nagesh Kumar 2014b). Gleckler *et al.* (2008) suggested developing metrics to characterise the GCM performance. This may facilitate identification of optimal subsets of GCMs for various applications. Guilyardi *et al.* (2009) viewed that metrics 'should be concise, physically informative, societally relevant & easy to understand, compute and compare'. Gu *et al.* (2015) expected that metrics should (a) have the ability to distinguish between good performance and poor performance, and (b) be computationally efficient, bounded and dimensionless. Similar views were presented by Moise & Delage (2011) and McMahon *et al.* (2015). Fahimi *et al.* (2017) summarised various evaluation indicators employed by various researchers in hydrological modelling. The following sections present studies over different geographical areas.

China

Gu *et al.* (2015) evaluated 27 CMIP5-GCMs for seasonal and annual surface air temperature and precipitation for five climate-based regions of China. Mielke measures, *M1*, *M2* and *M3* were explored to evaluate GCMs. The GCMs, BCC-CSM1.1(m), CanESM2, CMCC-CMS and CMCC-CM were preferred for precipitation, while the GCMs, CMCC, BCC-CSM1.1(m), IPSL-CM5A-MR, NCAR and MPI were preferred for temperature. Fine resolution improved the GCM simulating ability especially for temperature. Bao & Feng (2016) conducted a similar study using 16 CMIP5-GCMs for the Yellow River and Yangtze River basins in China for precipitation, evaporation and water vapour transport. Most GCMs had a tendency to overestimate precipitation in the Yellow River basin. However, simulating capability of GCMs was satisfactory in the Yangtze River basin. Jiang *et al.* (2016) evaluated 77 GCMs from Third Assessment Report (TAR), CMIP3 and CMIP5 in simulating the mean state and year-to-year climate variability over China and the East Asian monsoon. Simulating ability of GCMs improved from TAR to CMIP3 for both temperature and precipitation. However, simulating ability remained stable for temperature and decreased for precipitation for CMIP5 from CMIP3.

Bannister *et al.* (2017) evaluated 47 CMIP5-GCMs for Sichuan basin in China for mean temperature, minimum

temperature and maximum temperature with performance indicator skill score (SS). They also formulated MME with all GCMs with equal weightage. MIROC4h, IPSL-CM5A-MR, CESM1(FASTCHEM), MPI-ESM-MR and MIROC5 were the top five GCMs for mean temperature; MIROC4h, GISS-E2-R-CC and GISS-E2-H-CC in the case of minimum temperature and CESM1(BGC), CCSM4 and CESM1 (FASTCHEM) for maximum temperature. However, SS of MME was substantially lower than SS of the top GCMs for mean temperature, maximum temperature and minimum temperature. They concluded that determining the best overall GCM was difficult. Wu *et al.* (2018) evaluated six CMIP5-GCMs for the Huaihe River basin in China for variable precipitation. Indicators used were standard deviation (σ), deterministic coefficient (*DC*), correlation coefficient (*CC*), relative error (*RE*) and root mean square error (*RMSE*). MME was performed by arithmetic mean (MME-AM) and backpropagation neural network (MME-BP). The relative order of simulation ability of GCMs on precipitation process was BNU-ESM, MME-AM, CNRM-CM5, MRI-CGCM3, MIROC-ESM, BCC-CSM1.1 and MPI-ESM-LR.

He *et al.* (2019) evaluated nine CMIP5-GCMs, BCC-CSM1.1(m), CMCC-CMS, CNRM-CM5, FGOALS-g2, GFDL-ESM2G, INM-CM4, IPSL-CM5A-MR, HadGEM2-AO and MPI-ESM-MR over China for temperature. *RMSE* and *CC* were employed as indicators to evaluate GCMs. GCMs, CMCC-CMS and MPI-ESM-MR, had higher capability in simulating spatial pattern and its decadal change of climate zones. An MME with seven GCMs (except FGOALS-g2 and INM-CM4) was formulated which was found preferable to that of any single GCM. They preferred MME to reduce uncertainty of GCMs.

Chhin & Yoden (2018) evaluated 43 GCMs with 36 performance metrics for the Indo-China region for precipitation. They provided different patterns of ensemble average. Optimal ensemble subsets significantly improved the monthly precipitation, as compared to full model ensemble as well as best single GCM during the historical period.

Indian subcontinent

Jena *et al.* (2016) evaluated 20 CMIP5-GCMs for the Indian summer monsoon and CCSM4, CESM1(CAM5), GFDL-CM3 and GFDL-ESM2G were the preferred GCMs. Sarthi

et al. (2016) evaluated 34 CMIP5-GCMs for the Indian summer monsoon with indicators, Taylor diagram, *SS*, *CC* and *RMSE*. The GCMs, MPI-ESM-MR, CESM1(WACCM), CESM1(CAM5), CESM1(BGC), BCC-CSM1.1(m) and CCSM4, captured precipitation effectively. Jain *et al.* (2019) evaluated 28 CMIP5-GCMs for the Indian summer monsoon with indicator pattern correlation (*PC*). All GCMs simulated seasonal mean surface air temperatures well whereas performance was relatively poor for precipitation. MIROC4h was preferred as compared to the other GCMs. Raju & Nagesh Kumar (2014b) evaluated precipitation for 73 grid points for Indian conditions. Eleven GCMs of CMIP3 repository with indicators, *CC*, average absolute relative error (*AARE*), normalised root mean square error (*NRMSE*), absolute normalised mean bias error (*ANMBE*) and *SS* were considered. Weights of indicators were obtained by entropy method; equal weights of indicators were also used. PROMETHEE-2 (Preference Ranking Organisation METHod of Enrichment Evaluation) was employed to compute preference of GCMs. Ranking of GCMs for Cauvery, Godavari, Mahanadi and Krishna river basins was also performed. No single GCM was found suitable for any of the Indian conditions and river basins. An ensemble of HadGEM1, MPI-ECHAM4, HadCM3, BCCR-BCCM2.0, MIROC3 and GFDL2.0 was suggested for India; ensembles of GFDL2.0, CGCM2, GISS, HadCM3 for Cauvery river basin; GFDL2.0, MIROC3, BCCR, HadCM3, CGCM2, GFDL2.1 for Godavari river basin; GFDL2.0, HadCM3, GFDL2.1, CGCM2, ECHAM for Mahanadi river basin; and GFDL2.0, BCCR, CGCM2, MIROC3, GISS for Krishna river basin.

Raju & Nagesh Kumar (2015a) evaluated precipitation and temperature for 11 GCMs of CMIP3 repository for India as well as Krishna and Mahanadi basins with the indicator *SS*. Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was employed to prioritise GCMs. Ensemble of GFDL2.1, BCCR-BCCM2.0, MIROC3, HadCM3, INGV-ECHAM4 and GFDL2.0 was found preferable for India. The result of this study was contrary to the study of Raju & Nagesh Kumar (2014b) as the ensemble suggested was altogether different. This may be due to the additional variable (i.e., temperature) used, the chosen number of indicators and the chosen decision-making technique.

Anandhi & Nanjundiah (2015) evaluated 19 CMIP3-GCMs for daily precipitation for the Indian region consisting of six zones with indicator *SS* for three categories, annual, June–October and non-monsoon seasons (JFMAMND: January to May, November, December). No single GCM has been diagnosed as suitable for all categories and zones. Raju *et al.* (2017) evaluated 36 CMIP5-GCMs for India for maximum temperature and minimum temperature. Indicators chosen were *CC*, *NRMSE* and *SS*. Compromise programming (CP) technique was employed to prioritise GCMs; group decision-making for facilitating aggregate ranking for India. They also proposed MME, which they claimed simple and effective. Meher *et al.* (2017) evaluated 13 CMIP3-GCMs and 42 CMIP5 for the Western Himalayan region for precipitation. Signal-to-noise ratio, mean annual cycle, spatial patterns, trends and annual cycles of interannual variability were the employed indicators. CMIP3-GCMs, GFDL-CM2.1, GFDL-CM2.0 and MIROC3.2 (hires), and CMIP5-GCMs, INM-CM4, MIROC5 and CESM1(BGC), were graded as the most competent. Interestingly, some of the coarser resolution GCMs were found to have improved skills, compared to the finer resolution GCMs.

Panjwani *et al.* (2019) evaluated 12 CMIP5-GCMs for India for precipitation, minimum temperature and maximum temperature. They employed fuzzy analytic hierarchy process (FAHP) and reliability index to assess their ability. Indicators employed were agreement index (*AI*), *RMSE* and *CC*. FAHP was found suitable to rank GCMs. NorESM1-M for maximum temperature; MIROC5, GFDL-CM3, FIO-ESM and IPSL-CM5A-LR for minimum temperature; IPSL-CM5A-LR, GFDL-ESM2M, HadGEM2, MIROC5 and CSIRO for precipitation were found suitable. Pandey *et al.* (2019) evaluated 24 CMIP5-GCMs for Upper Narmada river basin (UNB), India. Six climate variables were chosen. Indicators were *SS*, *RMSE* and total index (*TI*). Three GCMs, MIROC5, CNRM-CM5 and MPI-ESM-LR, were suitable GCMs. Sreelatha & Raj (2020) evaluated average temperature for Telangana region, South India. Indicators chosen were *SS*, *CC*, normalised root mean square deviation (*NRMSD*), Nash–Sutcliffe efficiency (*NSE*), and absolute normalised mean bias deviation (*ANMBD*). CP and group decision-making were considered for ranking pattern. MIROC5, CNRM-CM5, ACCESS1.0 and BCC-CSM1.1(m) were found to be suitable GCMs.

Khan *et al.* (2018) evaluated 31 CMIP5-GCMs for the variables precipitation, minimum temperature and maximum temperature for Pakistan. The chosen main indicator was symmetrical uncertainty (*SU*). Six GCMs, MIROC5, HadGEM2-ES, HadGEM2-CC, CMCC-CM, CESM1(BGC) and ACCESS1.3, were the top ranked GCMs and they proposed the same for MME. Mahmood *et al.* (2018) used *CC* and *RMSE* to evaluate five CMIP5-GCMs for Jhelum River basin, Pakistan and India, for precipitation. GCMs were found to lack the capability of capturing variation in the precipitation pattern. Latif *et al.* (2018) evaluated 36 CMIP5-GCMs for the Indo-Pakistan subcontinent for precipitation. The indicator chosen was partial correlation (*PAC*). Three GCMs, HadGEM2-AO, CNRM-CM5, and CCSM4 were found relevant and expected relatively reliable future projections.

Ahmed *et al.* (2018) evaluated 20 CMIP5-GCMs for precipitation over Pakistan. Spatial indicators chosen were Goodman–Kruskal's lambda, Kling–Gupta efficiency, Map-curves, fractional skill score (*FSS*), spatial efficiency metric and Cramer's *V*. GFDL-ESM2G, GFDL-CM3, CESM1 (CAM5) and NorESM1-M were preferred and used for MME, performed with random forest (RF) and mean. RF-based MME was preferred. They advocated the use of an ensembling approach that would reduce uncertainties in climate projections. Ahmed *et al.* (2019) applied wavelet-based skill score (*WSS*), *SU* and *CP* for ranking 20 CMIP5-GCMs for precipitation and for minimum and maximum temperature over Pakistan. *SU* preferred CESM1(CAM5), HadGEM2-AO, NorESM1-M and HadGEM2-ES; *CP* preferred CESM1(CAM5), HadGEM2-AO, NorESM1-M and GFDL-CM3; *WSS* preferred CCSM4, CESM1(CAM5), GFDL-ESM2G and HadGEM2-ES. MME of better performing GCMs captured return periods associated with observed moderate and severe droughts. Ahmed *et al.* (2020) evaluated 36 CMIP5-GCMs for precipitation, minimum temperature and maximum temperature over Pakistan. They used Taylor skill score (*TSS*). HadGEM2-AO, CMCC-CM, CESM1(CAM5) were the preferred GCMs. Optimum ensemble of 18 top ranked GCMs was suggested.

In summary, for Pakistan, Khan *et al.* (2018), Latif *et al.* (2018), Ahmed *et al.* (2018), Ahmed *et al.* (2019) and Ahmed *et al.* (2020) considered six, two, three, four and 18 top ranked GCMs for forming MME, respectively in their

studies. Ahmed *et al.* (2020) carried out an extensive and informative literature review regarding optimum number of GCMs for MME.

