ORIGINAL ARTICLE

Improving precipitation estimates for Turkey with multimodel ensemble: a comparison of nonlinear artificial neural network method with linear methods

Buket Mesta[1] · O. Burak Akgun[2] · Elcin Kentel[2]

Abstract

Ensemble analysis is proven to provide advantages in climate change impact assessment based on outputs from climate models. Ensembled series are shown to outperform single-model assessments through increased consistency and stability. This study aims to test the improvement of precipitation estimates through the use of ensemble analysis for south and southwestern Turkey which is known to have complex climatic features due to varying topography and interacting climate forcings. The analysis covers an evaluation of the performance of eight regional climate models (RCMs) from the EUR-11 domain available from the CORDEX database. The historical outputs are evaluated for their representativeness of the current climate of the Mediterranean region and its surroundings in Turkey through a comparison with long-term monthly precipitation time series obtained from ground-based precipitation observations by the use of statistical performance indicators and Taylor diagrams. This is followed by a comparative evaluation of three ensemble methodologies, simple average of the models, multiple linear regression for superensemble, and artificial neural networks (ANN). The analysis results show that the overall performance of ensembled time series is better compared to individual RCMs. ANN generally provided the best performance when all RCMs are used as inputs. Improvement in the performance of ensembling due to the use of nonlinear models is further confirmed by fuzzy inference systems (FIS). Both ANN and FIS generated monthly precipitation time series with higher correlations with those of observations. However, extreme events are poorly represented in the ensembled time series, and this may result in inefficiency in the design of various water structures such as spillways and storm water drainage systems that are based on high return period events.

Keywords Precipitation · RCM · Artificial neural network · Multiple linear regression

1 Introduction

The Mediterranean region is expected to be affected most by climate change impacts [17]. The Mediterranean basin including southern Europe, Anatolia, and the Middle East has a very complex topography resulting in a milder climate in coastal areas while inland Anatolia experiences extreme weather conditions [48]. Moreover, the climate processes in that region are affected by teleconnections and regional processes, the interactions of which are not well understood [22, 45]. The local processes and the teleconnections (e.g., North Atlantic Oscillation and North Sea-Caspian) interacting with those, and the effects of the varying topography and the Mediterranean Sea itself create large inter-annual and intra-annual variability as well as significant spatial variability depending on local properties in the region [5, 8, 22, 36, 43, 45,].

High-resolution regional climate models (RCMs) are used for local-scale assessments of climate change impacts including impacts on hydrology [35] and parameters such as precipitation that are mainly controlled by local processes, particularly for regions with a complex orography [34, 46, 52]. However, RCM outputs may contain

✉ Elcin Kentel
    ekentel@metu.edu.tr

[1] Department of Earth System Science, Middle East Technical University, Ankara, Turkey

[2] Water Resources Laboratory, Department of Civil Engineering, Middle East Technical University, K4 Bldg., 06800 Ankara, Turkey

systematic errors or biases [20, 51] due to the boundary conditions or imperfect conceptualization, discretization, and spatial averaging within grid cells [15, 24, 33, 55]. Furthermore, studies indicate uncertainties and significant inter-model variability for RCM outputs [33, 52] which are even more pronounced for regions with complex terrain features [13]. Various factors such as the multiplicity of model designs or assumptions amplify the variability and divergence in outputs, particularly for evaluations based on single models. Hence, the uncertainties and significant divergence in precipitation outputs by various climate models create challenges in impact assessment and planning for adaptation.

To overcome the limitations of the single-model assessment in dealing with uncertainties, the multimodel ensemble (MME) assessment is preferred over single-model assessments. MME approach has been used in various fields such as finance [60], ecology [2], atmospheric air quality [16], streamflow forecasting [1], weather prediction [31, 47], and for climate simulations [29]. Studies have shown that ensemble analysis helps to improve the statistical indicators relative to single-model analysis, provides better consistency and reliability in simulations, and enables a better understanding of uncertainties [21, 52, 54, 56]. The MME approach overcomes the disadvantages of single-model-based assessments by using a set of models with different mechanisms that eliminate or reduce poorly represented processes through better represented processes from other models in the ensemble group [28, 54]. Thus, the main motivation of this study is to investigate the performances of ensembled versus single-model estimates of monthly precipitation time series and potential improvements in enmsembling that can be achieved due to the use of nonlinear approaches instead of linear ones.

The most common approach in MME is ensemble averaging. In this approach, weights are assigned to each member of the ensemble either equally to obtain a multi-model simple average (will be referred to as the simple average of the models, SAM) or according to some pre-defined criteria (e.g., weights for relative model performances or weights defined by certain statistical techniques) to obtain a multimodel weighted average [12, 18, 19, 54].

A linear statistical MME approach, superensemble (SE) method uses multiple linear regression (MLR) to assign a suitable weight for each model to build an ensemble model with a minimized squared deviation from observations [30–32, 50]. Studies verify that SE provides a better simulation efficiency than single models and SAM [11, 30–32, 59] by enabling a bias correction in ensembling. Comparison of SE with "individual bias removed ensemble mean" is reported to verify the better performance of SE providing a collective bias removal by

assigning different weights on member models depending on the degree of convergence to observations [30, 31, 59]. Sirdas et al. [49] tested the efficiency of SE, for the Euro-Mediterranean region, by the use of monthly and seasonal forecasts of precipitation, sea surface temperature (SST), and surface air temperature (SAT) from 13 global climate models (GCMs). The outputs obtained with the synthetic SE method for winter SST and SAT provided satisfactory results concerning the root-mean-square error (RMSE) and anomaly correlation (AC) coefficient values.

Another methodology for MME is the artificial neural network (ANN) approach [6, 7, 14, 26] which is also frequently used for statistical downscaling of climate models [10, 23, 37, 40–42, 57]. For example, Campozano et al. [10] compared the performances of two artificial intelligence methods, namely, ANN and least squares support vector machines in downscaling of monthly precipitation, and the authors concluded that they performed equally well. Regarding the use of neural networks in MME, Boulanger et al. [6, 7] worked on the ensembling of seven atmospheric-ocean global circulation models (AOGCMs). They used ANN and Bayesian statistics to attain MME with higher efficiency than single models for the simulations of temperature and precipitation in South America. Krasnopolsky and Lin [26] developed an ANN model for MME of 24-h precipitation forecasts and demonstrated improvements in the forecasts over continental US. Furthermore, their findings verified superior results of the nonlinear ANN approach compared to the linear approaches (including MLR) in MME for forecasting. In a more recent study, Fan et al. [14] used ANN to improve the climate forecast system (CFS) of the US National Oceanic and Atmospheric Administration (NOAA) for week 3–4 precipitation and 2-m air temperature forecasts. The study developed an ANN model for MME by the use of a set of predictors and the predictand. The study results revealed that MLR and ANN provided superior results over bias-adjusted CFS, and MME through ANN improved forecasts better than MLR in many aspects. However, despite improvements in forecasting skills through MME with ANN, some weaknesses regarding RMSE and AC skills are still notified [14].

Studies on local impacts of climate change regarding hydrology, crop yields, and reservoir inflows for various basins in Turkey use the benefit of the ensemble approach by including the SAM in the analysis [27, 41, 44]. In their study, on climate change in Gediz Basin in western Turkey, Okkan and Kirdemir [42] used a multi-GCM ensemble based on Bayesian model averaging (BMA) for 12 GCMs from Coupled Model Inter-comparison Project Phase 5 (CMIP5) after statistical downscaling with ANN and least squares support vector machine methods. In another study, Cakir et al. [9] used historical records from 50

meteorological stations across Turkey to test the ANN ensemble approach for temperature. Their study verified that the ANN ensemble generates more accurate results compared to the simple bias-corrected ensemble mean. On the other hand, the study does not provide any evaluation for a comparison of the nonlinear ANN approach with the linear MLR approach and concludes the necessity of comparison for future studies.

To the best of our knowledge, the use of ANN for MME for the Mediterranean region is very limited, and only a few studies are published on the use of ANN for ensemble analysis for Turkey. Moreover, for Turkey, there are still not many studies using multiple high-resolution CORDEX RCMs to evaluate relative model performances in comparison with MME. Mesta and Kentel [39] in their study evaluated the efficiency of the ensembles of raw and bias-adjusted RCMs in projections on precipitation for Mediterranean Turkey in comparison with the single models and gave evidence of certain advantages and weaknesses of linear ensembling methodology, superensembling. The objective of this study is to evaluate the efficiency of ensemble analysis with the nonlinear ANN approach and to compare it with conventional linear MME methodologies, namely, SAM and SE. The study uses CORDEX data of RCM outputs. The evaluation is done through comparison with reference data from local stations in south and southwestern Turkey that provide long-term daily precipitation time series. The comparison of the efficiencies of ANN, SAM, and SE showed that for most stations, correlation and normalized root-mean-square difference (RMSD) values improved by MME compared to the best-performing individual RCM, and the ANN approach showed relatively best skills, in general. Similar results are obtained when ensembling is carried out using another nonlinear approach, namely, fuzzy inference systems (FIS) in terms of correlations and RMSDs. However, certain weaknesses to represent the variability of precipitation due to the reduction of the extreme or low-frequency events are also detected for ANN, SAM, and SE through the analysis.

2 Case study

2.1 Study area

In this study, the historical simulation outputs of eight RCMs are used for the MME analysis. This paper makes a comparison of the efficiency of the linear and nonlinear ANN and FIS ensembling methodologies relative to the individual raw RCM outputs regarding the replication of the historical precipitation. For the analysis, the observed and simulated daily total precipitation for 14

meteorological stations (MSs) throughout south and southwestern Turkey including the Mediterranean coast and its hinterland are used (Fig. 1). The MSs are selected from different climatologic regions (i.e., coastal and inland Mediterranean, inland Aegean, and Central Anatolian regions) and major basins of Turkey to assess the potential effect of spatial differences. The characteristics of these MSs are given in Table 1.

2.2 Data

For the ensemble analysis, the daily precipitation data for the closest grid to the MSs (see Table 1, the last column gives the distance between MS and the center of the nearest model grid) are extracted from the historic simulation results of eight RCMs for the EUR-11 domain (with the horizontal grid spacing of 0.11°) available from the CORDEX database [61]. RCMs used in this study together with their driving GCMs are given in Table 2. All of the eight RCMs have daily historical simulation outputs for the period between 1951 and 2005. The historical daily total precipitation time series from 1951 to 2005 are extracted for the relevant modeling grids by the use of the R code developed by Kentel et al. [25].

