ORIGINAL ARTICLE

Multi-Model Ensemble Machine Learning Approaches to Project Climatic Scenarios in a River Basin in the Pyrenees

Nerea Bilbao-Barrenetxea[1,2] · Raquel Martínez-España[3] · Patricia Jimeno-Sáez[4] · Sergio Henrique Faria[2,5] · Javier Senent-Aparicio[4]

Abstract

This study employs machine learning algorithms to construct Multi Model Ensembles (MMEs) based on Regional Climate Models (RCMs) within the Esca River basin in the Pyrenees. RCMs are ranked comprehensively based on their performance in simulating precipitation (pr), minimum temperature (tmin), and maximum temperature (tmax), revealing variability across seasons and influenced by the General Circulation Model (GCM) driving each RCM. The top-ranked approach is used to determine the optimal number of RCMs for MME construction, resulting in the selection of seven RCMs. Analysis of MME results demonstrates significant improvements in precipitation on both annual and seasonal scales, while temperature-related enhancements are more subtle at the seasonal level. The effectiveness of the ML–MME technique is highlighted by its impact on hydrological representation using a Temez model, yielding outcomes comparable to climate observations and surpassing results from Simple Ensemble Means (SEMs). The methodology is extended to climate projections under the RCP8.5 scenario, generating more realistic information for precipitation, temperature, and streamflow compared to SEM, thus reducing uncertainty and aiding informed decision-making in hydrological modeling at the basin scale. This study underscores the potential of ML–MME techniques in advancing climate projection accuracy and enhancing the reliability of data for basin-scale impact analyses.

Keywords Multi-model ensemble · Machine learning algorithms · Complex orography · Pyrenees · Regional climate models

1 Introduction

Accurately characterising future climate is of crucial importance for medium and long-term water resource planning and management within the context of climate change (IPCC 2022). While General Circulation Models (GCMs) have

✉ Nerea Bilbao-Barrenetxea
  nbilbao034@ikasle.ehu.eus

1 Faculty of Science and Technology, University of the Basque Country UPV/EHU, 48940 Leioa, Spain

2 Basque Centre for Climate Change (BC3), 48940 Leioa, Spain

3 Information and Communications Engineering Department, University of Murcia (UM), 30100 Murcia, Spain

4 Department of Civil Engineering, Catholic University of San Antonio, 30107 Murcia, Spain

5 Basque Foundation for Science, IKERBASQUE, 48009 Bilbao, Spain

emerged as powerful tools for climate prediction (Semenov and Stratonovitch 2010), they still exhibit certain limitations when it comes to representing regional climates affected by small-scale processes (Torma et al. 2015). Addressing this need, Regional Climate Models (RCMs) have been developed, based on dynamic downscaling of GCM models to provide high-resolution data. Notably, the CORDEX (Giorgi et al. 2009) initiative has successfully brought together RCM projects from around the world, boasting more than 70 simulations for the European region (Jacob et al. 2014, http://www.euro-cordex.net/).

Despite the clear advantages of RCMs over GCMs in capturing the primary features of regional climate (Kotlarski et al. 2014; Ciarlo et al. 2021), inherent uncertainties persist, extending beyond the scope of downscaling. These uncertainties encompass structural disparities in both GCM and RCM models (Knutti et al. 2008), the downscaling technique itself (Zhu et al. 2019), model parametrizations in reference to physical processes (Chen et al.

2011), and initial conditions, among other factors (Knutti et al. 2008; Dey et al. 2022). Furthermore, in studies conducted at the catchment scale, such as those examining the impacts of climate change on water resources, a scale mismatch remains, at times leading to unresolved climatic dynamics beyond the capabilities of RCM resolutions (Crawford et al. 2019). Consequently, these uncertainties can result in significant discrepancies in climate change projections between different RCMs, even when considering identical emission scenarios (Ruane and McDermid 2017). This, coupled with the scale mismatch that introduces limitations in climate representation, hampers the effective utilisation of this data for catchment-scale planning and water resource management (Venkataraman et al. 2016).