Middle East (Iran, Iraq, Syria)

Farzaneh *et al.* (2012) evaluated four CMIP3-GCMs, CCSR, CGCM2, CSIRO and HadCM3, for the Northern Karoon region, Iran for precipitation and temperature and found that HadCM3 was suitable due to more appropriate correlation with the observed data. Afshar *et al.* (2017) analysed the Kashafrood mountainous watershed, Iran for precipitation and temperature. They evaluated 14 GCMs of CMIP5 repository with ratio of the root mean square error to the standard deviation of measured data (*RSR*), percent of bias (*PBIAS*), *NSE* and coefficient of determination (R^2). Preferred GCMs were GFDL-ESM2G, IPSL-CM5A-MR, MIROC-ESM and NorESM1-M.

Zamani & Berndtsson (2019) evaluated 20 CMIP5-GCMs for temperature and precipitation. TOPSIS was applied to rank GCMs for Bakhtegan (BKH), Zard River (ZR), and Ghareso (GH) in west and southwest Iran. Indicators were *NRMSE*, *TSS*, Brier score (*BS*) and *SS*. MIROC-ESM, MPI-ESM-MR, MPI-ESM-LR and GFDL-ESM2M were preferred for the ZR basin; BCC-CSM1.1, CanESM2, MIROC5 and ACCESS1.0 for the BKH basin; BCC-CSM1.1, CanESM2, ACCESS1.0 and NorESM1-M for the GH basin.

Abbasian *et al.* (2019) evaluated 37 CMIP5-GCMs for Iran for precipitation and temperature. Evaluation statistics were Kolmogorov–Smirnov (*KS*) statistic, *CC*, *NSE*, mean bias, *RMSE* and Sen's slope estimator. CMCC-CMS and MRI-CGCM3 were preferred GCMs. Ehteram *et al.* (2018) evaluated five CMIP3-GCMs for the Dez basin, Iran for precipitation and temperature. They used indicators *RMSE*, mean absolute error (*MAE*) and *CC*. HadCM3 was the preferred GCM.

Doulabian *et al.* (2020) evaluated 25 CMIP5-GCMs for precipitation and surface air temperature (SAT) for six synoptic stations in Iran. Indicators chosen were *RMSE* and *SS*. They suggested suitable GCMs for each station, variable and indicator. GCMs performed better in simulating SAT compared to precipitation. As part of the literature review, they also discussed various studies related to Iran.

Homsy *et al.* (2020) evaluated 20 CMIP5-GCMs for precipitation for a case study of Syria. The indicator chosen was occurrence frequency of GCM. Employed ranking techniques were *SU* and multicriteria decision technique. HadGEM2-AO, CSIRO-Mk3.6.0, NorESM1-M and CESM1(CAM5) were the preferred GCMs. RF-based ensembling algorithm was performed to generate MME with four selected GCMs. Khayyun *et al.* (2020) evaluated 20 CMIP5-GCMs for precipitation for a case study of Iraq with *SU* as the indicator. Four GCMs, HadGEM2-AO, HadGEM2-ES, CSIRO-Mk3.6.0 and MIROC5, were found to be suitable.

South Asia, East Asia and South East Asia

Le & Bae (2013) evaluated 14 CMIP3-GCMs for temperature and precipitation for the South Korean peninsula. Indicators were *CC* and *RMSE*. CNRM-CM5, HadCM3, CSIRO-Mk3.0, IPSL-CM4, NCAR-CCSM3, CCCMA-CGCM3_T47 were the preferred GCMs. Prasanna (2015) studied 12 CMIP5-GCMs for their capability over South Asia using mean and coefficient of variation (C_v). NORESM, MPI-ESM, GISS, GFDL-ESM2M, CanESM, MIROC5, HadGEM2-ES, CNRM and ACCESS were the preferred GCMs.

Tan *et al.* (2014) chose six CMIP5-GCMs out of 18 for Johor River basin, Malaysia for projection of precipitation and temperature for hydrological impact assessment. They strongly discouraged the use of projections produced by a single GCM. Hussain *et al.* (2018) evaluated 20 CMIP5-GCMs for Rajang River basin (RRB), Sarawak, Malaysia for precipitation. Indicators used were *MAE*, *RMSE*, *CC* and normalised standard deviation (*NSD*). GFDL-ESM2M, ACCESS1.3 and ACCESS1.0 were the preferred GCMs and discouraged the use of a single GCM for climate change assessment. An overview of recent studies over the South and South East Asian region was also provided by Hussain *et al.* (2018). Noor *et al.* (2019) evaluated 58 CMIP5-GCMs for Malaysia for projection of precipitation. Ratio of standard deviations (*rSD*), *FSS*, *NRMSE*, modified index of agreement and *PBIAS* were the indicators employed to evaluate GCMs. Four GCMs, HadGEM2-ES, CCSM4, BCC_CSM1.1(m) and CSIRO-Mk3.6.0 were preferred. MME of four GCMs was performed with RF.

Ruan *et al.* (2018) analysed 34 CMIP5-GCMs for precipitation over the Lower Mekong basin, South East Asia. They used improved rank score method that aggregates *RMSE*, *PBIAS*, *BS*, linear correlation coefficient, Sen's slope, Mann-Kendall (*MK*) statistic and the significance score (*SIS*). The top five preferred GCMs were MPI-ESM-LR, IPSL-CM5A-MR, CMCC-CMS, CESM1(CAM5) and BNU-ESM. Raghavan *et al.* (2018) analysed ten CMIP5-GCMs for South East Asia for precipitation. Indicators were annual cycles, mean climatological spatial distributions, regional area averages, *RMSE*, *CC*, empirical orthogonal function (*EOFs*) and interannual variability. No GCM was suitable for the chosen case study. Kamworapan & Surussavadee (2019) evaluated 40 CMIP5-GCMs for precipitation and temperature in South East Asia. They employed 19 different performance metrics which included *RMSE*. The ability of the preferred GCM was correlated with six best GCM-ensemble and 40 GCM-ensemble for four categories. They recommended CNRM-CM5-2 and 6-GCM-ensemble for climate studies in South East Asia. Sridhar *et al.* (2019) discussed the choice of GCMs for the case study of Mekong basin for precipitation and temperature. MIROC was preferred for the later part of the century. IPSL and GFDL were preferred through the year 2040 for projection of variables.

Australia

Perkins *et al.* (2007) evaluated 14 GCMs of CMIP3 for precipitation, 13 for minimum temperature and ten for maximum temperature for 12 regions of Australia employing *SS* to evaluate the efficacy of GCMs. MIROC-m, CSIRO and ECHO-G were the top three GCMs for all three chosen variables. They suggested omitting weak GCMs from MME as these strongly biased the skill of MME. Suppiah *et al.* (2007) judged the performance of 23 GCMs for mean sea level pressure, precipitation and temperature over the Australian continent. Indicators were *RMSE* and spatial correlation (*SC*) thresholds; accordingly, they identified 15 GCMs. Smith & Chandler (2010) evaluated 22 GCMs for precipitation for Australia.

Johnson & Sharma (2009) proposed variable convergence score (*VCS*) for Australia, which was used to test the outputs of nine CMIP5-GCMs, eight climate variables

and two emission scenarios. Pressure, temperature and humidity had high VCS that influence the climate in Australia. Johnson *et al.* (2011) evaluated 23 CMIP3-GCMs for Southeastern Australia. Wavelet-based skill scores (WSS) were used to compare GCM performance for sea surface temperature, precipitation and surface pressure anomaly. MPI-ECHAM5 was found to have the best performance for all the chosen climate variables. Fu *et al.* (2013) evaluated 25 CMIP3-GCMs for Southeastern Australia for air temperature, mean sea level pressure and precipitation. They used %RE, NRMSE, CC, trend magnitude, MK test, EOF, BS and SIS as indicators. Aggregation of all ranks for each GCM for each variable in the form of rank score (RS) was computed. CSIRO, MIROC-m and IPSL-CM4 were found to be suitable. They also compared 25 CMIP3-GCMs with 40 CMIP5-GCMs. Out of 65 GCMs, CMIP3-GCM was found to be the best. Wang *et al.* (2016) evaluated 28 CMIP5-GCMs for variable temperature. They used indicator, TSS and selected the seven best GCMs and used the same for MME for New South Wales (NSW) wheat belt in Southeastern Australia. Comparison of two MME approaches, arithmetic mean (AM) and independence weighted mean (IWM) was made.

Africa

Hughes *et al.* (2014) studied nine CMIP3-GCMs for their efficacy to simulate precipitation for 15 catchments in five regions of Africa. Evaluation measures included seasonal skill, statistical skill and serial correlation skill. GISS, CNRM and MPI were found to be more skilful with some exceptions. However, differences were observed in skills of GCMs between inland regions and coastal areas. They suggested these GCMs for exploring as ensembles. Aloysius *et al.* (2016) evaluated 25 CMIP5-GCMs for Central Africa for precipitation, mean surface air temperature, minimum surface air temperature and maximum surface air temperature. Mean square error (MSE), spatial pattern correlation (SPC) coefficient, spatial skill score were the indicators. It was observed that GCMs simulate temperature better than precipitation. Agyekum *et al.* (2018) evaluated 18 CMIP5-GCMs for precipitation in the Volta basin, West African region. Indicators were σ , CC, RMSE. CESM1(BGC), CCSM4, NorESM1-M, MPI-ESM-MR, and ensemble mean

of all the 18 GCMs performed relatively well with small biases over most parts of the basin.

Ongoma *et al.* (2019) evaluated 22 CMIP5-GCMs for historical simulations of precipitation for East Africa. CC, σ , bias, PBIAS, RMSE and trend were the indicators. Performances of individual GCMs varied. Eight GCMs, MICROC5, INM-CM4, EC-Earth, CSIRO-Mk3.6.0, CNRM-CM5, CMCC-CESM, CESM1(CAM5) and CanESM2, performed relatively well. However, they suggested improvement in rainfall-related process in GCMs. MME of all GCMs was formulated by assigning equal weights to all GCMs and it was found that individual GCMs performed better than MME. Similar views were expressed by Joubert & Hewitson (1997) for Southern Africa.

Europe

Jury *et al.* (2015) evaluated a total of 81 realisations of 20 CMIP5-GCMs for reproducing near-surface variables for European domain of the Coordinated Regional Climate Downscaling Experiment (EURO-CORDEX). The indicator was model performance index (MPI). MIROC4 h was found to be suitable for the chosen region. Basharin *et al.* (2016) evaluated 12 CMIP5-GCMs for the European region for precipitation and temperature. GCMs well reproduced historical tendencies of regional warming. CNRM-CM5, GFDL-CM3, HadGEM2-ES, MIROC5, CanESM2 and MPI-ESM-LR were the preferred GCMs.

America

Walsh *et al.* (2008) analysed 15 CMIP3-GCMs for seasonal cycle of precipitation, mean sea level pressure and temperature in Alaska and Greenland, North America with RMSE as the indicator. No single GCM outperformed the other GCMs for either all regions or all variables. GFDL-CM2.1, MPI-ECHAM5, MIROC3 and HadCM3 were the preferred GCMs for future projections. They also suggested a subset of GCMs to narrow down uncertainty as well as for projections that were more robust. Radić & Clarke (2011) evaluated 22 CMIP3-GCMs for number of climate variables. Indicators were RMSE, model variability index (MVI) and model climate performance index (MCPi). MRI-CGCM2.3.2,

ECHAM5-MPI-OM and MIROC3.2 (hires) were the top three ranked GCMs.

Anandhi *et al.* (2011) evaluated 41 GCMs/realisations of CMIP3 repository for the Catskill Mountain watersheds, New York, having variable snow water equivalent with the indicator SS. GFDL2.0 was found suitable. They also classified GCMs into three groups based on SS, high (0.83–0.93), medium (0.72–0.83) and low (0.26–0.72). Rupp *et al.* (2013) evaluated 41 CMIP5-GCMs and 24 CMIP3-GCMs for the Pacific Northwest and surrounding regions in the US. Climate variables chosen were precipitation and temperature. Metrics were amplitude of seasonal cycle, long-term persistence, annual- to decadal-scale variance, diurnal temperature range, variance of mean seasonal spatial patterns, correlation and regional teleconnections to El Niño Southern Oscillation (ENSO). CNRM-CM5, CESM1(CAM5) and CanESM2 favoured CMIP5-GCMs and few differences existed between CMIP5 and CMIP3 with respect to the analysed statistics.