The daily time series are converted to monthly mean precipitation time series to be used in ensembling. These will be referred to as RCM time series (RCMTS) from hereafter. The observed daily total precipitation data for the same historical period (1951 to 2005) from 14 meteorological stations operated by the Turkish State Meteorological Services are used as the reference data in this study. The observed precipitation is used for the training of SE, ANN, and FIS models and for the testing of the simulation efficiency of the ensembled time series. For this purpose, the observation data are initially screened through a quality check (QC) process (see Fig. 2). In the QC, the observation data are initially screened for continuous data gaps. The months, that are identified to have more than 10 days of missing records, are removed from the dataset. The observed monthly mean precipitation time series is formed by the use of the remaining months. As a second stage of the QC, the data for the months with less than 0.1-mm precipitation record outside the dry season raised an error flag and are excluded from the data series as well. In the study region, July, August, and September are the main dry months during which monthly mean precipitation is likely to drop to distinctly low levels or even to zero. Hence, annual precipitation is mostly received during the remaining months. Therefore, monthly mean precipitation records of zero for the months outside of the dry season raised an error flag. After the removal of the months with long data gaps and monthly means with an error flag, the final quality checked dataset (QCD) of the observed
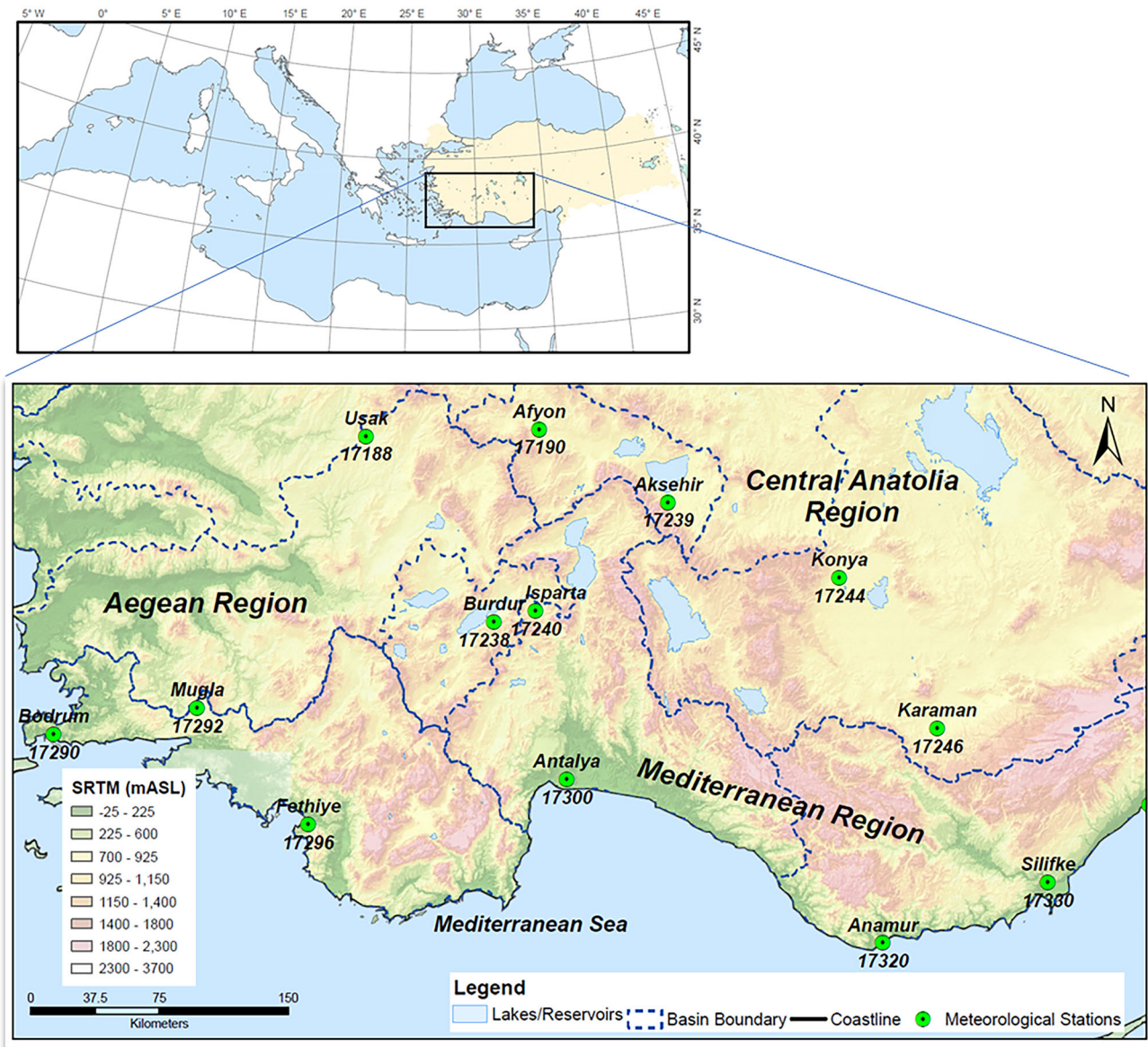
Fig. 1 The study region and the locations of meteorological stations

precipitation is obtained. The time series of observed precipitation generated using QCD will be referred to as the observed time series (OTS) from hereafter. The OTS is used for the application and testing of ensemble methodologies.

3 Methodology

3.1 Formation of training and validation datasets

The analysis includes a comparison of the efficiencies of linear methodologies with the nonlinear ANN approach for

the ensembling of the climate simulation outputs to replicate the observed precipitation. As the second stage, the advantage of the use of a nonlinear model for ensembling is further demonstrated by the FIS for the best-performing combination of RCMTS as well. The use of linear and nonlinear methodologies for ensembling of eight RCMTS is based on a benchmark with the aforementioned OTS. Training and validation datasets that are used in SE and ANN methodologies are formed from the randomly selected data points of eight RCMTS and OTS. The relevant datasets include the same data points (from the same grid point and the same point in time) of all time series. To test the robustness of the linear and nonlinear approaches, 5-fold cross-validation (80% for training and 20% for

Table 1 The characteristics of meteorological stations (MSs) used in this study

MS ID*	MS name	Elevation (m ASL)	Latitude (DD)	Longitude (DD)	Distance to the sea (km)	Distance to the center of the nearest RCM grid (km)
17290	BODRUM	26	37.03	27.44	0	4.4
17296	FETHIYE	3	36.63	29.12	0	5.3
17320	ANAMUR	2	36.07	32.86	0	3.8
17340	MERSIN	7	36.78	34.6	0	4.8
17300	ANTALYA	64	36.91	30.8	6	2.4
17330	SILIFKE	10	36.38	33.94	7	3.7
17292	MUGLA	646	37.21	28.37	17	3.3
17238	BURDUR	957	37.72	30.29	100	5.6
17240	ISPARTA	997	37.78	30.57	100	5.6
17246	KARAMAN	1018	37.19	33.22	115	4.2
17244	KONYA	1031	37.98	32.57	165	2.2
17239	AKSEHIR	1002	38.37	31.43	170	4.3
17188	USAK	919	38.67	29.4	195	3.9
17190	AFYON	1034	38.74	30.56	205	5.6

*MS ID: Meteorological station ID used by Turkish State Meteorological Services

Table 2 Overview of RCMs used in this study

Driving GCM	RCM		Model ID*
	Name	Institution	
CNRM-CM5 (CNRM-CERFACS)	CCLM4-8–17	Climate Limited-area Modelling Community (CLM-Community)	M1
CNRM-CM5 (CNRM-CERFACS)	ALADIN53	Centre National de Recherches Météorologiques (CNRM)	M2
EC-EARTH (ICHEC)	CCLM4-8–17	Climate Limited-area Modelling Community (CLM-Community)	M3
EC-EARTH (ICHEC)	RACMO22E	Royal Netherlands Meteorological Institute (KNMI)	M4
EC-EARTH (ICHEC)	HIRHAM5	Danish Meteorological Institute (DMI)	M5
CM5A-MR (IPSL)	WRF331F	Institut Pierre-Simon Laplace (IPSL-INERIS)	M6
HadGEM2-ES (MOHC)	CCLM4-8–17	Climate Limited-area Modelling Community (CLM-Community)	M7
HadGEM2-ES (MOHC)	RACMO22E	Royal Netherlands Meteorological Institute (KNMI)	M8

*Climate model ID used in this study

validation) is used. Training and subsequent validation for the ensembled datasets are done for each fold separately. Log-transformation and normalization are applied to datasets before ensembling to improve efficiency.

Ensembling is carried out for two different sets of inputs. The first set is composed of all eight RCMTS (referred to as AllMs in Fig. 2). The second set is composed of three best-performing RCMTS selected among available eight RCMTS (referred to as SMs in Fig. 2). However, as mentioned above, due to the multidimensional variability of climate outputs, the selection of the best or most representative model is a challenging task. Although there is an ample number of metrics that provide an evaluation from different aspects of climate features [28], no standard set of performance metrics is defined for specific

assessment purposes [4]. Taylor diagram [53] is among the most common means of depicting the relative performance skill of models. Thus, in this study, Taylor diagrams are used to select the three best-performing RCMTS (see Fig. 2) for each MS.