Impact modellers employ a wide array of methods to tackle these uncertainties and errors, encompassing a broad spectrum of complexities. These methods span from identifying the best-performing simulations within the study area (Crawford et al. 2019; Xu et al. 2020) to the utilisation of bias correction techniques with observational data (Dobor and Hlásny 2019; Teng et al. 2015; Piani et al. 2010), and extend to the development of Multi-Model Ensembles (MMEs)(Calì Quaglia et al. 2022; Salman et al. 2018). Bias correction methods have been instrumental in rectifying the systematic biases inherent in simulations (Piani et al. 2010). Nevertheless, they often prove less efficient in addressing non-stationary biases (White and Toumi 2013; Wang et al. 2018). A promising avenue for addressing the uncertainty of climate models lies in the development of MMEs, which have the potential to mitigate uncertainties and enhance the confidence in climate projections (Pavan and Doblas-Reyes 2000; Lutz et al. 2016; Sanderson et al. 2015; Keller et al. 2019). MMEs are categorised into two distinct groups: SEM (Simple Ensemble Mean) and WEM (Weighted Ensemble Mean). In the former approach, all ensemble members are uniformly assigned equal weights, whereas in the Weighted Ensemble Method (WEM), each member is allocated a distinct weight determined by its proficiency in replicating past climate conditions (Oh and Suh 2017; Ahmed et al. 2020). SEM, known for its simplicity, is a commonly employed method, which provides an overall better performance than individual members (Lambert and Boer 2001). However, it comes with certain limitations. Many of the models share model parameterizations and components, which can lead to interdependencies between different climate simulations (Sanderson et al. 2015). Failing to account for this interdependence may result in misleading model consensus, reduced accuracy, and a flawed estimation of uncertainty (Herger et al. 2018). Moreover, SEM may not be suitable for all applications, as it significantly diminishes the spatial and temporal variability of information when compared to individual members and observational data (Wang et al. 2018).

In contrast, WEM methods have demonstrated their capacity to mitigate the impact of systematic biases within individual members and even enhance the ensemble's predictive capabilities (Krishnamurti et al. 1999, 2000). The use of Machine Learning algorithms to generate a Multi-Model Ensemble (ML–MME) is an emerging technique in climate simulation (Zhu et al. 2023; Sand et al. 2023). These algorithms have a significant potential to enhance the outcomes of climate simulations, especially in relation to its potential advantages in dealing with non-linearity between response variables and predictors (Ahmed et al. 2020; Sachindra et al. 2018; Xu et al. 2020). Krishnamurti et al. (1999) established a precedent of an MME based on multiple regression techniques to improve the 850 hPa meridional wind speed and precipitation simulations of eight general circulation models, obtaining superior results over the ensemble mean. Wang et al. (2018) employed four Machine Learning (ML) techniques to develop MMEs for mean monthly temperature and mean monthly precipitation by considering 33 CMIP5 GCMs over Australia and reported that Random Forest (RF) and Support vector machine (SVM) demonstrated a significant improvement over the ensemble mean, which is in agreement with the results reported by Sa'adi et al. (2017) who employed a Generalised Linear Model (GLM) to construct their MMEs obtaining better results for the MMEs than for the 20 individual members of the CMIP5 GCMs over Borneo Island, Malaysia. Results along these lines have been reported in studies in Iraq for monthly mean temperature (Salman et al. 2018), in Pakistan for monthly precipitation (Ahmed et al. 2020), or in the Gulf Basin and North America for both (Crawford et al. 2019). Daily scale studies also show favourable results for ML–MME techniques (Jose et al. 2022). Likewise, Dey et al. (2022) obtained significant improvements in the characterisation of these climate variables with data from CMIP6 GCMs.

In our study, we ventured into a novel approach by applying various ML–MME methods to RCMs for the first time. These methods were then further applied to a hydrological model. We subjected them to a comparative analysis against the SEM (Simple Ensemble Mean) approach, focusing on monthly precipitation (pr), the monthly average of daily maximum temperature (tmax), and the monthly average of daily minimum temperature (tmin). Specifically, the ML–MME techniques encompassed Linear Regression (LR), Gradient Boosting (GB), and Random Forest (RF). This investigation is particularly noteworthy as we apply it to a complex topography region, which adds a layer of novelty to our research given the challenges it presents for simulation (Torma et al. 2015; Reder et al. 2020). First, a ranking of the RCMs has been developed based on their skill to characterize the past climate and the optimal number of RCMs to be included in the ML-MMEs has been determined. Once the final ML–MMEs for the three variables

have been defined, the monthly series were analysed in detail by comparing them with the climate observations. To illustrate the practical utility of the ML–MMEs in the application of impact studies at the watershed scale, we employed them as input data for the Temez hydrological model, both for historical periods and future climate projections within the study area.