Gulizia & Camillonia (2015) evaluated 19 CMIP5-GCMs and 19 CMIP3-GCMs as well as a MME of eight GCMs for South America. Variables were summer precipitation, winter precipitation and annual precipitation. Indicators were RMSE, SC, RE and relative bias. Best representation of the observed patterns in most seasons and regions was reflected in MME. However, MIROC4h of CMIP5 and MIROC3.2 (hires) of CMIP3 repository is better performed than the ensemble in some regions and seasons. Venkataraman *et al.* (2016) analysed 21 CMIP5-GCMs for climate variables, minimum surface air temperature, maximum surface air temperature, average surface air temperature and monthly precipitation in the state of Texas. Performance indicators chosen were MAE and NSD. GCMs simulated historical temperature better than precipitation. They suggested MME of all 21 GCMs as compared to subset of GCMs.

Bhowmik *et al.* (2017) evaluated ten CMIP5-GCMs for conterminous United States with variable precipitation having indicator RMSE. ACCESS and BCC-CSM emerged as preferred GCMs. They proposed MME using equal weighting, percentile and non-percentile-based optimal weighting. Sensitivity analysis on ranking was also performed. Ahmadalipour *et al.* (2017) evaluated 20 CMIP5-GCMs for temperature and precipitation in the Columbia River basin (CRB) in the Pacific Northwest US. Indicators used were (i) mean, (ii) σ , (iii) C_v , (iv) relative change

(variability), (v) Mann–Kendall (MK) trend and (vi) KS test. GCMs in the order of decreasing ranking were BCC_CSM1.1, GFDL-ESM2M, CCSM4, GFDL-ESM2G, MIROC5, CanESM2, IPSL-CM5A-MR, IPSL-CM5B-LR, IPSL-CM5A-LR and MIROC-ESM. Cheng *et al.* (2017) evaluated six CMIP5-GCMs and their ensemble mean for Athabasca River basin, Canada for precipitation, minimum temperature and maximum temperature. Ensemble mean did not outperform any GCM even though its overall accuracy was higher. It was suggested that GCMs be integrated according to accuracy variations. Bhowmik *et al.* (2017) evaluated ten CMIP5-GCMs for climate regions of the conterminous United States for projection of precipitation. No GCM was found to be dominant and a MME of ten GCMs using equal weighting, percentile and non-percentile-based optimal weighting was developed. Anandhi *et al.* (2019) evaluated 20 CMIP5-GCMs for New York City. Variables analysed were precipitation, minimum temperature, maximum temperature, average temperature and wind speed. The indicator used was SS. No single GCM was identified as superior because different GCMs reacted differently for different variables.

Other regions

Moise & Delage (2011) evaluated 23 CMIP3-GCMs for the South Pacific Convergence Zone for precipitation. Indicators chosen were location metric and shape metric. It was observed that a group of GCMs performed well for one metric and not so well for the other metric. However, GFDL-CM2.0, CCSM3 and PCM were found consistent over both metrics.

Errasti *et al.* (2011) evaluated 24 CMIP3-GCMs for the Iberian Peninsula. Variables considered were precipitation, temperature and mean sea level pressure; the indicator used was SS. HadGEM1, BCCR-BCM2.0, GFDL-CM2.1, MPI-ECHAM5 and MIROC3.2 (hires) occupied top positions. Barfus & Bernhofer (2015) evaluated 12 CMIP3-GCMs for the Arabian Peninsula for tropospheric stability. Indicators used were cross totals index, sweat, k-index, show alter index, vertical and total totals index. No GCM was found to outperform other GCMs.

Su *et al.* (2013) analysed performance of 24 CMIP5-GCMs over the eastern Tibetan Plateau for temperature

and precipitation. Climatological patterns and spatial variations of the observed temperature were reasonably captured by GCMs. [Salunke et al. \(2019\)](#) evaluated 28 CMIP5-GCMs for surface air temperature and precipitation for the Himalaya-Tibetan Plateau. They employed *PC* and *RMSE* as indicators for evaluating GCMs. CCSM4, CESM1(CAM5), EC-Earth and IPSL-CM5A-MR best simulated the surface air temperature, whereas EC-Earth and MIROC5 best simulated the precipitation with high *SPC* values and low *RMSE*. They suggested improvement in the resolution and parameterisation schemes. [Jia et al. \(2019\)](#) evaluated 33 CMIP5-GCMs for precipitation over the Tibetan Plateau. They used improved rank score method which is the amalgamation of mean value, σ , *SC*, temporal correlation coefficient, *MK* test statistics, *BS* and *SIS*. CSIRO-Mk3.6.0, EC-Earth, MRI-CGCM3, CNRM-CM5 and CanESM2 were the preferred GCMs.

[Perez et al. \(2014\)](#) evaluated 26 CMIP3-GCMs and 42 CMIP5-GCMs for the North-east Atlantic region. Variables were precipitation, snow, storm surge and wave height. Indicators used were scatter index (*SI*) and relative entropy (*RE*). Three GCMs, MIROC3.2 (hires), ECHAM5/MPI-OM and HadGEM2 of CMIP3 and seven CMIP5-GCMs, CMCC-CM, MPI-ESM-P, HadGEM2-ES, HadGEM2-CC HadGEM2-AO, ACCESS1.0 and EC-Earth were found to be the best. [Ashofteh et al. \(2016\)](#) analysed seven CMIP5-GCMs for simulating runoff in the Aidoghmoush basin, East Azerbaijan. The performance indicators used were *NSE*, *CC*, *RMSE* and *MAE*. HadCM3 was the preferred GCM.

Global studies

[Reifen & Toumi \(2009\)](#) evaluated 17 CMIP3-GCMs for projection of temperature. They recommended MMEs (six GCMs to 16 GCMs) which produced better prediction capability than any single GCM. [Cai et al. \(2009\)](#) evaluated 17 CMIP3-GCMs for their simulating capability of temperature and precipitation. The indicator was *SS*. No single GCM was preferred for the whole world. Some GCMs are better for certain particular regions. [Schaller et al. \(2011\)](#) preferred five best CMIP3-GCMs out of 24 for MME worldwide. They used MME for projection of precipitation and temperature.

[Macadam et al. \(2010\)](#) evaluated 17 CMIP3-GCMs for the variables temperature and temperature anomaly for worldwide, USA and Europe. The indicator used was *SS*. They used the turnover concept to identify the best and the weakest GCMs. First ranked GCMs for temperature in the case of worldwide, USA and Europe were GISS-ER, ECHAM5 and CCSM3, respectively. A similar trend was observed for temperature anomaly. [Watterson et al. \(2014\)](#) explored precipitation, temperature and mean sea level pressure with non-dimensional arcsin Mielke measure *M* as indicator. Twenty-five CMIP5-GCMs and 24 CMIP3-GCMs were analysed for each continent and worldwide. Overall, CMIP5 MME represented a modest improvement in skill over CMIP3 for global land (excluding Antarctica) and six continents. [Mehran et al. \(2014\)](#) evaluated 34 CMIP5-GCMs for several parts of the world with Global Precipitation Climatology Project (GPCP) data. CMIP5 simulations and GPCP patterns were in close agreement in many regions. However, their replication is problematical over arid regions and certain subcontinental regions. [Grose et al. \(2014\)](#) evaluated 27 CMIP5-GCMs and 24 CMIP3-GCMs for the western tropical Pacific for projection of precipitation, temperature and mean sea level pressure. Indicators used to evaluate GCM performance were *RMSE*, *SC* and σ . CMIP3-GCMs that performed well were CSIRO-Mk3.5, ECHO-G, GFDL-CM2.0 and MRI-CGCM2.3.2 whereas CMIP5-GCMs that performed well were ACCESS1.0, CCSM4, CNRM-CM5 and NorESM1-M. They cautioned that selection of the best GCMs should not be perceived as guideline for weighting or sub-setting GCMs.

[McMahon et al. \(2015\)](#) evaluated 22 GCMs of CMIP3 category for temperature and precipitation for global land surface data obtained from the Climatic Research Unit. Indicators used were *NSE*, *RMSE* and R^2 . HadCM3, MIROC-m, MIUB, MPI and MRI were found to be preferred GCMs. They also provided insights from 15 related papers with various features. According to them, *RMSE* was the most preferred indicator. [Baker & Taylor \(2016\)](#) analysed 34 CMIP5-GCMs for top-of-atmosphere and surface radiative flux variance with 44 performance indicators for clouds, and the Earth's radiant energy system observations and GISS surface temperature analysis. CESM, ACCESS and NorESM were the best performing GCMs.

MULTI-MODEL ENSEMBLE OF GCMs WITHOUT EXPLICIT EVALUATION

There are several studies where multi-model ensembles were considered without explicit evaluation of GCMs. The authors of these studies suggested MMEs based on performance of GCMs in previous studies, data availability, holistic understanding of the case study and related analyses. Relevant case studies in this perspective are as follows.

China

Tian *et al.* (2017) employed four CMIP5-GCMs, BNU-ESM, GISS-E2-R, MIROC5 and MPI-ESM-LR, for Xiangjiang river basin in central China. These four GCMs are found to capture the 'major features of distribution and variability of temperature and precipitation throughout China'. Wang *et al.* (2019) studied an ensemble of 29 CMIP5-GCMs for Xiangjiang watershed and Manicouagan-5 watersheds in China for precipitation, maximum temperature and minimum temperature. They employed several weighting schemes including reliability ensemble averaging (REA), upgraded REA, Bayesian model averaging, etc. They assigned weights to GCM simulations and investigated the impacts of weights on quantification of hydrological impacts.

Indian subcontinent

Bae *et al.* (2015) employed MME of three CMIP3-GCMs, CGCM3_T47, CGCM2.3.2 and CM4 out of nine GCMs for ascertaining the Asian monsoon region. Das & Umamahesh (2017) employed REA, Bayesian analysis and delta method for combining six CMIP5-GCMs for Wainganga River basin, India for discharge. Abeysingha *et al.* (2018) used a hybrid-delta method for combining 22 CMIP5-GCMs projections for Gomti River basin, India for precipitation, minimum temperature, maximum temperature and average temperature. Saeed & Athar (2018) evaluated 22 CMIP3-GCMs for the projection of temperature and precipitation in Pakistan and suggested MMEs with all 22 GCMs. Bisht *et al.* (2019) performed MME of nine CMIP5-GCMs using the Taylor diagram statistics for the projection of precipitation and temperature for different homogeneous monsoon regions of India. MME was proposed

based on the competence of GCMs in dealing with the climatic cycle for different homogeneous regions of India. MME was found to simulate the seasonal cycle of the regions with reasonable accuracy compared to IMD data. Vandana *et al.* (2019) used a hybrid-delta ensemble method for combining 16 CMIP3-GCMs for Brahmani River basin, India for precipitation and temperature. All these researchers employed MME for impact and related studies using hydrological models. Mustafa *et al.* (2019) employed MME of 22 CMIP5-GCMs for a drought-prone study area in Bangladesh.

Middle East (Iran and Iraq)

Osman *et al.* (2014) projected daily precipitation based on ensemble of seven CMIP3-GCMs, for three time periods for a case study of Central Iraq. They treated each GCM prediction as an equally possible evolution of climate. Zamani *et al.* (2017) developed an MME framework of 14 CMIP3-GCMs for precipitation and temperature for southwest Iran. They used a mean observed temperature-precipitation (MOTP) approach for MME. Saki *et al.* (2018) considered MME of 14 CMIP5-GCMs for projecting mean annual precipitation, maximum temperature and minimum temperature for Isfahan province, central Iran. They mentioned that selection of 14 GCMs is mainly due to the availability of the data for the chosen RCPs. Sayadi *et al.* (2019) examined the impact of climatic change on maximum temperature, minimum temperature and rainfall using MME of 15 CMIP3-GCMs for Doroudzan catchment, northeast Fars province, Iran for three time periods. Vaghefi *et al.* (2019) employed MME of five CMIP5-GCMs, namely, NorESM1-M, GFDL-ESM2M, MIROC, IPSL-CM5A-LR, HadGEM2-ES for rainfall, maximum temperature, minimum temperature and occurrences of extreme temperatures with reference to flooding for Iran. Nourani *et al.* (2019) considered CMIP5-GCMs, CanESM2, INM-CM4, BNU-ESM and their ensembles for projecting temperature and mean monthly precipitation for Tabriz and Ardabil in northwest Iran. These GCMs were chosen due to their successful application in the case study area.