As an example, Fig. 3 provides the Taylor diagram for Usak MS (17188) comparing the OTS with eight RCMTS for the entire period between 1951 and 2005. Each RCMTS is represented by a symbol on the Taylor diagram. The correlation coefficient, which is represented by the azimuthal angle, the centered RMSD, which is represented by the distance from the OTS (i.e., the point marked with a plus in a circle on the x-axis) to the symbol of the RCMTS, and the standard deviation (SD), which is represented by the radial distance from the origin, are plotted on the
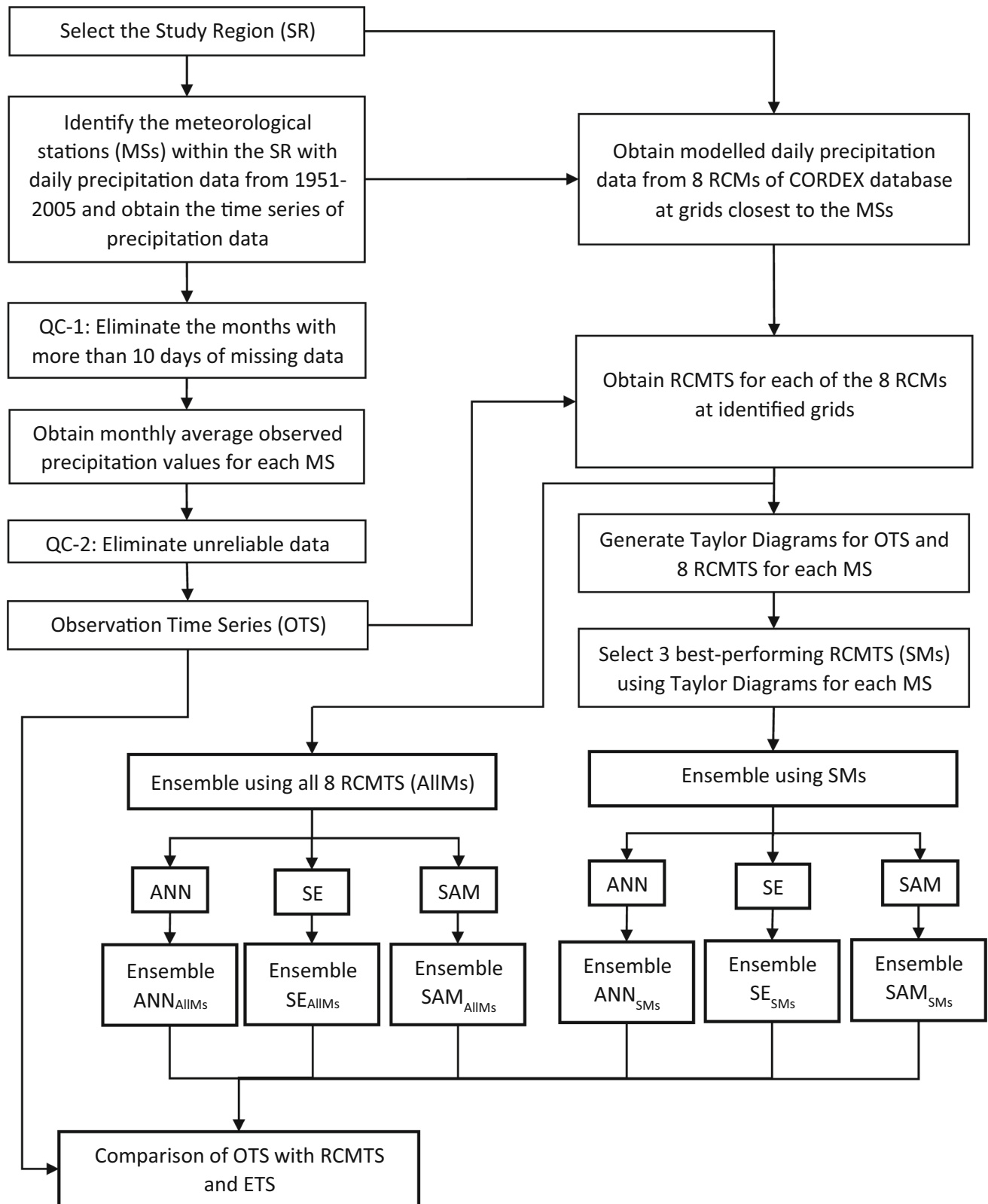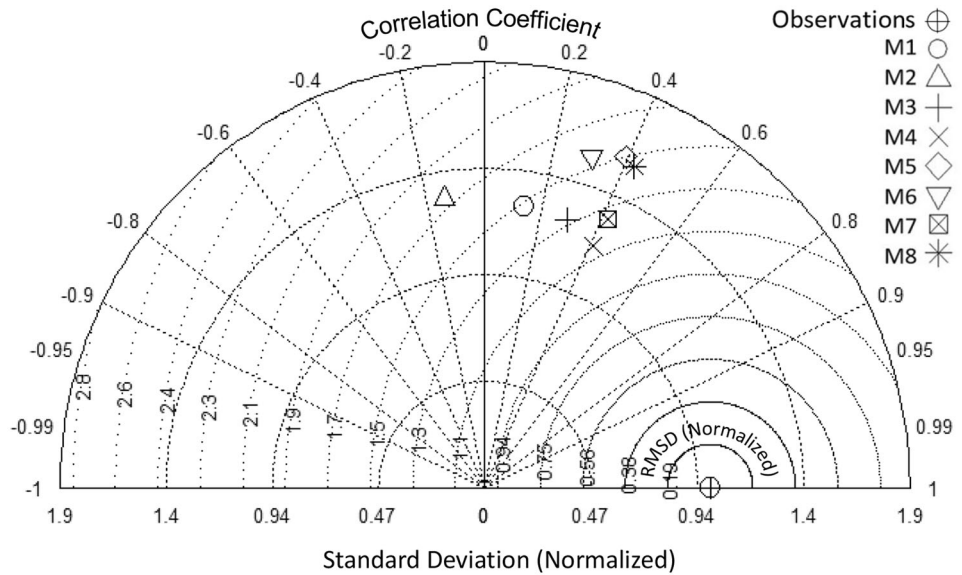
Fig. 2 Flowchart of the methodology

Fig. 3 Taylor diagram of Usak MS (17188)

Taylor diagram. The position of each symbol indicates how similar that RCMTS and OTS are. Euclidean distance from the OTS to the symbol of the RCMTS is calculated, and three best-performing RCMs are identified as the models having three shortest Euclidean distances. For Usak MS, three best-performing RCMs are M4, M7, and M3. Table 3 gives three best-performing RCMs for each MS in the descending order of simulation skills based on the corresponding Taylor diagrams. In this study, two types of MMEs (i.e., MME of all RCMs and MME of three best-performing RCMs) are generated by using ANN, SE, and SAM methods as shown in Fig. 2.

Table 3 Selected models (SMs) for meteorological stations (MSs)

MS ID	MS name	SMs		
17290	BODRUM	M7	M3	M4
17296	FETHIYE	M5	M4	M2
17320	ANAMUR	M5	M4	M8
17340	MERSIN	M4	M6	M5
17300	ANTALYA AIRPORT	M2	M4	M6
17330	SILIFKE	M5	M4	M6
17292	MUGLA	M7	M3	M4
17238	BURDUR	M5	M4	M8
17240	ISPARTA	M4	M3	M7
17246	KARAMAN	M5	M4	M8
17244	KONYA AIRPORT	M5	M4	M8
17239	AKSEHIR	M4	M3	M8
17188	USAK	M4	M7	M3
17190	AFYON	M4	M8	M5

3.2 Artificial neural networks for ensembling

The ensembling process used to simulate monthly mean precipitation values at a selected grid i using RCMTS at the same grid using all RCMs (AllMs) or three selected RCMs (SMs) can be mathematically represented as follows:

$$\text{ANN}_{i,t}^{\text{AllMs}} = f\left(\text{RCM}_{i,t}^1, \text{RCM}_{i,t}^2, \ldots, \text{RCM}_{i,t}^8\right), \forall i, \forall t \quad (1)$$

$$ANN_{i,t}^{SMs} = f\left(\text{RCM}_{i,t}^1, \text{RCM}_{i,t}^2, \text{RCM}_{i,t}^3\right), \forall i, \forall t \quad (2)$$

where $\text{ANN}_{i,t}^{\text{AllMs}}$ is the ANN ensemble generated using all RCMs at grid i for month t, and $\text{RCM}_{i,t}^j$ is the jth RCMs at grid i for month t. When all models are used in generating the ensemble, $j = 1, 2, \ldots, 8$. $ANN_{i,t}^{SMs}$ is the ANN ensemble generated using three selected RCMs based on Taylor diagrams at grid i for month t. For each MS, selected three RCMs are given in Table 3. In this study, 14 MSs are used; thus, $i = 1, 2, \ldots, 14$. The simulation period is from 1951 to 2005, but the months that could not pass the quality check are eliminated, thus ensembles are calculated for $t = 1, 2, \ldots, T$ where T is at most 660.

As shown in Fig. 2, two ANN models—one using outputs of eight RCMs (i.e., eight input nodes) and the other using outputs of three selected RCMs (i.e., three input nodes) as inputs—are built for each MS. One hidden layer with six and two hidden nodes for AllMs and SMs, respectively, is used in all ANN models in this study. The number of hidden nodes is selected after multiple trials to avoid overfitting. The architecture for the ANN model built for Anamur MS (17320) which uses all RCMs as inputs is shown in Fig. 4. In this study, the sigmoid activation function is used in all ANN models, and all data are scaled
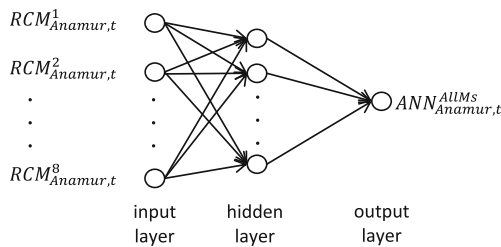
Fig. 4 ANN model architecture for Anamur MS (17320)

to the 0.1–0.9 range. The momentum, the learning rate, and maximum iterations are identified through trial and error. Various combinations of momentum and learning rates from the ranges of 0.001–0.9 and 0.001–0.5, respectively, are tested. Finally, the momentum and learning rates are selected as 0.5 and 0.05, respectively. The maximum number of iterations is set to 2000 for all ANN models. Larger number of iterations have been tested; however, no significant improvement is achieved when all cross-validation runs are considered.

3.3 Simple average of the models for ensembling

SAM of all RCMs (Eq. 3) and three selected RCMs (Eq. 4) are calculated based on the arithmetic average of the anomalies simulated by the RCMs by the use of the following equations [11]:

$$\text{SAM}_{i,t}^{\text{AllMs}} = \overline{O_i} + \frac{1}{8}\sum_{j=1}^{8}\left(\text{RCM}_{i,t}^{j} - \overline{\text{RCM}_i^{j}}\right), \forall i, \forall t \tag{3}$$

$$\text{SAM}_{i,t}^{\text{SMs}} = \overline{O_i} + \frac{1}{3}\sum_{j=1}^{3}\left(\text{RCM}_{i,t}^{j} - \overline{\text{RCM}_i^{j}}\right), \forall i, \forall t \tag{4}$$

where $\text{SAM}_{i,t}^{\text{AllMs}}$ is the SAM ensemble generated using all RCMs at grid i for month t, $\text{SAM}_{i,t}^{\text{SMs}}$ is the SAM ensemble generated using three best-performing RCM outputs at grid i for month t, $\overline{\text{RCM}_i^{j}}$ is the climatology determined by RCMj, and $\overline{O_i}$ is the mean observation or observed climatology value at grid i.

$$\overline{\text{RCM}_i^{j}} = \frac{1}{T}\sum_{t=1}^{T}\text{RCM}_{i,t}^{j}, \forall i, \forall j$$

$$\overline{O_i} = \frac{1}{T}\sum_{t=1}^{T}O_{i,t}, \forall i$$

where $O_{i,t}$ is the observed precipitation at grid i at month t.

3.4 Multiple linear regression for ensembling

The superensemble method suggested by Krishnamurti et al. [30, 31] uses the MLR approach for ensembling based on the following equations:

$$\text{SE}_{i,t}^{\text{AllMs}} = \overline{O_i} + \sum_{j=1}^{8}a_j\left(\text{RCM}_{i,t}^{j} - \overline{\text{RCM}_i^{j}}\right), \forall i, \forall t \tag{5}$$

$$\text{SE}_{i,t}^{\text{SMs}} = \overline{O_i} + \sum_{j=1}^{3}a_j\left(\text{RCM}_{i,t}^{j} - \overline{\text{RCM}_i^{j}}\right), \forall i, \forall t \tag{6}$$

where $\text{SE}_{i,t}^{\text{AllMs}}$ is the SE generated using all RCMs at grid i for month t, $\text{SE}_{i,t}^{\text{SMs}}$ is the SE generated using three selected RCMs at grid i for month t, and a_i is the weight of anomaly simulated by RCMj optimized for the training period to minimize the squared difference between observed and modeled precipitation at grid i based on the conventional MLR. In order to enable benchmarking between three methods, the same training and validation datasets are used for all three ensemble approaches for each fold.