Canada

Minville *et al.* (2010) used an ensemble of five CMIP3-GCMs for Peribonka water resource system, Quebec, Canada. Most

of these GCMs were used earlier as part of the Atmospheric Model Inter-Comparison Project. [Chen *et al.* \(2017\)](#) employed five weighting methods, random weights, equal weights, *REA*, representation of annual cycle (RAC) in precipitation and temperature and upgraded *REA* for MME of 28 CMIP5-GCMs for Manicouagan 5 watershed, centre of the province of Quebec, Canada. It was observed that weighting of GCMs had limited impact.

Australia

[Schepen & Wang \(2013\)](#) employed MME of six GCMs for projecting Australian seasonal rainfall through Bayesian model averaging (BMA). [Yan *et al.* \(2018\)](#) considered MME of seven CMIP5 – third phase of the Paleoclimate Modeling Intercomparison Project (PMIP3) GCMs, namely, MRI-CGCM3, MIROC-ESM, MPI-ESM-P, GISS-E2-R, IPSL-CM5A-LR, CCSM4 and CNRM-CM5 for interpreting the Australian monsoon. These GCMs were found to have a better performance in simulating the Australian monsoon. [Al-Safi & Sarukkalgige \(2017\)](#) formulated MME of eight CMIP5-GCMs for Richmond River catchment, Australia. According to the authors, MME ‘effectively represents the Australian future climate’. [Wang *et al.* \(2018\)](#) considered monthly rainfall and temperature for Australia with the ensemble of 33 CMIP5-GCMs. Bayesian model averaging (BMA), support vector machine (SVM), arithmetic ensemble mean (EM) and RF were considered for ensembling. RF and SVM are found to be preferred MME approaches. They also classified GCMs into top, middle and bottom category based on Taylor skill score observation. [Al-Safi & Sarukkalgige \(2019\)](#) considered MME of eight CMIP5-GCMs for projecting rainfall and temperature at Harvey, Beardy and Goulburn catchments, Australia.

Africa

[Adhikari & Nejadhashemi \(2016\)](#) employed an ensemble of six CMIP5-GCMs, CanESM2, CCSM4, CSIRO-Mk3.6, IPSL-CM5A-LR, MPI-ESM-LR and MRI-CGCM3, for Malawi located in southeastern Africa. These GCMs were chosen due to their capability to predict the seasonal migration of the Inter-Tropical Convergence Zone that governs the precipitation in the Greater Horn of Africa.

[Akinsanola & Zhou \(2019\)](#) proposed MME of four CMIP5-GCMs, CanESM2, INM-CM4, IPSL-CM5A-MR and MIROC5, for the West African summer monsoon (WASM) with respect to the present state and future changes. Selection of these GCMs were based on ‘their remarkable performance in reproducing the rainfall characteristics over West Africa’. Results from the historical period indicated that ensemble mean reasonably reproduced the characteristics of WASM rainfall.

Europe

[Jackson *et al.* \(2011\)](#) used ensembling of 13 CMIP3-GCMs for a case study area 70 km west of London for analysing groundwater resources in the uncertain perspective. [Hagemann *et al.* \(2013\)](#) employed MME of three CMIP3-GCMs, ECHAM5, IPSL, LMDZ-4 and CNRM-CM3 for precipitation, temperature and other climate variables along with eight hydrological models and used the data of WATCH project. The selection of GCMs was based on the available data. However, they demonstrated that uncertainty in selection of hydrology model was larger than that of selection of GCMs. [Claudia *et al.* \(2017\)](#) employed four CMIP5-GCMs for case study of Alto Sabor watershed, northeast Portugal. [Saraiva *et al.* \(2019\)](#) studied an ensemble of 21 scenario simulations driven by four CMIP5-GCMs, MPI-ESM-LR, EC-Earth, IPSL-CM5A-MR and HadGEM2-ES for the Baltic Sea and uncertainty in projections was due to GCMs. [Niel *et al.* \(2019\)](#) employed MME of 24 CMIP5-GCMs with 93 simulation runs for the case study of the Grote Nete catchment and the Dijle catchment, Belgium, for assessing flow extremes. [Amraoui *et al.* \(2019\)](#) employed seven CMIP3-GCMs, ARPV3, CCCMA-CGCM 3, MRI-CGCM2.3.2, GFDL-CM2.0, GFDL-CM2.1, GISS and MPI-ECHAM5 for ensembling and projection of precipitation and other climate variables for the Somme River basin, northern France.

America

[Schuster *et al.* \(2012\)](#) used MME of 14 CMIP3-GCMs for projection of precipitation for the state of Wisconsin, US. [Acharya *et al.* \(2012\)](#) used MME of 16 CMIP3-GCMs for projection of temperature and precipitation for North Platte River watershed, in the states of Wyoming and Colorado,

US. [Daraio \(2017\)](#) used MME of 13 CMIP5-GCMs for the case study of the outer coastal plain of New Jersey, the Upper Maurice River and the Batsto River, US. [Acharya \(2017\)](#) employed 36 CMIP5-GCMs for the case study of Flint River watershed in Northern Alabama, US. [Sultana & Choi \(2018\)](#) used seven CMIP5-GCMs for an American river basin, a snow-dominated alpine watershed in Northern California and GCMs were chosen based on their previous performance. [Mani & Tsai \(2017\)](#) combined 13 CMIP5-GCMs for south Arkansas and north Louisiana, US using hierarchical Bayesian model averaging (HBMA), simple model averaging (SMA) and REA. HBMA performed marginally superior to REA and SMA in simulating the historical mean discharge.

OBSERVATIONS AND DISCUSSION

As discussed earlier, several researchers worked on performance indicators, ranking of GCMs and ensembling of GCMs on different aspects. Related observations and discussion are presented in the following sections.

Number and type of indicators

Observation: Most of these studies did not report the basis of selecting indicators. [Smith & Chandler \(2010\)](#) concluded that no universally agreed criterion existed for the assessment of GCMs.

Discussion: The fundamental question arises whether these many indicators are necessary to evaluate GCMs. Four perspectives exist in this direction, which are independent of each other and can be considered as possible alternatives:

1. Any one indicator will suffice for evaluating the performance ability of GCMs to simulate observed data on the assumption that all indicators are equally capable. The indicator can be either SS, based on probability density functions, or a much simpler indicator such as RMSE, or any other similar indicator.
2. The indicator can be selected category-wise (example: one from error, one from correlation coefficient, one from skill score). If required, a composite indicator can be formulated.

3. Compute weights of chosen indicators using either a rating method, analytic hierarchy process (AHP) or their fuzzy extension. Accordingly, for evaluation, the highest weighted indicator can be considered.
4. Lastly, the principal component analysis of indicators can be explored for dimension reduction, to arrive at (say) two components which will be sufficient and useful for evaluating the GCMs.

We suggest, these perspectives can be considered as a future research area.

Handling uncertainty in indicators

Observation: There may be imprecision in observed data and GCM simulated data leading to imprecision in the evaluation of GCMs. In addition, interpolation, averaging procedures and approximations can also cause imprecision. These inherent uncertainties were ignored by many researchers ([Gu et al. 2015](#); [Ahmed et al. 2018](#)).

Discussion: One way of representing uncertainty relates to probability and another way to fuzzy logic. The first approach is based on distribution functions and hypothesis, whereas the second approach is based on membership functions and natural language. Fuzzy logic is becoming prominent due to its simple mathematics and seamless extension with classical logic. For example, crisp-based decision-making techniques can be easily extended to fuzzy-based by incorporating membership functions with classical logic mathematics intact.

One of the remedies is to provide an uncertainty allocation for each indicator in a fuzzy logic framework in the form of a triangular membership function, as shown in [Figure 1](#), or trapezoidal, Gaussian, etc. In the triangular membership function, q represents the most likely value, whereas p and r represent the extreme values (lower and upper bounds). For example, let the evaluation indicator have most likely value of 0.6 for a GCM; by assuming deviation of (say) $\pm 10\%$, i.e., 0.06. Then the indicator value in the triangular membership framework is (0.54, 0.6, 0.66), respectively, representing (p , q , r). If the researcher feels that all simulated and observed values are precise, data will be represented as (0.6, 0.6, 0.6). [Raju & Nagesh Kumar \(2015b\)](#) used the standard deviation as the uncertain

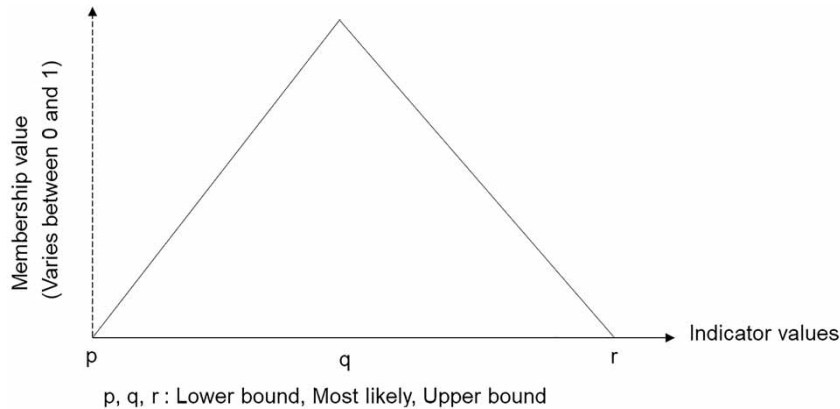


Figure 1 | Typical triangular membership function.

measure. However, the uncertainty allocation depends on the individual choice of the end user. Table 1 presents hypothetical data representing uncertain allocation of indicators that can be utilised while ranking GCMs. For example, GCM1, GCM2, GCM N can be any GCM of CMIP3 or CMIP5 repository.

However, the impact of these variations can be tested as part of sensitivity analysis, if necessary. Mathematical expressions for *NSE*, *CC*, *RMSE* and procedure of relevant computations are presented in Raju & Nagesh Kumar (2014b).

Ranking of GCMs

Jain *et al.* (2019) suggested selection of GCMs which would work for all climate variables as well as for all seasons. They cautioned that the outcome would be based on the chosen diagnostics and GCMs. The following common observations have emanated from the studied research papers on ranking of GCMs.

Observation:

- No logic of choosing X GCMs from Y
- No unanimity on the choice of GCM selection approach
- No complex ranking techniques to handle imprecision in data or resulting imprecision in indicators.

Discussion: Simple crisp decision-making techniques can handle unique value for each indicator. They may not suffice for the problem when data are imprecise that require consideration of uncertainty (Raju & Nagesh Kumar 2014a). Keeping this in view, (a) fuzzy logic-based decision-making techniques and (b) stochastic-based decision-making techniques are proved to be useful. Some of them are:

(a) Fuzzy-based decision-making techniques

- Similarity analysis – GCM with a higher degree of resemblance relative to a bench marking GCM was treated as suitable (Chen 1994). The performance indicators were expressed by interval valued fuzzy sets.

Table 1 | Hypothetical data representing uncertain allocation of indicators

Model	NSE			CC			RMSE		
	[p	q	r]	[p	q	r]	[p	q	r]
GCM 1	0.649	0.806	0.964	0.390	0.466	0.578	0.714	0.788	0.863
GCM 2	0.670	0.828	0.985	0.436	0.534	0.687	0.704	0.778	0.853
.....
GCM N									

- Fuzzy Technique for Order Preference by Similarity to an Ideal Solution (F-TOPSIS) – It is based on the philosophy that suitable GCM should have the shortest distance from a positive ideal solution and the farthest distance from the negative ideal solution. F-TOPSIS computes positive separation measure D_i^+ from positive ideal solution, the negative separation measure D_i^- from negative ideal solution, and the closeness index C_i . The higher the C_i value, the better the GCM. Here, the payoff matrix can follow triangular, trapezoidal or any appropriate format (Opricovic & Tzeng 2004; Behzadian *et al.* 2012).
- Fuzzy ViseKriterijumska Optimizacija I Kompromisno Resenje (F-VIKOR) – It is based on normalised fuzzy difference. F-VIKOR computes separation measures from the ideal and negative ideal and the summation operator Q_i with the given strategy weight. GCM with lower Q_i value was preferred (Wu *et al.* 2016; Ploskas & Papathanasiou 2019).

Fuzzy-based algorithms discussed here and elsewhere (Chen & Wang 2013; Marttunen *et al.* 2017; Chu & Yeh 2018; Pei *et al.* 2019) help to tackle uncertainty in ranking of GCMs.

(b) Stochastic-based decision-making techniques

In this type of approach, mean and standard deviation play major roles along with other parameters. Notable techniques in this category are extended TOPSIS in stochastic environment (Xiong & Huan 2010), stochastic VIKOR (Tavana *et al.* 2016) and Stochastic Preference Ranking Organization METHod for Enrichment of Evaluations (PROMETHEE) (Mareschal 1986; Goodwin *et al.* 2019). Celik *et al.* (2019) carried out a comprehensive discussion on stochastic decision-making and probability distribution functions. They also discussed the relevant potential research areas. In our opinion, ranking of GCMs can also be performed in the stochastic decision-making arena which was not explored by many researchers.

Ensembling of GCMs

Numerous researchers suggested multi-model ensembling of GCMs. One group of researchers evaluated GCMs based on performance measures and ranking, whereas other groups of

researchers evaluated GCMs based on their expertise, holistic understanding about the case study and related results.

Observation: No logic of how many GCMs can be part of ensembling

Discussion: Limited research work was performed on this aspect. Some of the relevant and promising studies worth replicating are as follows.

Raju & Nagesh Kumar (2016) applied cluster analysis in association with cluster validation methods, namely, F-statistic and Davies–Bouldin Index for identifying optimal ensemble. They evaluated 36 CMIP5-GCMs for minimum temperature, maximum temperature and the combination of maximum and minimum temperature over India. The indicator was SS. Optimal clusters for minimum temperature, maximum temperature and the combination of maximum and minimum temperature scenarios were of two, three and two GCMs, respectively. Accordingly, respective ensembles of GCMs suggested were (ACCESS1.3, HadCM3), (HadCM3, IPSL-CM5A-LR, GFDL-ESM2M) and (MPI-ESM-MR, HadCM3).

Mendlik & Gobiet (2016) developed a subset of representative GCMs using principal component analysis and cluster analysis. They studied their methodology using ENSEMBLES project. Herger *et al.* (2018) analysed 81 CMIP5 simulations for identifying ensemble subset K. They proposed three approaches: (a) random ensemble, (b) performance ranking ensemble and (c) optimal ensemble. Optimal ensemble approach was found computationally efficient, with sufficient spread in projections as compared to (a) and (b). They used RMSE for their computations. More details can be obtained from Herger *et al.* (2018).

Ahmed *et al.* (2020) employed machine learning algorithms, namely, relevance vector machine (RVM), support vector machine (SVM), and artificial neural network (ANN), K-nearest neighbour (KNN) to establish MMEs for annual, monsoon and winter climate variables and found that KNN and RVM based MMEs showed better skills. They also used top ranked approach and bottom ranked approach to ascertain robustness of MMEs. It is suggested to refer to Ahmed *et al.* (2020) for extensive and informative details regarding machine learning algorithms that can be used as a guideline for MME of GCMs. Readers can refer to Ahmed *et al.* (2018, 2019, 2020) regarding various challenges and extensive discussion and logic of choosing GCMs from repositories.

Some data mining algorithms that facilitate clustering of GCMs are also briefly described below:

- K-means cluster analysis can group GCMs into relatively homogeneous clusters, that is, GCMs in a cluster expectedly more analogous to each other than those in the other clusters (Raju & Nagesh Kumar 2007, 2018). In brief, the procedure of K-means is as follows: (1) initiating with a random number of clusters and allotment of each GCM into a cluster; (2) calculation of group average for each cluster and total error; (3) iteration and classification of GCMs such that termination criterion can be met; (4) estimation of total error for K clusters; (5) repeating of steps 1–4 for various clusters; (6) analysis for optimal K if required.
- Fuzzy cluster analysis (FCA) is a classifying algorithm wherein each GCM is related to a cluster with some degree of membership. The procedure of computations is similar to that of K-means algorithm. Degree of membership of each GCM may alter from cluster to cluster bounded by zero to one, on the philosophy that the sum of the membership values for each GCM is one (Raju & Nagesh Kumar 2007).
- Kohonen artificial neural networks (KANN) is a self-organising mapping technique comprising competitive layers that use the learning rule to cluster GCMs. Each neuron in the output layer is inter-connected to all those in the input layer by a set of weights (Kohonen 1989; Raju & Nagesh Kumar 2007, 2014a).
- ELECTRE-TRI assigns GCMs to some predefined ordered clusters (Rogers *et al.* 2000; Raju *et al.* 2000). The limit between two consecutive clusters is demarcated by a profile. Clusters are mutually exclusive which means that one GCM cannot be entrusted to two different clusters. Construction and exploitation of an outranking relation is the basis for allotting GCMs to clusters.

For clustering algorithms, the number of clusters is an input. However, it is challenging to manually estimate optimum clusters for a set of GCMs. In this context, cluster validation techniques play a significant role in finding the optimal number of clusters. Detailed information about cluster validation techniques and their applications are available (Halkidi *et al.* 2001; Hämäläinen *et al.* 2017). Details of some of the validation indices, Dunn's,

Calinski–Harabasz, Davies–Bouldin, Silhouette and external validation indices are available in Halkidi *et al.* (2001), Dalton *et al.* (2009) and Hämäläinen *et al.* (2017). Most of these validation techniques work with inter-cluster and intra-cluster distances.

To augment the earlier studies on the topic:

- Researchers can use published literature to select suitable GCMs and ensembling for their study area.
- Research institutions can take the lead, and region-wise ranking of GCMs may be done (if not done already or outcomes of previous studies not confirmed) in a robust manner. These GCMs can be frozen to avoid repetitive efforts of researchers. Time thus saved, can be utilised for analysing the projected outcomes and validating them.
- One of the crucial aspects in assessment of GCMs is comparison with observed data which may have such limitations such as, accessibility to all researchers, authenticity, confidence in the data, acquisition procedures and permissions to use the data. It is suggested to have a central database at least for each country where data can be easily accessible. In our opinion, more accessibility of data to users may provide better outcomes which may provide confidence zones to planners for impact studies.

SUMMARY AND CONCLUSIONS

This state-of-the-art review paper has provided insights into mainly three aspects – performance indicators, ranking of GCMs and ensembling of GCMs. This study discussed the role of performance indicators, various types of performance indicators, necessity of GCM evaluation and relevant challenges and the basis of MME. Future research directions can be in terms of handling uncertainty in indicators, and ranking of GCMs both in fuzzy- and stochastic-based decision-making perspectives. In addition, various relevant MME-based approaches that provide optimum ensembling are also presented. In addition, a number of techniques which facilitate ranking and ensembling are also part of the paper.

Brief but relevant conclusions are as follows:

1. Most of the researchers opined that selection of suitable GCM(s) is necessary for impact studies with the understanding that past skill assessment of GCM(s) might help improved projections of climate variables.

Accordingly, related components such as performance measures and suitable decision-making technique(s) played a leading role in the studied papers.

2. Research papers studied here and elsewhere identified a number of sources of uncertainty which, in fact, effect the accuracy of projections. MME approaches proposed by various researchers have been found promising and replicable in reducing the uncertainties of projections. Subsets of GCMs/efficient GCMs can be part of MME that serves the purpose.

It may be noted that the inferences provided in this review paper are based on interpretation from the reviewed research papers.

ACKNOWLEDGEMENTS

This work is supported by the Council of Scientific and Industrial Research, New Delhi, through the Project no. 22(0782)/19/EMR-II dated 24.7.19. The second author would like to acknowledge the funding support provided by Ministry of Earth Sciences, Government of India, through a project with reference number MoES/PAMC/H&C/41/2013-PC-II. The authors would like to thank Prof. Chris Perera, Editor-in-Chief, for encouraging them to write this article and also for his constructive criticism throughout the review process. Some views of esteemed researchers were quoted verbatim to convey their views without losing the meaning. Acknowledgements are due to the esteemed reviewers for providing valuable suggestions which helped us to think more critically while revising this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this paper is available online at <https://dx.doi.org/10.2166/wcc.2020.128>.

REFERENCES

Abbasian, M., Moghim, S. & Abrishamchi, A. 2019 Performance of the general circulation models in simulating temperature and

- precipitation over Iran. *Theor. Appl. Climatol.* **135**, 1465–1483.
- Abeyasingha, N. S., Islam, A. & Singh, M. 2018 Assessment of climate change impact on flow regimes over the Gomti river basin under IPCC AR5 climate change scenarios. *J. Water Clim. Chan.* **11**, 303–426.
- Acharya, A. 2017 Quantification of modeled streamflows under climate change over the Flint River watershed in Northern Alabama. *J. Hydrol. Eng.* **22**, 04017032.
- Acharya, A., Piechota, T. C. & Tootle, G. 2012 Quantitative assessment of climate change impacts on the hydrology of the North Platte river watershed, Wyoming. *J. Hydrol. Eng.* **17**, 1071–1083.
- Adhikari, U. & Nejadhashemi, A. P. 2016 Impacts of climate change on water resources in Malawi. *J. Hydrol. Eng.* **21**, 05016026.
- Afshar, A. A., Hasanzadeh, Y., Besalatpour, A. A. & Pourreza-Bilondi, M. 2017 Climate change forecasting in a mountainous data scarce watershed using CMIP5 models under representative concentration pathways. *Theor. Appl. Climatol.* **129**, 683–699.
- Agyekum, J., Annor, T., Lamptey, B., Quansah, E. & Agyeman, R. Y. K. 2018 Evaluation of CMIP5 global climate models over the Volta Basin: precipitation. *Adv. Meteorol.* **2018**, 1–24.
- Ahmadalipour, A., Rana, A., Moradkhani, H. & Sharma, A. 2017 Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis. *Theor. Appl. Climatol.* **128**, 71–87.
- Ahmed, K., Sachindra, D. A., Shahid, S., Demirel, M. C. & Chung, E. S. 2018 Selection of multi-model ensemble of GCMs for the simulation of precipitation based on spatial assessment metrics. *Hydrol. Earth Syst. Sci. Discuss.* **23**, 4803–4824.
- Ahmed, K., Shahid, S., Sachindra, D. A., Nawaz, N. & Chung, E. S. 2019 Fidelity assessment of general circulation model simulated precipitation and temperature over Pakistan using a feature selection method. *J. Hydrol.* **573**, 281–298.
- Ahmed, K., Sachindra, D. A., Shahid, S., Iqbal, Z., Nawaz, N. & Khan, N. 2020 Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. *Atmos. Res.* **236**, 1–20.
- Akinsanola, A. A. & Zhou, W. 2019 Ensemble-based CMIP5 simulations of West African summer monsoon rainfall: current climate and future changes. *Theor. Appl. Climatol.* **136**, 1021–1031.
- Aloysius, N. R., Sheffield, J., Saiers, J. E., Li, H. & Wood, E. F. 2016 Evaluation of historical and future simulations of precipitation and temperature in Central Africa from CMIP5 climate models. *J. Geophys. Res. Atmos.* **121**, 130–152.
- Al-Safi, H. I. J. & Sarukkalige, P. R. 2017 Evaluation of the impacts of future hydrological changes on the sustainable water resources management of the Richmond river catchment. *J. Water Clim. Chan.* **9**, 137–155.
- Al-Safi, H. I. J. & Sarukkalige, P. R. 2019 The application of conceptual modelling approach to evaluate the impacts of climate change on the future streamflow in three unregulated catchments of the Australian hydrologic reference stations. *Int. J. Hydrol. Sci. Tech.* **9** (5), 494–525.