4 Results and discussion

In the tables and figures given in this section, MSs are sorted based on their proximity to the sea, in order to observe the potential spatial effect. 5-fold validation is used to test the performance of each approach (i.e., each data point is used in one of the validation datasets). The comparison of the performance ranges of ensembles with the performance range of individual RCMs regarding five validation datasets is given in Figs. 5 and 6 for SMs and AllMs, respectively. The performance ranges in Figs. 5 and 6 represent the range of correlation (Pearson correlation coefficient) values of each ensemble method (ANN, SE, and SAM) with the observed for SMs and AllMs, respectively.

As shown from Figs. 5 and 6, ensembling with ANN, SE, or SAM improves the correlations for most of the validation datasets. Moreover, more consistent estimates are obtained from ensembling compared to the single RCMs in terms of correlations. Better improvement is achieved when all the models are used in the ensemble. Although the model performances generally are better for MSs closer to the sea, variation of the correlation of the models varies regardless of spatial characteristics of the MSs.

In the rest of this study, the ensembled time series (ETS) is used to evaluate the ensemble performance of each approach (i.e., ANN, SE, and SAM). The ETS is generated by combining together the validation datasets of fivefold.

Fig. 5 Comparison of correlation performances when selected models (SMs) are used as inputs
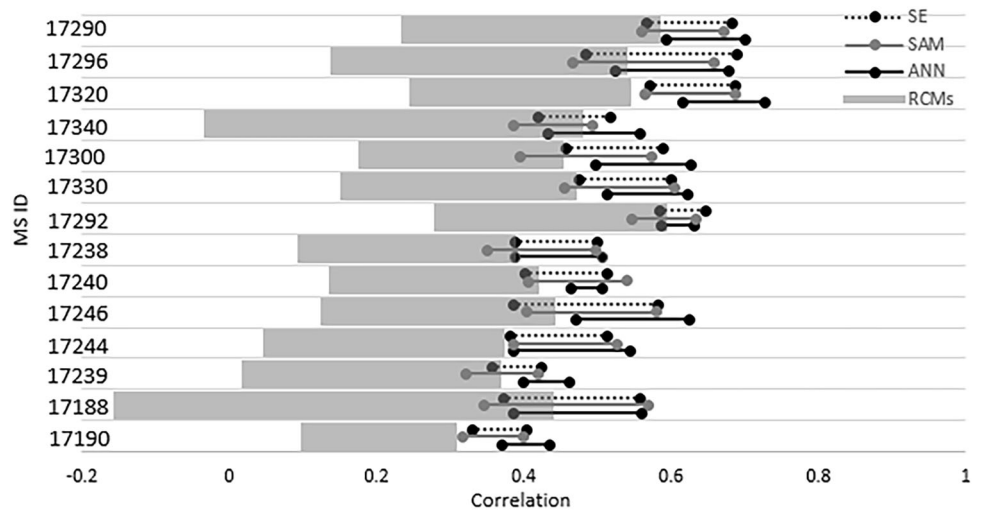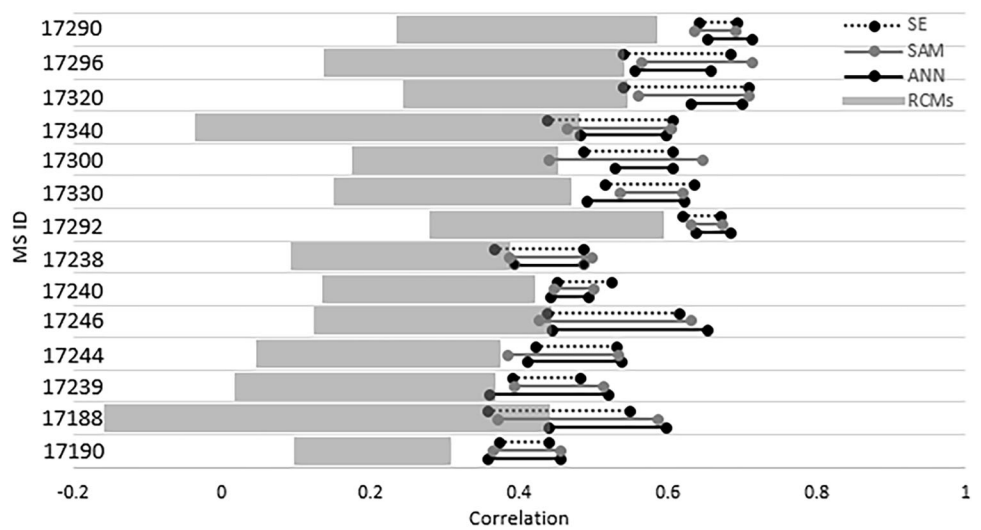

Fig. 6 Comparison of correlation performances when all models (AllMs) are used as inputs

The correlation, RMSD, and percent bias (PBIAS) values of the best- and the worst-performing RCMs, the ETS obtained from ANN, SE, and SAM approaches with selected, and all models are given in Tables 4, 5, and 6, respectively. Stations are listed according to their approximate distances from the sea (see Table 3), and the best-performing model for each station is given in bold.

As shown in Table 4, the ETS obtained from ANN_AllMs generally resulted in the best correlations, except for two stations (i.e., 17238 and 17240) for which ANN_SMs, and for two other stations (i.e., 17340 and 17330) for which SAM_AllMs provided the best correlations. The performances of both ANN models (selected and all) are very similar to each other. For all stations, other than 17238, 17239, and 17240, when all models are used, correlations improved but less than 5%. On the other hand, for SE, correlation values improved by more than 5% for eight of the stations when all models are used instead of three best-performing RCMs as inputs. The improvement is

more pronounced for SAM. When all models are used instead of three selected models, correlations of 11 stations improved by more than 5%. In fact, for 17340 and 17239, the improvement was more than 20%. Thus, it can be concluded that generally, performance in terms of correlations is better when all models instead of the selected three best-performing models are used as inputs and SAM benefits the most from ensembling all available RCMs.

For correlation, the ETS generated using all models or selected three models performs better than the best RCMTS for all stations. As can be seen in the third and fifth columns of Table 4, RCMTS has a wide range of correlation performances. The difference in model performance according to the atmospheric instability on local scale is also reported by Baghanam et al. [3]. Thus, the utilization of a single RCM for future precipitation projections is prone to high uncertainty. Moreover, stations closer to the sea generally perform better compared to the inland ones (i.e., correlations have a decreasing trend from

Table 4 Correlation values of the best and the worst RCMTS and the ETS obtained from ANN, SE, and SAM approaches

MS ID*	RCMTS				ETS					
	Best RCM		Worst RCM		ANN		SE		SAM	
	Model	Value	Model	Value	SMs	AllMs	SMs	AllMs	SMs	AllMs
17290	M4	0.57	M1	0.25	0.64	**0.67****	0.62	0.66	0.61	0.66
17296	M5	0.52	M2	0.16	0.60	**0.62**	0.57	0.61	0.54	0.61
17320	M5	0.53	M2	0.26	0.66	**0.66**	0.61	0.62	0.60	0.62
17340	M7	0.46	M2	-0.02	0.51	0.52	0.47	0.51	0.45	**0.54**
17300	M4	0.43	M6	0.19	0.54	**0.56**	0.51	0.55	0.48	0.55
17330	M8	0.45	M2	0.17	0.56	0.57	0.54	0.56	0.51	**0.58**
17292	M8	0.57	M2	0.30	0.61	**0.66**	0.61	0.64	0.59	0.65
17238	M8	0.37	M6	0.11	**0.47**	0.45	0.45	0.43	0.42	0.45
17240	M4	0.40	M2	0.15	**0.47**	0.46	0.43	0.47	0.46	0.45
17246	M7	0.42	M2	0.14	0.54	**0.55**	0.48	0.50	0.47	0.51
17244	M7	0.35	M2	0.07	0.46	**0.48**	0.44	0.47	0.44	0.46
17239	M7	0.35	M2	0.04	0.41	**0.44**	0.39	0.44	0.37	0.45
17188	M8	0.42	M2	-0.14	0.51	**0.53**	0.51	0.49	0.48	0.53
17190	M3	0.29	M2	0.12	0.39	**0.40**	0.36	0.40	0.37	0.40

*MSs are given in the order of proximity to the sea, **the best-performing model for each station is given in bold

Table 5 RMSD values of the best and the worst RCMTS and the ETS obtained from the ANN, SE, and SAM approaches

MS ID*	RCMTS				ETS					
	Best RCM		Worst RCM		ANN		SE		SAM	
	Model	Value	Model	Value	SMs	AllMs	SMs	AllMs	SMs	AllMs
17290	M4	2.26	M6	2.94	2.00	**1.85****	1.97	1.94	1.99	1.89
17296	M5	2.76	M7	4.10	2.49	**2.43**	2.56	2.51	2.62	2.45
17320	M8	3.12	M7	3.86	2.68	**2.66**	2.79	2.81	2.81	2.75
17340	M4	2.25	M1	3.06	1.96	1.93	2.01	1.96	2.05	**1.91**
17300	M4	4.16	M7	6.71	3.79	**3.67**	3.84	3.79	3.91	3.70
17330	M5	2.15	M8	2.81	1.83	**1.81**	1.89	1.87	1.95	1.82
17292	M7	3.59	M6	6.53	3.13	**2.93**	3.09	3.05	3.13	2.96
17238	M4	1.11	M1	1.61	0.96	0.96	**0.93**	0.95	1.00	0.93
17240	M4	1.72	M2	3.12	1.32	1.32	1.34	**1.31**	1.41	1.31
17246	M8	0.86	M1	1.19	0.70	**0.69**	0.72	0.71	0.74	0.71
17244	M5	0.83	M2	1.27	0.67	**0.66**	0.69	0.67	0.72	0.68
17239	M4	1.44	M2	2.18	1.25	1.23	1.27	1.24	1.37	**1.22**
17188	M4	1.49	M2	2.31	1.09	**1.07**	1.09	1.11	1.20	1.07
17190	M4	0.96	M2	2.21	0.79	0.79	0.80	0.80	0.85	**0.78**

*MSs are given in the order of proximity to the sea, **the best-performing model for each station is given in bold

the top to the bottom in Table 4). Milder climatologic conditions experienced in Mediterranean coasts compared to the inland Anatolia might be the reason for it. As shown in Fig. 6, the performance of the ETS is influenced by the performances of the RCMTS in the ensemble set. Hence,

using better-performing RCMs as inputs results in the ETS having better performance in terms of correlations. Percent improvements in correlation, when the ETS is evaluated with respect to the best RCMTS, are given in columns 2, 3,

Table 6 PBIAS values of the best and the worst RCMTS and the ETS obtained from the ANN, SE, and SAM approaches

| MS ID* | RCMTS | | | | ETS | | | | | |
| | Best RCM | | Worst RCM | | ANN | | SE | | SAM | |
	Model	Value	Model	Value	SMs	AllMs	SMs	AllMs	SMs	AllMs
17290	M5	**4.07****	M6	− 48.64	25.12	14.20	17.91	21.11	13.33	15.40
17296	M6	**− 3.13**	M7	− 47.57	18.30	15.11	19.71	22.84	21.57	18.56
17320	M1	**− 0.06**	M4	47.21	21.10	19.15	17.47	23.91	15.74	16.88
17340	M3	**6.61**	M6	45.79	20.77	19.85	21.27	20.72	15.27	19.47
17300	M3	**− 3.81**	M7	− 66.78	26.61	23.10	25.59	29.06	26.54	22.71
17330	M4	**-0.37**	M8	− 52.73	9.11	6.94	19.18	20.69	13.91	17.90
17292	M2	**− 2.14**	M6	− 100.89	23.48	17.03	21.14	22.10	13.48	18.63
17238	M2	− 7.05	M1	− 52.07	23.01	22.47	12.31	9.33	**5.33**	11.97
17240	M3	− 16.99	M6	− 121.52	20.37	17.64	12.05	9.11	**1.39**	12.54
17246	M6	**0.40**	M5	57.04	18.06	15.51	10.43	10.42	6.65	11.44
17244	M8	− 3.55	M5	24.21	**-1.93**	1.97	10.85	9.06	4.82	10.33
17239	M8	**1.58**	M2	− 34.38	13.28	13.48	14.21	10.03	4.83	13.03
17188	M4	− 6.73	M6	− 79.15	9.33	9.80	12.68	10.72	**2.56**	12.09
17190	M8	− 2.88	M2	− 116.00	12.56	11.72	9.87	7.16	**4.80**	10.34

*MSs are given in the order of proximity to the sea, **the best-performing model for each station is given in bold

Table 7 Percent changes in correlation, RMSD, and PBIAS values of the ETS relative to the best RCMTS

| MS ID* | Correlation | | | RMSD | | | PBIAS | | |
	ANN	SE	SAM	ANN	SE	SAM	ANN	SE	SAM
17290	**17.5****	**15.8**	**15.8**	**− 18.1**	**− 14.2**	**− 16.4**	248.9	418.7	278.4
17296	**19.2**	**17.3**	**17.3**	**− 12.0**	**− 9.1**	**− 11.2**	382.7	629.7	493.0
17320	**24.5**	**17.0**	**17.0**	**− 14.7**	**− 9.9**	**− 11.9**	31816.7	39750.0	28033.3
17340	**13.0**	**10.9**	**17.4**	**− 14.2**	**− 12.9**	**− 15.1**	200.3	213.5	194.6
17300	**30.2**	**27.9**	**27.9**	**− 11.8**	**− 8.9**	**− 11.1**	506.3	662.7	496.1
17330	**26.7**	**24.4**	**28.9**	**− 15.8**	**− 13.0**	**− 15.3**	1775.7	5491.9	4737.8
17292	**15.8**	**12.3**	**14.0**	**− 18.4**	**− 15.0**	**− 17.5**	695.8	932.7	770.6
17238	**21.6**	**16.2**	**21.6**	**− 13.5**	**− 14.4**	**− 16.2**	218.7	32.3	69.8
17240	**15.0**	**17.5**	**12.5**	**− 23.3**	**− 23.8**	**− 23.8**	3.8	**− 46.4**	**− 26.2**
17246	**31.0**	**19.0**	**21.4**	**− 19.8**	**− 17.4**	**− 17.4**	3777.5	2505.0	2760.0
17244	**37.1**	**34.3**	**31.4**	**− 20.5**	**− 19.3**	**− 18.1**	**− 44.5**	155.2	191.0
17239	**25.7**	**25.7**	**28.6**	**− 14.6**	**− 13.9**	**− 15.3**	753.2	534.8	724.7
17188	**26.2**	**16.7**	**26.2**	**− 28.2**	**− 25.5**	**− 28.2**	45.6	59.3	79.6
17190	**37.9**	**37.9**	**37.9**	**− 17.7**	**− 16.7**	**− 18.8**	306.9	148.6	259.0

*MSs are given in the order of proximity to the sea, **improved values are given in bold

and 4 of Table 7 for ANN, SE, and SAM, respectively. Up to 38% improvement is achieved due to ensembling.

Similar to correlations, the ETS obtained from ANN_AllMs generally resulted in the best RMSD values (see Table 5). For 17340, 17240, 17239, and 17190, either SE or SAM with all models resulted in better RMSD values, while for 17238, SE_SMs provided the minimum RMSD. The performances of both ANN models are very similar to each other. Improvements in RMSD when all models are used instead of selected three are less than 3% for all stations, other than 17290, 17292, and 17300. Similarly, for SE, improvements are less than 3% for all the stations, and in fact for 17320, 17238, and 17188, SE_SMs performed better than SE_AllMs. On the contrary, improvement in RMSD is more than 5% for 12 stations when SAM is used for ensembling. Thus, it can be

concluded that generally, all methods perform better in terms of RMSD when all models instead of the selected three models are used as inputs, but SAM benefits the most from ensembling a larger number of inputs.

For RMSD, the ETS generated using all models or selected three models performs better than the best RCMTS for all stations. Percent improvements in RMSD (i.e., decrease in RMSD) when the ETS is evaluated with respect to the best RCM are given in columns 5, 6, and 7 of Table 7 for ANN, SE, and SAM, respectively. The improvements in RMSD range between 9 and 28% when the ETS is used instead of the best RCMTS.

As shown from Table 6, the bias correction performance of ensembling is not as good as those of correlation and RMSD. All ensemble results are in the range of best and worst RCMTS performances for all stations. Thus, the ETS performs better than some of RCMTS but generally not as well as the best RCMTS. PBIAS values of the best RCMTS are better than all the ETS for all the stations other than 17238, 17240, 17244, 17188, and 17190. For ANN, utilization of all models compared to selected models improves PBIAS values for all the stations other than 17239 and 17188, but not to the level of the best RCMTS. For SE, PBIAS values improve when all models are used for half of the stations. In contrast, PBIAS values are better when selected models are used for all stations other than 17296 and 17300 for SAM. So, it can be concluded that PBIAS values of the best RCMTS cannot be achieved with either of the three ensembling methods with all or selected models. Thus, when the goal is to obtain precipitation time series with minimum bias, the best-performing RCMTS should be preferred over the ETS when ensembling is carried out using the procedures outlined in this study.

Percent deterioration in PBIAS when the ETS obtained by using the better of the selected or all models is evaluated with respect to the best RCMTS and is given in the last three columns of Table 7. For most of the stations, higher PBIAS values are obtained for the ETS compared to those obtained for the best RCMTS. This outcome does not support the literature where the multimodel ensemble is reported to reduce model biases [30, 32, 54]. In this study, we believe that the poor performance of ensembling for PBIAS is partially due to the utilization of log-transformed inputs. In order to test this hypothesis, the ETS is generated for MS 17320 which has the worst PBIAS performance

Fig. 7 Taylor diagrams of RCMTS and the ETS obtained from ANN, SE, and SAM for all MSs (legend and explanation of axis are provided on the bottom right diagram)

(see Table 7) using only normalized RCMTS as inputs (i.e., log-transformation is not applied to input data). AllMs are used to generate the ETS, and the results are given in Table 8. Log-transformation does not significantly affect correlation and RMSD performances of ensembling. However, PBIAS values decrease significantly, almost to the level of the best-performing RCMTS, when log-transformation is not used. Thus, when bias is the critical performance criterion, it is beneficial to use RCMTS directly, without log-transformation.

The superior results for simulation skills regarding correlation and RMSD indicate the advantage of the use of MME over single-model analysis for long-term climate assessments. Furthermore, the MME obtained by the use of the nonlinear ANN method has a relatively better performance compared to the linear ensembling methods regarding both MLR and simple ensemble averaging. Hence, the improved representation skills by nonlinear ANN over linear methods obtained as a result are in agreement with the findings from a study by Krasnopolsky and Lin [26] for the continental US. Krasnopolsky and Lin [26] denote that nonlinear approaches provide better representation for long-term simulations, and for precipitation fields with high gradients and sharp, localized features for which the linearity assumption is not applicable, and therefore, ANN generates better skills compared to the conventional linear methods.

To provide a visual comparison of RCMTS and ETS, Taylor diagrams for all MSs are given in Fig. 7. The legend is given at the bottom right corner of the figure. The outputs that are close to the reference/observed are the better-performing ones. As shown from Fig. 7, RCMTS (marked with gray circles) always has lower correlation and higher RMSD values compared to those of the ETS as demonstrated in Tables 4 and 5. However, improvement in correlation and RMSD is maintained with ensembling at the cost of losing variance (see normalized standard deviation performances in Fig. 7).

To represent the performance of ensembling in terms of reproducing the mean and variation of the OTS, the relative

	Best RCM	Worst RCM	Log-transformed			Without log-transformation		
			ANN	SE	SAM	ANN	SE	SAM
Corr	0.53	0.26	0.66	0.62	0.62	0.64	0.62	0.61
RMSD	3.12	3.86	2.66	2.81	2.75	2.65	2.70	2.72
PBIAS	− 0.06	47.21	19.15	23.91	16.88	− 0.33	− 0.56	− 0.29