- Amraoui, N., Sbai, M. A. & Stollsteiner, P. 2019 Assessment of climate change impacts on water resources in the Somme River basin (France). *Water Resour. Manage.* **33**, 2073–2092.
- Anandhi, A. & Nanjundiah, R. S. 2015 Performance evaluation of AR4 climate models in simulating daily precipitation over the Indian region using skill scores. *Theor. Appl. Climatol.* **119**, 551–566.
- Anandhi, A., Frei, A., Pradhanang, S. M., Zion, M. S., Pierson, D. C. & Schneiderman, E. M. 2011 AR4 climate model performance in simulating snow water equivalent over Catskill Mountain watersheds, New York, USA. *Hydrol. Process.* **25**, 3302–3311.
- Anandhi, A., Pierson, D. C. & Frei, A. 2019 Evaluation of climate model performance for water supply studies: case study for New York city. *J. Water Res. Plan. Manage.* **145**. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001054](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001054).
- Ashofteh, P., Bozorg-Haddad, O. & Mariño, M. A. 2016 Performance evaluation of a developed hybrid AOGCM model under climate change. *J. Irrig. Drain. Eng.* **142**. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001107](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001107).
- Bae, D. H., Koike, T., Awan, J. A., Lee, M. H. & Sohn, K. H. 2015 Climate change impact assessment on water resources and susceptible zones identification in the Asian monsoon region. *Water Resour. Manage.* **29**, 5377–5393.
- Baker, N. C. & Taylor, P. C. 2016 A framework for evaluating climate model performance metrics. *J. Climate* **29**, 1773–1782.
- Bannister, D., Herzog, M., Graf, H. F., Hosking, J. S. & Short, C. A. 2017 An assessment of recent and future temperature change over the Sichuan Basin, China, using CMIP5 climate models. *J. Climate* **30**, 6701–6722.
- Bao, J. & Feng, J. 2016 Intercomparison of CMIP5 simulations of summer precipitation, evaporation, and water vapor transport over Yellow and Yangtze River basins. *Theor. Appl. Climatol.* **123**, 437–452.
- Barfus, K. & Bernhofer, C. 2015 Assessment of GCM capabilities to simulate tropospheric stability on the Arabian Peninsula. *Int. J. Climatol.* **35**, 1682–1696.
- Basharin, D., Polonsky, A. & Stankūnavičius, G. 2016 Projected precipitation and air temperature over Europe using a performance-based selection method of CMIP5 GCMs. *J. Water Clim. Change* **7**, 103–113.
- Behzadian, M., Otaghsara, S. K., Yazdani, M. & Ignatius, J. 2012 A state-of-the-art survey of TOPSIS applications. *Exp. Sys. Appl.* **39**, 13051–13069.
- Benedict, I., van Heerwaarden, C. C., Weerts, A. H. & Hazeleger, W. 2019 The benefits of spatial resolution increase in global simulations of the hydrological cycle evaluated for the Rhine and Mississippi basins. *Hydrol. Earth Syst. Sci.* **23**, 1779–1800.
- Bhattacharjee, P. S. & Zaitchik, B. F. 2015 Perspectives on CMIP5 model performance in the Nile River headwaters regions. *Int. J. Climatol.* **35**, 4262–4275.
- Bhowmik, R. D., Sharma, A. & Sankarasubramanian, A. 2017 Reducing model structural uncertainty in climate model projections – a rank-based model combination approach. *J. Climate* **30**, 10139–10154.
- Bisht, D. S., Sridhar, V., Mishra, A., Chatterjee, C. & Raghuvanshi, N. S. 2019 Drought characterization over India under projected climate scenario. *Int. J. Climatol.* **39**, 1889–1911.
- Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M. & Schär, C. 2013 Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Res. Res.* **49**, 1523–1536.
- Cai, X., Wang, D., Zhu, T. & Ringler, C. 2009 Assessing the regional variability of GCM simulations. *Geophys. Res. Lett.* **36**, L02706.
- Celik, E., Gul, M., Yucesan, M. & Mete, S. 2019 Stochastic multi-criteria decision-making: an overview to methods and applications. *Beni-Suef Univ. J. Basic Appl. Sci.* **8**, 1–11.
- Chen, S. M. 1994 A new method for handling multicriteria fuzzy decision making problems. *Cyber. Systems* **25**, 409–420.
- Chen, S. M. & Wang, C. Y. 2013 Fuzzy decision making systems based on interval type-2 fuzzy sets. *Inform. Sci.* **242**, 1–21.
- Chen, J., Brissette, F. P., Lucas-Picher, P. & Caya, D. 2017 Impacts of weighting climate models for hydro-meteorological climate change studies. *J. Hydrol.* **549**, 534–546.
- Cheng, G. H., Huang, G. H., Dong, C., Zhu, J. X., Zhou, X. & Yao, Y. 2017 An evaluation of CMIP5 GCM simulations over the Athabasca river basin, Canada. *J. Amer. Water Works Assoc.* **33**, 823–843.
- Chhin, R. & Yoden, S. 2018 Ranking CMIP5 GCMs for model ensemble selection on regional scale: case study of the Indochina Region. *J. Geophys. Res. Atmos.* **123**, 8949–8974.
- Chu, T. C. & Yeh, W. C. 2018 Fuzzy multiple criteria decision-making via an inverse function-based total utility approach. *Soft Comput.* **22**, 7423–7433.
- Claudia, C. S., Monteiro, A. T., Azevedo, J. C., Honrado, J. P. & Nunes, J. P. 2017 Climate change impacts on water resources and reservoir management: uncertainty and adaptation for a mountain catchment in Northeast Portugal. *Water Resour. Manage.* **31**, 3355–3370.
- Cook, L. M., Anderson, C. J. & Samaras, C. 2017 Framework for incorporating downscaled climate output into existing engineering methods: application to precipitation frequency curves. *J. Infra. Sys.* **23**. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000382](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000382).
- Dalton, L., Ballarin, V. & Brun, M. 2009 Clustering algorithms: on learning, validation, performance, and applications to genomics. *Curr. Genom.* **10**, 430–445.
- Daraio, J. A. 2017 Potential climate change impacts on streamflow and recharge in two watersheds on the New jersey coastal plain. *J. Hydrol. Eng.* **22**, 05017002.
- Das, J. & Umamahesh, N. V. 2017 Uncertainty and nonstationarity in streamflow extremes under climate change scenarios over a river basin. *J. Hydrol. Eng.* **22**. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001571](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001571).
- Doulabian, S., Golian, S., Toosi, A. S. & Murphy, C. 2020 Evaluating the effects of climate change on precipitation and

- temperature for Iran using RCP scenarios. *J. Water Clim. Change*. <https://doi.org/10.2166/wcc.2020.114>.
- Duan, H., Zhang, G., Wang, S. & Fan, Y. 2019 Robust climate change research: a review on multi-model analysis. *Environ. Res. Lett.* **14**, 033001.
- Ehteram, M., Mousavi, S. F., Karami, H., Farzin, S., Singh, V. P., Chau, K. W. & El-Shafie, A. 2018 Reservoir operation based on evolutionary algorithms and multi-criteria decision-making under climate change and uncertainty. *J. Hydroinform.* **20**, 332–355.
- Errasti, I., Ezcurra, A., Sáenz, J. & Ibarra-Berastegi, G. 2011 Validation of IPCC AR4 models over the Iberian Peninsula. *Theor. Appl. Climatol.* **103**, 61–79.
- Fahimi, F., Yaseen, Z. M. & El-shafie, A. 2017 Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. *Theor. Appl. Climatol.* **128**, 875–903.
- Farzaneh, M. R., Eslamian, S., Samadi, S. Z. & Akbarpour, A. 2012 An appropriate general circulation model (GCM) to investigate climate change impact. *Int. J. Hydrol. Sci. Tech.* **2**, 34–47.
- Fu, G., Liu, Z., Charles, S. P., Xu, Z. & Yao, Z. 2013 A score-based method for assessing the performance of GCMs: a case study of southeastern Australia. *J. Geophys. Res. Atmos.* **18**, 4154–4167.
- Gleckler, P. J., Taylor, K. E. & Doutriaux, C. 2008 Performance metrics for climate models. *J. Geophys. Res.* **113**, D06104.
- Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P. & Zorita, E. 2012 What is the role of the observational dataset in the evaluation and scoring of climate models? *Geophys. Res. Lett.* **39**, L24701.
- Goodwin, D., Raffin, M., Jeffrey, P. & Smith, H. M. 2019 Stakeholder evaluations of risk interventions for non-potable recycled water schemes: a case study. *Sci Total Environ.* **674**, 439–450.
- Große, M. R., Brown, J. N., Narsey, S., Brown, J. R., Murphy, B. F., Langlais, C., Gupta, A. S., Moise, A. F. & Irving, D. B. 2014 Assessment of the CMIP5 global climate model simulations of the western tropical Pacific climate system and comparison to CMIP5. *Int. J. Climatol.* **34**, 3382–3399.
- Gu, H., Yu, Z., Wang, J., Wang, G., Yang, T., Ju, Q., Yang, C., Xu, F. & Fan, C. 2015 Assessing CMIP5 general circulation model simulations of precipitation and temperature over China. *Int. J. Climatol.* **35**, 2431–2440.
- Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., Oldenborgh, G. J. V. & Stockdale, T. 2009 Understanding El Niño in ocean–atmosphere general circulation models. *Bull. Amer. Meteor. Soc.* **90**, 325–340.
- Gulizia, C. & Camillonia, I. 2015 Comparative analysis of the ability of a set of CMIP5 and CMIP5 global climate models to represent precipitation in South America. *Int. J. Climatol.* **35**, 583–595.
- Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F. & Wiltshire, A. J. 2013 Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dyn.* **4**, 129–144.
- Halkidi, M., Batistakis, Y. & Vazirgiannis, M. 2001 On clustering validation techniques. *J. Intel. Inform. Sys.* **17**, 107–145.
- Hämäläinen, J., Jauhiainen, S. & Kärkkäinen, T. 2017 Comparison of internal clustering validation indices for prototype-based clustering. *Algorithms* **10**, 105.
- Hargreaves, J. C. & Annan, J. D. 2014 Can we trust climate models? *WIREs Clim. Change* **5**, 435–440.
- He, W., Zhao, S., Wu, Q., Jiang, Y. & Wan, S. 2019 Simulating evaluation and projection of the climate zones over China by CMIP5 models. *Clim. Dyn.* **52**, 2597–2612.
- Heger, N., Abramowitz, G., Knutti, R., Angéllil, O., Lehmann, K. & Sanderson, B. M. 2018 Selecting a climate model subset to optimise key ensemble properties. *Earth Syst. Dyn.* **9**, 135–151.
- Homsí, R., Shiru, M. S., Shahid, S., Ismail, T., Harun, S. B., Al-Ansari, N., Chau, K. W. & Yaseen, Z. M. 2020 Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. *Eng. Appl. Comput. Fluid. Mech.* **14**, 90–106.
- Hughes, D. A., Kingston, D. G. & Todd, M. C. 2011 Uncertainty in water resources availability in the Okavango River Basin as a result of climate change. *Hydrol. Earth Syst. Sci.* **15**, 931–941.
- Hughes, D. A., Mantel, S. & Mohobane, T. 2014 An assessment of the skill of downscaled GCM outputs in simulating historical patterns of rainfall variability in South Africa. *Hydrol. Res.* **45**, 134–147.
- Hussain, M., Yusof, K. W., Mustafa, M. R. U., Mahmood, R. & Jia, S. 2018 Evaluation of CMIP5 models for projection of future precipitation change in Bornean tropical rainforests. *Theor. Appl. Climatol.* **134**, 423–440.
- Jackson, C. R., Meister, R. & Prudhomme, C. 2011 Modelling the effects of climate change and its uncertainty on UK chalk groundwater resources from an ensemble of global climate model projections. *J. Hydrol.* **399**, 12–28.
- Jain, S., Salunke, P., Mishra, S. K. & Sahany, S. 2019 Performance of CMIP5 models in the simulation of Indian summer monsoon. *Theor. Appl. Climatol.* **137**, 1429–1447.
- Jena, P., Azad, S. & Rajeevan, M. N. 2016 CMIP5 projected changes in the annual cycle of Indian monsoon rainfall. *Climate* **4**, 14.
- Jia, K., Ruan, Y., Yang, Y. & Zhang, C. 2019 Assessing the performance of CMIP5 global climate models for simulating future precipitation change in the Tibetan Plateau. *Water* **11**, 1771.
- Jiang, D., Tian, Z. & Langa, X. 2016 Reliability of climate models for China through the IPCC third to fifth assessment reports. *Int. J. Climatol.* **36**, 1114–1133.
- Jobst, A. M., Kingston, D. G., Cullen, N. J. & Schmidt, J. 2018 Intercomparison of different uncertainty sources in hydrological climate change projections for an alpine catchment (upper Clutha River, New Zealand). *Hydrol. Earth Syst. Sci.* **22**, 3125–3142.