Table 8 Effect of log-transformation on ensemble performance for Anamur MS (17320)
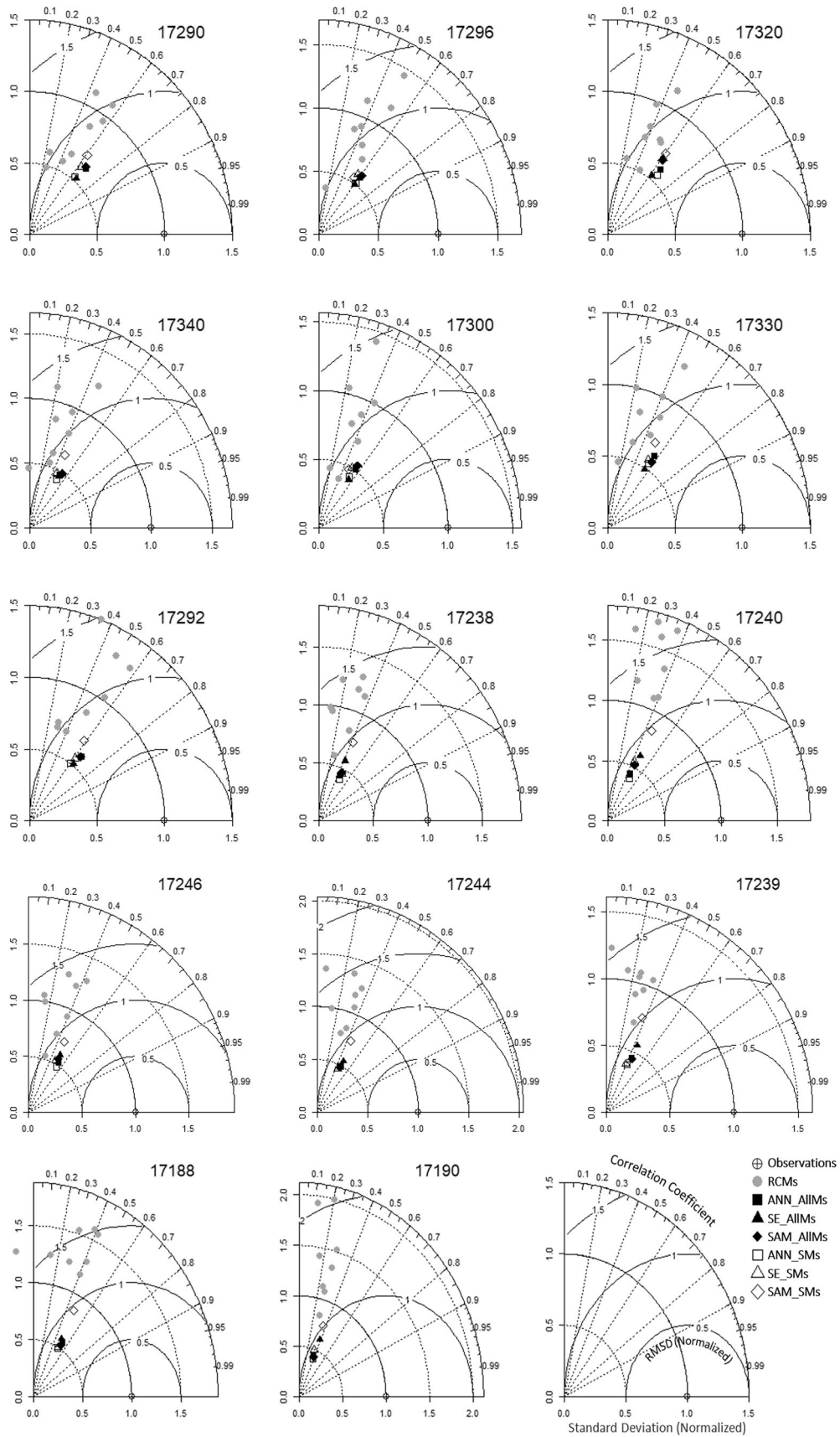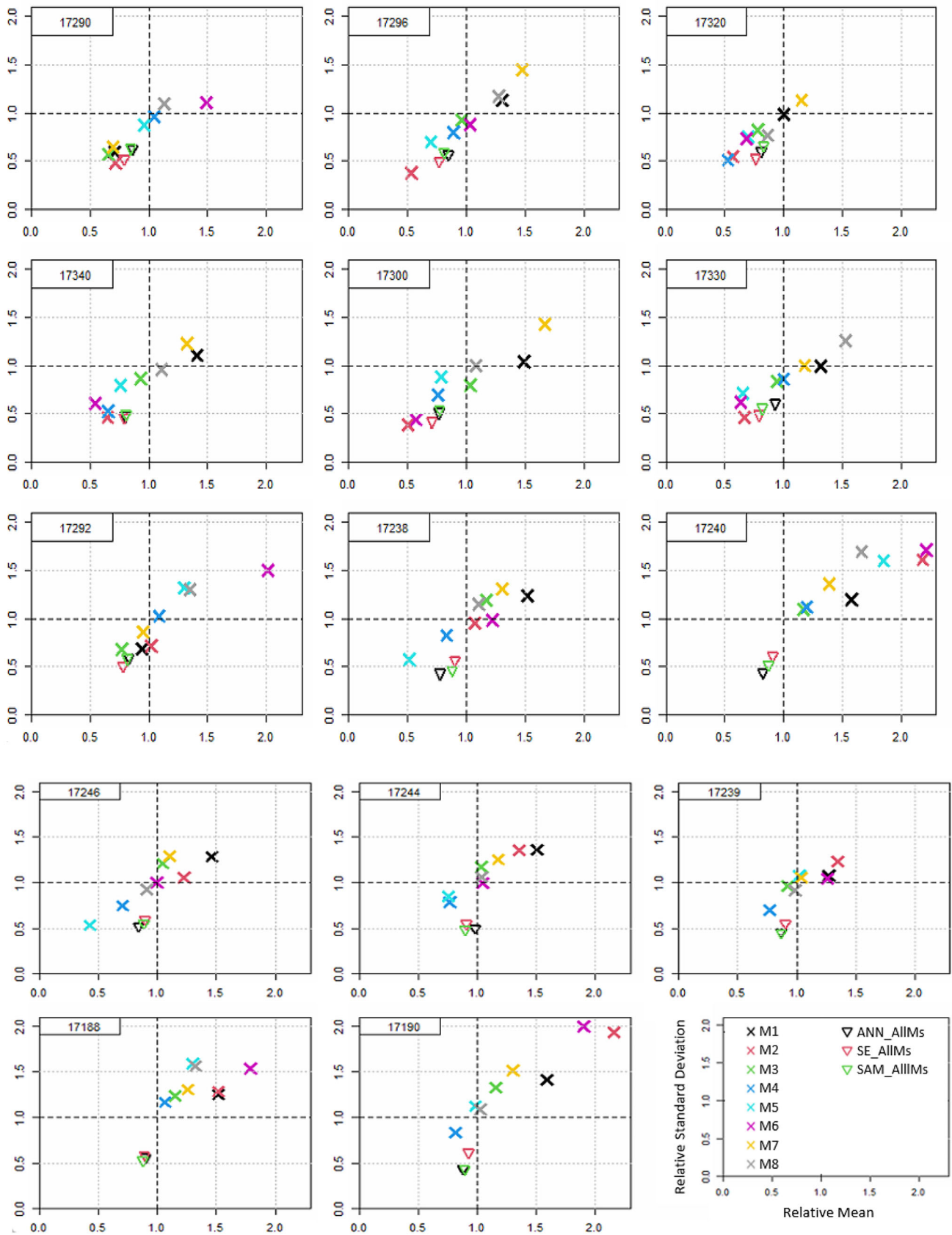
◄**Fig. 8** The relative mean versus relative standard deviation of RCMTS and the ETS obtained from ANN_AllMs, SE_AllMs, and SAM_AllMs for all MSs (legend and explanation of axis are provided on the bottom right diagram)

mean (i.e., the mean of RCMTS or ETS divided by the mean of the OTS) versus the relative standard deviation (i.e., the standard deviation of RCMTS or ETS divided by the standard deviation of the OTS) plots are given in Fig. 8. The relative mean values of the ETS range between 0.75 and 1.00, while relative standard deviations are around 0.5 for all MSs. On the other hand, the relative standard deviations of RCMTS are between 0.5 and 1.5, while the relative means of RCMTS are highly scattered (especially for MS 17240 and 17190). One noteworthy drawback of ensembling is that it causes time series to accumulate around the mean as shown in Fig. 8. Moreover, the standard deviation values of the ETS are significantly lower than those of the RCMTS in comparison with the OTS (see Figs. 7 and 8). Thus, it can be concluded that through ensembling, higher variations in RCMTS are mapped into ETS with lower variations.

The ETS obtained by ANN_AllMs for 17290 which has a very good performance (i.e., correlation = 0.67, RMSD = 1.85 mm, and PBIAS = 14.2) is given in Fig. 9. Eight RCMTS used in ensembling and the OTS are marked with gray and dashed black lines, respectively. As shown in Fig. 9, despite following the general trend, the ETS obtained by ANN_AllMs does not show the variation that exists in the OTS. For MS 17290, the relative standard deviations for four RCMTS (M1, M2, M3, and M7) have similar values to those of the ETS, but three RCMTS (M4, M5, and M8) have normalized standard deviations very close to the OTS (see Fig. 8). Although their correlation and RMSD values are lower than those of the ETS, these three RCMTS which have higher normalized standard deviations represent the variability in observed precipitation better. Hence, a set of individual climate models with different mechanisms provides a better representation of
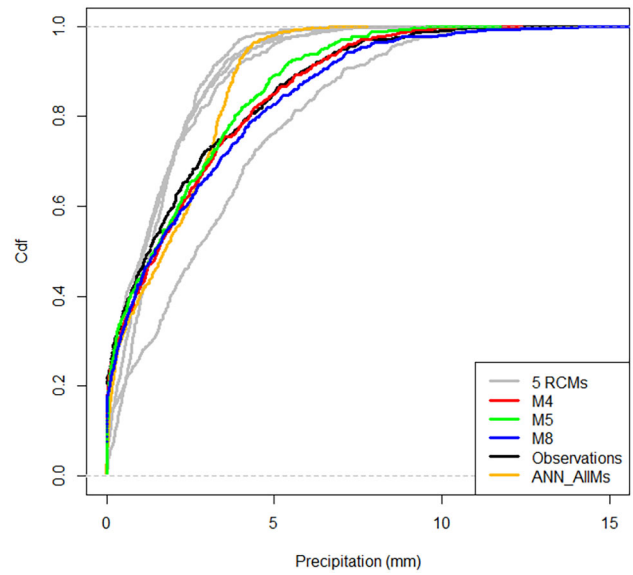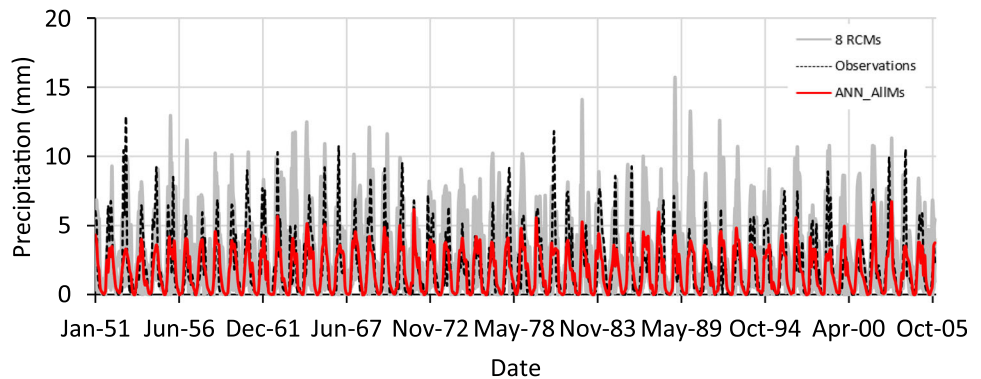