- Johnson, F. & Sharma, A. 2009 Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. *J. Climate* **22**, 4373–4382.
- Johnson, F., Westra, S., Sharma, A. & Pitman, A. J. 2011 An assessment of GCM skill in simulating persistence across multiple time scales. *J. Climate* **24**, 3609–3623.
- Joubert, A. M. & Hewitson, B. C. 1997 Simulating present and future climates of southern Africa using general circulation models. *Progress in Phys. Geog.* **21**, 51–78.
- Jury, M. W., Prein, A. F., Truhetz, H. & Gobiet, A. 2015 Evaluation of CMIP5 models in the context of dynamical downscaling over Europe. *J. Climate* **28**, 5575–5582.
- Kamworapan, S. & Surussavadee, C. 2019 Evaluation of CMIP5 global climate models for simulating climatological temperature and precipitation for Southeast Asia. *Adv. Meteorol.* **2019**, 1–18.
- Khan, N., Shahid, S., Ahmed, K., Ismail, T., Nawaz, N. & Son, M. 2018 Performance assessment of general circulation model in simulating daily precipitation and temperature using multiple gridded datasets. *Water* **10**, 1793.
- Khayyun, T. S., Alwan, I. A. & Hayder, A. M. 2020 Selection of suitable precipitation CMIP-5 sets of GCMs for Iraq using a symmetrical uncertainty filter. *IOP Conf. Series: Mat. Sci. Eng.* **671**, 1–12.
- Knutti, R. 2008 Why are climate models reproducing the observed global surface warming so well? *Geophys. Res. Lett.* **35**, L18704.
- Knutti, R. & Sedláček, J. 2013 Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Change* **3**, 369–373.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G. A. 2010 Challenges in combining projections from multiple climate models. *J. Climate* **23**, 2739–2758.
- Kohonen, T. 1989 *Self Organization and Associative Memory*. Springer-Verlag, Berlin, Germany.
- Latif, M., Hannachi, A. & Syed, F. 2018 Analysis of rainfall trends over Indo-Pakistan summer monsoon and related dynamics based on CMIP5 climate model simulations. *Int. J. Climatol.* **38** (S1), e577–e595.
- Le, T. & Bae, D. H. 2013 Evaluating the utility of IPCC AR4 GCMs for hydrological application in South Korea. *Water Res. Manage.* **27**, 3227–3246.
- Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Res. Res.* **35**, 233–241.
- Liu, Z., Xu, Z., Fu, G. & Yao, Z. 2012 Assessing the hydrological impacts of climate change in the headwater catchment of the Tarim River basin. *China. Hydrol. Res.* **44**, 834–849.
- Lutz, A. F., ter Maat, H. W., Biemans, H., Shrestha, A. B., Wester, P. & Immerzeel, W. W. 2016 Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *Int. J. Climatol.* **36**, 3988–4005.
- Macadam, I., Pitman, A. J., Whetton, P. H. & Abramowitz, G. 2010 Ranking climate models by performance using actual values and anomalies: implications for climate change impact assessments. *Geophys. Res. Lett.* **37**, L16704.
- Mahmood, R., Jia, S., Tripathi, N. & Shrestha, S. 2018 Precipitation extended linear scaling method for correcting gem precipitation and its evaluation and implication in the Transboundary Jhelum River Basin. *Atmosphere* **9**, 160.
- Mandal, S., Arunkumar, R., Breach, P. A. & Simonovic, S. P. 2019 Reservoir operations under changing climate conditions: hydropower-production perspective. *J. Water Res. Plan. Manage.* **145**. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001061](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001061).
- Mani, A. & Tsai, F. T. C. 2017 Ensemble averaging methods for quantifying uncertainty sources in modeling climate change impact on runoff projection. *J. Hydrol. Eng.* **22**. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001487](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001487).
- Mareschal, B. 1986 Stochastic multicriteria decision making and uncertainty. *Eur. J. Oper. Res.* **26**, 58–64.
- Marttunen, M., Lienert, J. & Belton, V. 2017 Structuring problems for multi-criteria decision analysis in practice: a literature review of method combinations. *Eur. J. Oper. Res.* **263**, 1–17.
- McMahon, T. A., Peel, M. C. & Karoly, D. J. 2015 Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation. *Hydrol. Earth Syst. Sci.* **19**, 361–377.
- Meher, J. K., Das, L., Akhter, J., Benestad, R. E. & Mezghani, A. 2017 Performance of CMIP3 and CMIP5 GCMs to simulate observed rainfall characteristics over the Western Himalayan region. *J. Climate* **30**, 7777–7799.
- Mehran, A., Kouchak, A. A. & Phillips, T. J. 2014 Evaluation of CMIP5 continental precipitation simulations relative to satellite based gauge-adjusted observations. *J. Geophys. Res. Atm.* **119** (4), 1695–1707.
- Mendlik, T. & Gobiet, A. 2016 Selecting climate simulations for impact studies based on multivariate patterns of climate change. *Clim. Change* **135**, 381–393.
- Miao, C., Duan, Q., Sun, Q., Huang, Y., Kong, D., Yang, T., Ye, A., Di, Z. & Gong, W. 2014 Assessment of CMIP5 climate models and projected temperature changes over northern Eurasia. *Environ. Res. Lett.* **9**, 055007.
- Minville, M., Brissette, F. & Leconte, R. 2010 Impacts and uncertainty of climate change on water resource management of the Peribonka River System (Canada). *J. Water Res. Plan. Manage.* **136**, 376–385.
- Moise, A. F. & Delage, F. P. 2011 New climate model metrics based on object-orientated pattern matching of rainfall. *J. Geophys. Res.* **116**, D12108.
- Mujumdar, P. P. & Nagesh Kumar, D. 2012 *Floods in A Changing Climate: Hydrologic Modeling*. International Hydrology Series, Cambridge University Press, Cambridge, UK.
- Mustafa, S. M. T., Hasan, M. M., Saha, A. K., Rannu, R. P., Uytven, E. V., Willems, P. & Huysmans, M. 2019 Multi-model approach to quantify groundwater-level prediction uncertainty using an ensemble of global climate models and multiple abstraction scenarios. *Hydrol. Earth Syst. Sci.* **23**, 2279–2303.
- Najafi, M. R., Moradkhani, H. & Jung, I. W. 2011 Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrol. Process.* **25** (18), 2814–2826.

- Niel, J. D., Uytven, E. V. & Willems, P. 2019 Uncertainty analysis of climate change impact on river flow extremes based on a large multi-model ensemble. *Water Res. Manage.* **33**, 4319–4333.
- Noor, M., Ismail, T., Shahid, S., Ahmed, K., Chung, E. S. & Nawaz, N. 2019 Selection of CMIP5 multi-model ensemble for the projection of spatial and temporal variability of rainfall in peninsular Malaysia. *Theor. Appl. Climatol.* **138**, 999–1012.
- Northrop, P. J. & Chandler, R. E. 2014 Quantifying sources of uncertainty in projections of future climate. *J. Climate* **27**, 8793–8808.
- Nourani, V., Paknezhad, N. J., Sharghi, E. & Khosravid, A. 2019 Estimation of prediction interval in ANN-based multi-GCMs downscaling of hydro-climatologic parameters. *J. Hydrol.* **579**, 124226.
- Ongoma, V., Chen, H. & Gao, C. 2019 Evaluation of CMIP5 twentieth century rainfall simulation over the equatorial East Africa. *Theor. Appl. Climatol.* **135**, 893–910.
- Opricovic, S. & Tzeng, G. H. 2004 Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. *Eur. J. Oper. Res.* **156**, 445–455.
- Osman, Y., Al-Ansari, N., Abdellatif, M., Aljawad, S. B. & Knutsson, S. 2014 Expected future precipitation in Central Iraq using LARS-WG stochastic weather generator. *Engineering* **6**, 948–959.
- Otieno, V. O. & Anyah, R. O. 2013 CMIP5 simulated climate conditions of the Greater Horn of Africa (GHA) part 1: contemporary climate. *Clim. Dyn.* **41**, 2081–2097.
- Pandey, B. K., Khare, D., Kawasaki, A. & Mishra, P. K. 2019 Climate change impact assessment on blue and green water by coupling of representative CMIP5 climate models with physical based hydrological model. *Water Res. Manage.* **33**, 141–158.
- Panjwani, S., Naresh Kumar, S., Ahuja, L. & Islam, A. 2019 Prioritization of global climate models using fuzzy analytic hierarchy process and reliability index. *Theor. Appl. Climatol.* **137**, 2381–2392.
- Pei, Z., Liu, J., Hao, F. & Zhou, B. 2019 FLM-TOPSIS: The fuzzy linguistic multiset TOPSIS method and its application in linguistic decision making. *Inform. Fus.* **45**, 266–281.
- Perez, J., Menendez, M., Mendez, F. J. & Losada, I. J. 2014 Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Clim. Dyn.* **43**, 2663–2680.
- Perkins, S. E., Pitman, A. J., Holbrook, N. J. & McAneney, J. 2007 Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate* **20**, 4356–4376.
- Pierce, D. W., Barnett, T. P., Santer, B. D. & Gleckler, P. J. 2009 Selecting global climate models for regional climate change studies. *Proc. Natl Acad. Sci. USA* **106**, 8441–8446.
- Ploukas, N. & Papathanasiou, J. 2019 A decision support system for multiple criteria alternative ranking using TOPSIS and VIKOR in fuzzy and nonfuzzy environments. *Fuz. Sets Sys.* **377**, 1–30.
- Prasanna, V. 2015 Regional climate change scenarios over South Asia in the CMIP5 coupled climate model simulations. *Met. Atm. Phys.* **127**, 561–578.
- Radić, V. & Clarke, G. K. C. 2011 Evaluation of IPCC models' performance in simulating late-twentieth-century climatologies and weather patterns over North America. *J. Climate* **24**, 5257–5274.
- Raghavan, S. V., Liu, J., Nguyen, N. S., Vu, M. T. & Liong, S. Y. 2018 Assessment of CMIP5 historical simulations of rainfall over Southeast Asia. *Theor. Appl. Climatol.* **132**, 989–1002.
- Raju, K. S. & Nagesh Kumar, D. 2007 Classification of Indian meteorological stations using cluster and fuzzy cluster analysis, and kohonen artificial neural networks. *Nordic Hydrol.* **38**, 303–314.
- Raju, K. S. & Nagesh Kumar, D. 2014a *Multicriterion Analysis in Engineering and Management*. Prentice Hall India, New Delhi, India.
- Raju, K. S. & Nagesh Kumar, D. 2014b Ranking of global climatic models for India using multicriterion analysis. *Clim. Res.* **60**, 103–117.
- Raju, K. S. & Nagesh Kumar, D. 2015a Ranking general circulation models for India using TOPSIS. *J. Water Clim. Change* **6**, 288–299.
- Raju, K. S. & Nagesh Kumar, D. 2015b Fuzzy approach to rank global climate models. *Advances in Intelligent Systems and Computing*, 415 Springer, pp. 53–61.
- Raju, K. S. & Nagesh Kumar, D. 2016 Selection of global climate models for India using cluster analysis. *J. Water Clim. Change* **7**, 764–774.
- Raju, K. S. & Nagesh Kumar, D. 2018 *Impact of Climate Change on Water Resources-Modeling Techniques and Case Studies*. Springer Climate Series, Springer, Singapore.
- Raju, K. S., Duckstein, L. & Arondel, C. 2000 Multicriterion analysis for sustainable water resources planning: a case study in Spain. *Water Res. Manage.* **14**, 435–456.
- Raju, K. S., Sonali, P. & Nagesh Kumar, D. 2017 Ranking of CMIP5 based global climate models for India using compromise programming. *Theor. Appl. Climatol.* **128**, 563–574.
- Reifen, C. & Toumi, R. 2009 Climate projections: past performance no guarantee of future skill? *Geophys. Res. Lett.* **36**, L13704.
- Rogers, M., Bruen, M. & Maystre, L. Y. 2000 *ELECTRE and Decision Support Methods and Applications in Engineering and Infrastructure Investment*. Kluwer Academic Publishers, Dordrecht, the Netherlands.
- Ruan, Y., Yao, Z., Wang, R. & Liu, Z. 2018 Ranking of CMIP5 GCM skills in simulating observed precipitation over the Lower Mekong Basin, using an improved score-based method. *Water* **10**, 1868.
- Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C. & Mote, P. W. 2013 Evaluation of CMIP5 20th century climate simulations