Fig. 10 Cumulative distribution functions for the ETS obtained from ANN_AllMs, observations, and eight input RCMs for Bodrum MS (17290)

the potential variability in observations. The cumulative distribution functions of the ETS obtained from ANN_AllMs, observations, and RCMTS demonstrate this behavior (Fig. 10). As shown in Fig. 10, the cumulative distribution functions of three RCMTS (M4, M5, and M8) are more similar to that of the OTS compared to the ANN_AllMs for MS 17290. However, as reported in Mascaro et al. [36], the performance skills of RCMs in capturing the variation of climatological precipitation patterns at small temporal scales are limited. Hence, the use of a single model for the assessment may cause higher biases in evaluations due to unconsidered uncertainties, especially for regions having complex spatial characteristics like the Mediterranean region. Given that, ensembled time series can be considered to be particularly useful in assessment when evaluated in integration with the likely ranges including relatively low frequency and/or extreme events
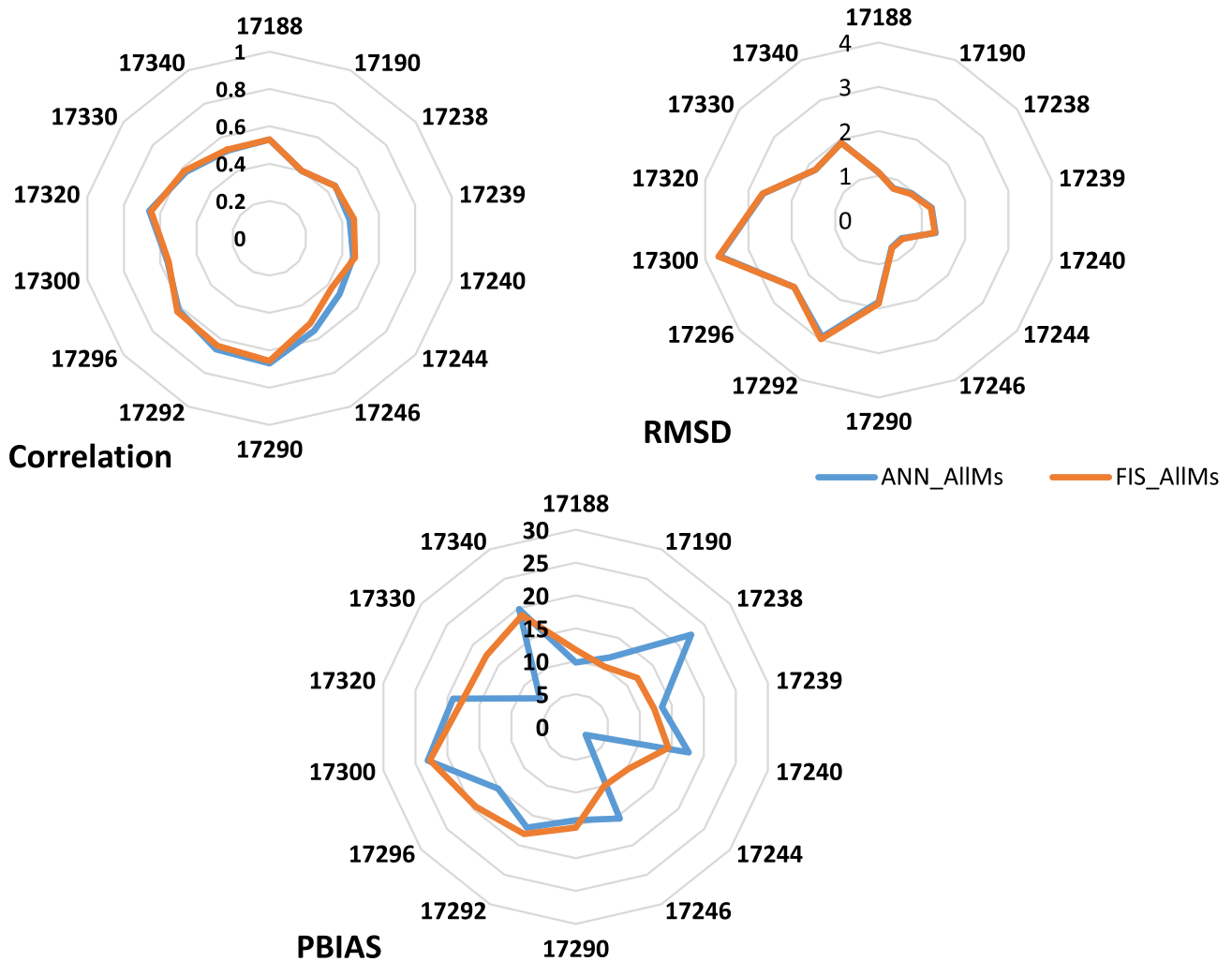
Fig. 9 ETS obtained from ANN_AllMs, observations, and eight input RCMs for Bodrum MS (17290)

Fig. 11 Comparison of the performances of ANN_AllMs and FIS_AllMs in terms of correlation, RMSD, and PBIAS

identified through a large ensemble set of individual models with diverse approaches or assumptions.

Providing more consistent estimates, ensembling is advisable for station simulation [58]. Ensembling, on the other hand, improves correlation and RMSD performances at the cost of reducing the variability in the precipitation. Thus, it can be concluded that in the estimation of the total depth/volume of precipitation over a long period of time, utilization of ensembled time series will be beneficial. However, when the goal is to select extreme precipitations to be used in design discharge calculations of water structures, peak values of individual RCMs will be more conservative, thus, may be better preferred.

It is shown that the nonlinear ANN approach resulted in improved performances in terms of correlations and RMSDs (Tables 4 and 5) for ensembling for most of MSs. As a final analysis, FIS which is another data-driven method is used for ensembling to investigate the added benefit of nonlinear approaches. FIS models using all

RCMs (FIS_AllMs) as inputs are built for each MS, and the performances of two nonlinear approaches are compared in terms of correlation, RMSD, and PBIAS. In this study, Takagi–Sugeno type FISs are constructed, and fuzzy rules are identified via subtractive clustering. For each MS, 400 different FIS models with various cluster radii and numbers of cluster centers are constructed by using training data, and the best-performing model is selected according to the performance of FIS for validation data (e.g., for fivefold validation, 80–20 training-validation ratio is used). Once the best-performing model is selected for the first fold, the predictions of other folds are obtained by using the parameters of the selected model to obtain ETS. The reader may refer to Mesta et al. [38] for details about FIS modeling. The comparison of FIS with ANN is presented in Fig. 11, and the information about the parameters of FIS is given in Appendix A. Figure 11 shows that the performances of ANN_AllMs and FIS_AllMs are almost the same for all MS in terms of correlations and RMSDs.

However, ANN_AllMs performs slightly better for stations 17244 and 17246 in terms of correlation. On the other hand, although the PBIAS values obtained from ANN_AllMs and FIS_AllMs for MS significantly differ from each other, both models underestimate the precipitation for all MS according to the PBIAS measure. To summarize, both nonlinear approaches, ANN and FIS, provided very similar performances in terms of correlations and RMSDs which are better than those of linear models for almost all the MSs.

5 Summary and main findings of the study

The objective of this study is to examine the effect of the multimodel ensembling using linear and nonlinear approaches on the precipitation simulations for a study region in south and southwestern Turkey with complex climatic features. In the analysis, three ensembling approaches, namely, SAM, SE, and ANN, are applied by the use of the historical simulation outputs of eight RCMs from the CORDEX database. Long-term monthly precipitation time series obtained from ground-based precipitation observations from fourteen meteorological stations is used as the benchmark. This study provides a comparison of the improvement in replication skills of historical precipitation simulations from individual RCMs by the use of linear and nonlinear MME methods. The main findings of the study are as follows:

- The analysis results show that the overall performance of the ensembled time series is better compared to individual RCMs. Generally, stations in the coastal strip have better-performing RCMs, thus ensembling works better for these stations compared to the inland ones.
- ANN generally provided the best performance among the three ensembling methodologies, particularly regarding the correlation and RMSD values. Additionally, it is seen to perform better when all RCMTS are used as inputs instead of the three best-performing RCMTS.
- Analysis of the individual RCMTS in a multimodel analysis generates a wide range of correlations and RMSD, hence, high uncertainty, however, the use of multimodel ensembling reduces the model uncertainty. On the other hand, PBIAS values for MME are seen to be higher than the value for the best individual model. It is observed that bias reduction, almost to the level of the best RCMTS performance, is achieved when input time series is directly used (i.e., without log-transformation) in ensembling.

- Despite the advantages of reducing uncertainty and improvement in correlation and RMSD values, ensembling is verified to increase the simulation performance at the cost of reducing the variability in the precipitation. It should be realized that extreme events are poorly represented in the ensembled time series, and this may result in the inefficient design of various water structures such as spillways and storm water drainage systems that are based on high return period events. Thus, it can be concluded that ensembling is more useful in the estimation of the total depth/volume of precipitation over long time periods, rather than for the simulation of extreme precipitations.

6 Conclusions

The main conclusions of the study are as follows:

- Ensembled monthly mean precipitation time series of RCM outputs has higher correlations and lower RMSDs with ground-based observations compared to single RCM outputs, and thus should be preferred in climate impact analysis.
- Nonlinear models such as ANN or FIS that as well have linear modeling capabilities should be preferred for ensembling instead of linear models such as SE and SAM.
- When the goal is to assess the impacts of extreme precipitation, all available single RCM outputs should be used instead of ensembled time series as this will lead to a better understanding of possible variability.

The use of nonlinear methods such as ANN and FIS for ensembling will lead to better quantification of climate change impacts and in return better design of new hydraulic structures and development of suitable adaptation measures. Of course, the above-stated conclusions are derived based on the study conducted in south and southwestern Turkey. Confirmation of these results in other parts of the world is a matter for future work.

Appendix 1: The parameters of fuzzy inference system (FIS)

See Table 9.

Table 9 The number of cluster center (cc) and cluster radius (r_a) used in the FIS model for each meteorological station (MS)

MS	cc	r_a*
17188	2	0.25
17190	2	0.45
17238	1	0.2
17239	2	0.75
17240	1	0.25
17244	4	0.65
17246	1	0.2
17290	1	0.25
17292	2	0.4
17296	2	0.5
17300	1	0.25
17320	2	0.65
17330	3	0.65
17340	2	0.3

*In all FIS models, r_b is taken as $1.5 r_a$

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest Other than the funding information provided under Acknowledgements title, the authors have no other competing interests to declare that are relevant to the content of this article.