- for the Pacific Northwest USA. *J. Geophys. Res. Atmos.* **118**, 10884–10906.
- Saeed, F. & Athar, H. 2018 Assessment of simulated and projected climate change in Pakistan using IPCC AR4-based AOGCMs. *Theor. Appl. Climatol.* **134**, 967–980.
- Saki, M., Esfahani, M. T. & Soltani, S. 2018 A scenario-based modeling of climate change impacts on the aboveground net primary production in rangelands of central Iran. *Environ. Earth Sci.* **77**, 670.
- Salman, S. A., Shahid, S., Ismail, T., Ahmed, K. & Wang, X. J. 2018 Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. *Atmos Res.* **213**, 509–522.
- Salunke, P., Jain, S. & Mishra, S. K. 2019 Performance of the CMIP5 models in the simulation of the Himalaya-Tibetan Plateau monsoon. *Theor. Appl. Climatol.* **137**, 909–928.
- Saraiva, S., Meier, H. E. M., Andersson, H., Höglund, A., Dieterich, C., Gröger, M., Hordoir, R. & Eilola, K. 2019 Uncertainties in projections of the Baltic Sea ecosystem driven by an ensemble of global climate models. *Front. Earth Sci.* **6**, 1–18.
- Sarathi, P., Kumar, P. & Ghosh, S. 2016 Possible future rainfall over Gangetic Plains (GP), India, in multi-model simulations of CMIP3 and CMIP5. *Theor. Appl. Climatol.* **124**, 691–701.
- Sayadi, A., Beydokhti, N. T., Najarchi, M. & Najafzadeh, M. M. 2019 Investigation into the effects of climatic change on temperature, rainfall, and runoff of the Doroudzan catchment, Iran, using the ensemble approach of CMIP3 climate models. *Adv. Meteorol.* **2019** (1), 1–16.
- Schaller, N., Mahlstein, I., Cermak, J. & Knutti, R. 2011 Analyzing precipitation projections: a comparison of different approaches to climate model evaluation. *J. Geophys. Res.* **116**, D10118.
- Schepen, A. & Wang, Q. J. 2013 Toward accurate and reliable forecasts of Australian seasonal rainfall by calibrating and merging multiple coupled GCMs. *Monthly Weather Review* **141**, 4554–4563.
- Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Genio, A. D., Faluvegi, G., Friend, A. D., Hall, T. M., Hu, Y., Kelley, M., Kiang, N. Y., Koch, D., Lacis, A. A., Lerner, J., Lo, K. K., Miller, R. L., Nazarenko, L., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Shindell, D. T., Stone, P. H., Sun, S., Tausnev, N., Thresher, D. & Yao, M. S. 2006 Present-day atmospheric simulations using GISS model: comparison to in situ, satellite and reanalysis data. *J. Climate* **19**, 153–192.
- Schuster, Z. T., Potter, K. W. & Liebl, D. S. 2012 Assessing the effects of climate change on precipitation and flood damage in Wisconsin. *J. Hydrol. Eng.* **17**, 888–894.
- Sikder, S., Chen, X., Hossain, F., Roberts, J. B., Robertson, F., Shum, C. K. & Turk, F. J. 2016 Are general circulation models ready for operational streamflow forecasting for water management in the Ganges and Brahmaputra river basins? *J. Climate* **17**, 195–210.
- Smith, I. & Chandler, E. 2010 Refining rainfall projections for the Murray Darling Basin of south-east Australia– the effect of sampling model results based on performance. *Clim. Change* **102**, 377–393.
- Song, J. Y. & Chung, E. S. 2016 Robustness, uncertainty and sensitivity analyses of the TOPSIS method for quantitative climate change vulnerability: a case study of flood damage. *Water Res. Manage.* **30**, 4751–4771.
- Sood, A. & Smakhtin, V. 2015 Global hydrological models: a review. *Hydrol. Sci. J.* **60**, 549–565.
- Sreelatha, K. & Raj, P. A. 2020 Ranking of CMIP5-based global climate models using standard performance metrics for Telangana region in the southern part of India. *ISH J. Hyd. Eng.* <https://doi.org/10.1080/09715010.2019.1634648>
- Sridhar, V., Kang, H. & Ali, S. A. 2019 Human-induced alterations to land use and climate and their responses on hydrology and water management in the Mekong River basin. *Water* **11**, 1307.
- Su, F., Duan, X., Chen, D., Hao, Z. & Cuo, L. 2013 Evaluation of the Global Climate Models in the CMIP5 over the Tibetan Plateau. *J. Climate* **26**, 3187–3208.
- Sultana, R. & Choi, M. 2018 Sensitivity of streamflow response in the snow-dominated Sierra Nevada watershed using projected CMIP5 data. *J. Hydrol. Eng.* **23**, 05018015.
- Suppiah, R., Hennessy, K. J., Whetton, P. H., McInnes, K., Macadam, I., Bathols, J., Ricketts, J. & Page, C. M. 2007 Australian climate change projections derived from simulations performed for the IPCC 4th assessment report. *Australian Meteorological Magazine* **56**, 131–152.
- Tan, M. L., Ficklin, D. L., Ibrahim, A. L. & Yusop, Z. 2014 Impacts and uncertainties of climate change on streamflow of the Johor River Basin, Malaysia using a CMIP5 general circulation model ensemble. *J. Water Clim. Change* **5**, 676–695.
- Tavana, M., Mavi, R. K., Santos-Arteaga, F. J. & Doust, E. R. 2016 An extended VIKOR method using stochastic data and subjective judgments. *Comp. Indus. Eng.* **97**, 240–247.
- Tebaldi, C. & Knutti, R. 2007 The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. Roy. Soc. A, Math. Phys. Eng. Sci.* **365**, 2053–2075.
- Tian, Y., Xu, Y. P., Boojj, M. J. & Cao, L. 2015 Impact assessment of multiple uncertainty sources on high flows under climate change. *Hydrol. Res.* **47**, 61–74.
- Tian, Y., Xu, Y. P., Ma, C. & Wang, G. 2017 Modeling the impact of climate change on low flows in Xiangjiang river basin with Bayesian averaging method. *J. Hydrol. Eng.* **22**, 04017035.
- Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F. & Knutti, R. 2020 Past warming trend constrains future warming in CMIP6 models. *Sci. Adv.* **6** (12), eaaz9549.
- Vaghefi, S. A., Keykhai, M., Jahanbakhshi, F., Sheikholeslami, J., Ahmadi, A., Yang, H. & Abbaspour, K. C. 2019 The future of extreme climate in Iran. *Sci. Rep.* **9**, 1464.
- Vandana, K., Islam, A., Sarathi, P. P., Sikka, A. K. & Kapil, H. 2019 Assessment of potential impact of climate change on

- streamflow: a case study of the Brahmani River basin, India. *J. Water Clim. Change* **10** (3), 624–641.
- Venkataraman, K., Spandana, T., Medina, A. & Perry, J. 2016 21st century drought outlook for major climate divisions of Texas based on CMIP5 multimodel ensemble: implications for water resource management. *J. Hydrol.* **534**, 300–316.
- Walsh, J. E., Chapman, W. L., Romanovsky, V., Christensen, J. H. & Stendel, M. 2008 Global climate model performance over Alaska and Greenland. *J. Climate* **21**, 6156–6174.
- Wang, B., Liu, D. L., Macadam, I., Alexander, L. V., Abramowitz, G. & Yu, Q. 2016 Multi-model ensemble projections of future extreme temperature change using a statistical downscaling method in south eastern Australia. *Clim. Change* **138**, 85–98.
- Wang, B., Zheng, L., Liu, D. L., Ji, F., Clark, A. & Yu, Q. 2018 Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *Int. J. Climatol.* **38**, 4891–4902.
- Wang, H. M., Chen, J., Xu, C. Y., Chen, H., Guo, S., Xie, P. & Li, X. 2019 Does the weighting of climate simulations result in a more reasonable quantification of hydrological impacts? *Hydrol. Earth Syst. Sci. Discuss.* **23**, 4033–4050. <https://doi.org/10.5194/hess-2019-24>.
- Watterson, I. G., Bathols, J. & Heady, C. 2014 What influences the skill of climate models over the continents? *Bull. Am. Meteorol. Soc.* **95**, 689–700.
- Weiland, F. S., Van Beek, L. P. H., Weerts, A. H. & Bierkens, M. F. P. 2012 Extracting information from an ensemble of GCMs to reliably assess future global runoff change. *J. Hydrol.* **412–413**, 66–75.
- Wilby, R. L. & Harris, I. 2006 A framework for assessing uncertainties in climate change impacts: low flow scenarios for the river Thames, UK. *Water Res. Res.* **42**, W02419.
- Wu, Z., Ahmad, J. & Xu, J. 2016 A group decision making framework based on fuzzy VIKOR approach for machine tool selection with linguistic information. *Appl. Soft Comput.* **42**, 314–324.
- Wu, Y., Zhong, P., Xu, B., Zhu, F. & Fu, J. 2018 Evaluation of global climate model on performances of precipitation simulation and prediction in the Huaihe River basin. *Theor. Appl. Climatol.* **133**, 191–204.
- Xiong, W. & Huan, Q. H. 2010 A extended TOPSIS method for the stochastic multi-criteria decision making problem through interval estimation. In: *2nd International Workshop on Intelligent Systems and Applications*, 22–23 May 2010, Wuhan, China, pp. 1–4.
- Xuan, W., Ma, C., Kang, L., Gu, H., Pan, S. & Xu, Y. P. 2017 Evaluating historical simulations of CMIP5 GCMs for key climatic variables in Zhejiang Province, China. *Theor. Appl. Climatol.* **128**, 207–222.
- Yan, R., Huang, J., Wang, Y., Gao, J. & Qi, L. 2015 Modeling the combined impact of future climate and land use changes on streamflow of Xinjiang Basin, China. *Hydrol. Res.* **47**, 356–372.
- Yan, M., Wang, B., Liu, J., Zhu, A., Ning, L. & Cao, J. 2018 Understanding the Australian monsoon change during the Last Glacial Maximum with a multi-model ensemble. *Clim. Past.* **14**, 2037–2052.
- You, Q., Jiang, Z., Wang, D., Pepin, N. & Kang, S. 2018 Simulation of temperature extremes in the Tibetan Plateau from CMIP5 models and comparison with gridded observations. *Clim. Dyn.* **51**, 355–369.
- Zamani, R. & Berndtsson, R. 2019 Evaluation of CMIP5 models for west and southwest Iran using TOPSIS-based method. *Theor. Appl. Climatol.* **137**, 533–543.
- Zamani, R., Akhond-Ali, A., Roozbahani, A. & Fattahi, R. 2017 Risk assessment of agricultural water requirement based on a multi-model ensemble framework, southwest of Iran. *Theor. Appl. Climatol.* **129**, 1109–1121.
- Zhang, X. L. & Yan, X. D. 2016 Deficiencies in the simulation of the geographic distribution of climate types by global climate models. *Clim. Dyn.* **46**, 2749–2757.

First received 2 July 2019; accepted in revised form 7 April 2020. Available online 13 May 2020