References

1. Ajami NK, Duan Q, Gao X, Sorooshian S (2006) Multimodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results. J Hydrometeorol 7(4):755–768

2. Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22(1):42–47

3. Baghanam AH, Eslahi M, Sheikhbabaei A, Seifi AJ (2020) Assessing the impact of climate change over the northwest of Iran: an overview of statistical downscaling methods. Theoret Appl Climatol 141(3):1135–1150

4. Baker NC, Taylor PC (2016) A framework for evaluating climate model performance metrics. J Clim 29(5):1773–1782

5. Black E, Brayshaw DJ, Rambeau CM (2010) Past, present and future precipitation in the Middle East: insights from models and observations. Philosoph Transact Royal Soc A: Math, Phys Eng Sci 368(1931):5173–5184

6. Boulanger JP, Martinez F, Segura EC (2006) Projection of future climate change conditions using IPCC simulations, neural networks and Bayesian statistics. Part 1: temperature mean state and seasonal cycle in South America. Clim Dyn 27(2–3):233–259

7. Boulanger JP, Martinez F, Segura EC (2007) Projection of future climate change conditions using IPCC simulations, neural networks and Bayesian statistics. Part 2: precipitation mean state and seasonal cycle in South America. Clim Dyn 28(2–3):255–271

8. Bozkurt D, Sen OL (2011) Precipitation in the Anatolian Peninsula: sensitivity to increased SSTs in the surrounding seas. Clim Dyn 36(3–4):711–726

9. Cakir S, Kadioglu M, Cubukcu N (2013) Multischeme ensemble forecasting of surface temperature using neural network over Turkey. Theoret Appl Climatol 111(3–4):703–711

10. Campozano L, Tenelanda D, Sanchez E, Samaniego E, Feyen J (2016) Comparison of statistical downscaling methods for monthly total precipitation: case study for the paute river basin in Southern Ecuador. Adv Meteorol. https://doi.org/10.1155/2016/6526341

11. Cane D, Milelli M (2010) Multimodel superensemble technique for quantitative precipitation forecasts in Piemonte region. Nat Hazards Earth Syst Sci 10(2):265–273

12. Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. Climate Res 44(2–3):179–194

13. Evans JP (2009) 21st Century climate change in the Middle East. Clim Change 92(3–4):417–432

14. Fan Y, Krasnopolsky VM, van den Dool H, Ch-Y W, Gottschalck J (2021) Using artificial neural networks to improve CFS week 3–4 precipitation and 2-Meter air temperature forecasts. Weather Forecast. https://doi.org/10.1175/WAF-D-20-0014.1

15. Fujihara Y, Tanaka K, Watanabe T, Nagano T, Kojiri T (2008) Assessing the impacts of climate change on the water resources of the Seyhan River Basin in Turkey: use of dynamically downscaled data for hydrologic simulations. J Hydrol 353(1–2):33–48

16. Galmarini S, Bianconi R, Klug W, Mikkelsen T, Addis R, Andronopoulos S, Bellasio R (2004) Ensemble dispersion forecasting—Part I: concept approach and indicators. Atmosph Environ 38(28):4607–4617

17. Gao X, Giorgi F (2008) Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model. Global Planet Change 62(3–4):195–209

18. Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM

simulations via the reliability ensemble averaging (REA) method. J Clim 15(10):1141–1158

19. Giorgi F, Mearns LO (2003) Probability of regional climate change based on the reliability ensemble averaging (REA) method. Geophys Res Lett. https://doi.org/10.1029/2003GL017130

20. Gutiérrez JM, Maraun D, Widmann M, Huth R, Hertig E, Benestad R, San Martin D (2019) An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. Int J Climatol 39(9):3750–3785

21. Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—I Basic concept. Tellus A: Dyn Meteorol Oceanogr 57(3):219–233

22. Hemming D, Buontempo C, Burke E, Collins M, Kaye N (2010) How uncertain are climate model projections of water availability indicators across the Middle East? Philosoph Transact Royal Soc A: Math, Phys Eng Sci 368(1931):5117–5135

23. Kang B, Do Kim Y, Lee JM, Kim SJ (2015) Hydro-environmental runoff projection under GCM scenario downscaled by Artificial Neural Network in the Namgang Dam watershed. Korea KSCE J Civil Eng 19(2):434–445

24. Kara F, Yucel I (2015) Climate change effects on extreme flows of water supply area in Istanbul: utility of regional climate models and downscaling method. Environ Monit Assess 187(9):580

25. Kentel, E., Akgun, O. B., & Mesta, B. (2019, December). User-Friendly R-Code for Data Extraction from CMIP6 outputs. In: AGU Fall Meeting Abstracts (Vol. 2019, pp. PA33C-1098).

26. Krasnopolsky VM, Lin Y (2012) A neural network nonlinear multimodel ensemble to improve precipitation forecasts over continental US. Adv Meteorol. https://doi.org/10.1155/2012/649450

27. Kitoh, A. (2007). Future climate projections around Turkey by global climate models. Final report of ICCAP.

28. Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010) Challenges in combining projections from multiple climate models. J Clim 23(10):2739–2758

29. Kotlarski S, Keuler K, Christensen OB, Colette A, Déqué M, Gobiet A, Nikulin G (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7:1297–1333

30. Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Surendran S (1999) Improved weather and seasonal climate forecasts from multimodel superensemble. Science 285(5433):1548–1550

31. Krishnamurti TN, Kishtawal CM, Zhang Z, LaRow T, Bachiochi D, Williford E, Surendran S (2000) Multimodel ensemble forecasts for weather and seasonal climate. J Clim 13(23):4196–4216

32. Krishnamurti TN, Kumar V, Simon A, Bhardwaj A, Ghosh T, Ross R (2016) A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. Rev Geophys 54(2):336–377

33. López-Moreno JI, Goyette S, Beniston M (2008) Climate change prediction over complex areas: spatial variability of uncertainties and predictions over the Pyrenees from a set of regional climate models. Int J Climatol: A J Royal Meteorol Soc 28(11):1535–1550

34. Lun Y, Liu L, Cheng L, Li X, Li H, Xu Z (2021) Assessment of GCMs simulation performance for precipitation and temperature from CMIP5 to CMIP6 over the Tibetan Plateau. Int J Climatol. https://doi.org/10.1002/joc.7055

35. Maharana P, Kumar D, Das S, Tiwari PR (2021) Present and future changes in precipitation characteristics during Indian summer monsoon in CORDEX-CORE simulations. Int J Climatol 41(3):2137–2153

36. Mascaro G, Viola F, Deidda R (2018) Evaluation of precipitation from EURO-CORDEX regional climate simulations in a small-scale mediterranean site. J Geophys Res: Atmosph 123(3):1604–1625

37. Mendes D, Marengo JA (2010) Temporal downscaling: a comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios. Theoret Appl Climatol 100(3):413–421

38. Mesta B, Akgun OB, Kentel E (2021) Alternative solutions for long missing streamflow data for sustainable water resources management. Int J Water Resour Dev 37(5):882–905

39. Mesta B, Kentel E (2022) Superensembles of raw and bias-adjusted regional climate models for Mediterranean region. Turkey Int J Climatol 42(4):2566–2585

40. Okkan U (2015) Assessing the effects of climate change on monthly precipitation: proposing of a downscaling strategy through a case study in Turkey. KSCE J Civ Eng 19(4):1150–1156

41. Okkan U, Inan G (2015) Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: use of machine learning methods, multiple GCMs and emission scenarios. Int J Climatol 35(11):3274–3295

42. Okkan U, Kirdemir U (2016) Downscaling of monthly precipitation using CMIP5 climate models operated under RCPs. Meteorol Appl 23(3):514–528

43. Onol B, Unal YS (2014) Assessment of climate change simulations over climate zones of Turkey. Reg Environ Change 14(5):1921–1935

44. Ozdogan M (2011) Modeling the impacts of climate change on wheat yields in Northwestern Turkey. Agr Ecosyst Environ 141(1–2):1–12

45. Palatella L, Miglietta MM, Paradisi P, Lionello P (2010) Climate change assessment for Mediterranean agricultural areas by statistical downscaling. Nat Hazard 10(7):1647

46. Park C, Lee G, Kim G, Cha DH (2021) Future changes in precipitation for identified sub-regions in East Asia using bias-corrected multi-RCMs. Int J Climatol 41(3):1889–1904

47. Roebber PJ, Butt MR, Reinke SJ, Grafenauer TJ (2007) Real-time forecasting of snowfall using a neural network. Weather Forecast 22(3):676–684

48. Sensoy, S. (2004). The mountains influence on Turkey climate. In: *Balwois Conference on Water Observation and Information System for Decision Support.*

49. Sirdas S, Ross RS, Krishnamurti TN, Chakraborty A (2007) Evaluation of the FSU synthetic superensemble performance for seasonal forecasts over the Euro-Mediterranean region. Tellus A: Dyn Meteorol Oceanogr 59(1):50–70

50. Stefanova L, Krishnamurti TN (2002) Interpretation of seasonal climate forecast using Brier skill score, the Florida State University superensemble, and the AMIP-I dataset. J Clim 15(5):537–544

51. Stocker, T. F., Dahe, Q., Plattner, G. K., & Tignor, M. (2015, September). IPCC Workshop on Regional Climate Projections and their Use in Impacts and Risk Analysis Studies. In: Workshop Report (Vol. 15, p. 18).

52. Sunyer Pinya MA, Hundecha Y, Lawrence D, Madsen H, Willems P, Martinkova M, Vormoor K, Bürger G, Hanel M, Kriaučiuniene J, Loukas A, Osuch M, Yücel I (2015) Intercomparison of statistical downscaling methods for projection of extreme precipitation in Europe. Hydrol Earth Syst Sci 19(4):1827–1847. https://doi.org/10.5194/hess-19-1827-2015

53. Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmosph 106(D7):7183–7192

54. Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philosoph Transact Royal Soc A: Math, Phys Eng Sci 365(1857):2053–2075

55. Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J Hydrol 456:12–29

56. Vautard R, Kadygrov N, Iles C, Boberg F, Buonomo E, Bülow K, Wulfmeyer V (2021) Evaluation of the large EURO-CORDEX regional climate model ensemble. J Geophys Res: Atmosph 126(17):e2019JD032344

57. Vu MT, Aribarg T, Supratid S, Raghavan SV, Liong SY (2016) Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? Theoret Appl Climatol 126(3–4):453–467

58. Xu R, Chen N, Chen Y, Chen Z (2020) Downscaling and projection of multi-CMIP5 Precipitation using machine learning methods in the upper Han River Basin. Adv Meteorol 2020:1–7

59. Yun WT, Stefanova L, Mitra AK, Vijaya Kumar TSV, Dewar W, Krishnamurti TN (2005) A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. Tellus A: Dyn Meteorol Oceanogr 57(3):280–289

60. Zhang GP, Berardi VL (2001) Time series forecasting with neural network ensembles: an application for exchange rate prediction. J Operation Res Soc 52(6):652–664. https://doi.org/10.1057/palgrave.jors.2601133

61. ESGF, Earth System Grid Federation website, https://esgf-node.llnl.gov/search/esgf-llnl/, (CoG version v4.0.0b2, ESGF P2P Version v4.0.4), [Accessed 5th June 2020]