2 Data and Study Area

We considered the EURO-CORDEX ensemble (Jacob et al. 2014, 2020), with a total of 72 RCM simulations (Table 1) with a spatial resolution of 0.11° × 0.11°. These simulations cover a time period of 130 years, from 1980 to 2100 for RCP8.5 and is based on the combination of two models, the RCM and the driver model, the GCM, forming an incomplete matrix of 12 RCM and 8 GCM models.

CLIMPY observational dataset (Cuadrat et al. 2020) is used as reference, with a spatial resolution of 1 km × 1 km on a daily basis covering the period 1980–2015. It is a reconstruction (Serrano-Notivoli et al. 2017) of the variables based on the information from 1,343 meteorological stations located in Spain, France, and Andorra. This dataset was created in the framework of the transboundary project CLIMPY and has already been validated in different studies (Amblar-Francés et al. 2020; Lemus-Canovas et al. 2019). For the proper comparison between the data from simulations and observations, both must be on the same grid. Thus, an bilinear interpolation to a the rectilinear grid o the RCMs of 0.11° × 0.11° resolution of CLIMPY has been performed.

The Esca River basin is located in the western Pyrenees, northeastern Spain, and covers an area of 425 km^2, which corresponds to four grid-cells of the climate datasets. Characterised by a large altitudinal gradient, the elevation of the highest point of the basin is 2,100 m, while its lowest point is 595 m above sea level. Orographic characteristics make this type of basin remarkably difficult to simulate its climate dynamics (Kotlarski et al. 2014; Smiatek et al. 2016) Therefore, they are particularly problematic areas for accurately predicting the future climate and its related impacts in hydrology (Fatichi et al. 2016). It is important to make efforts to overcome these difficulties, particularly in cases such as the Esca river basin, since it is a key tributary feeding the Yesa reservoir, the primary reservoir in the western Pyrenees. Data on streamflows of the Esca river were available from the website of the Spanish Centre for Hydrographic Studies (CEDEX) (https://ceh.cedex.es/anuarioaforos/default.asp), where data are updated to September 2017.

The selection of the variables pr, tmax and tmin is motivated by two primary considerations. Firstly, their availability within the CLIMPY database (Cuadrat et al. 2020). Secondly, these variables are pivotal for characterizing the climate system, as emphasized in prior studies (Meehl et al. 2000; Perkins et al. 2007; Careto et al. 2022b, a), and play a crucial role in influencing diverse hydrological (Piani et al. 2010), biological, and industrial systems (Colombo et al. 1999; Coppola et al. 2021).

3 Methodology

This study follows a specific methodology, which progresses through several phases:(1) a ranking of the RCMs was developed according to the performance of the three analyzed variables—tmax, tmin, and pr—on a seasonal scale (Sect. 3.1), (2) the SEM and the ML–MMEs were constructed (Sect. 3.2) and (3) the optimal number of RCMs to form the MMEs was chosen (Sect. 3.3). (4) The definitive MMEs were evaluated (Sect. 3.4). Then, (5) to assess the impact of climate variables MMEs on flow characterization, we utilized these MMEs as input data for the Temez hydrological model (Sect. 3.5). Finally, (4) as an illustrative example of application of ML–MME results for climate change impact assessment, the definitive ML–MME algorithms were applied to the climate projections of the RCP8.5 emissions scenario.

The methodology proposed in the described 1, 2, 3 and 4 steps follows an outline of the data analysis processes (Berthold et al. 2010) presented in Fig. 1. The methodology initiates with a feature selection process aimed at eliminating noise-inducing features (RCMs) from the dataset, thus ensuring the development of a stable and reliable prediction model. This involves conducting an RCM ranking followed by the application of a filter-wrapper technique to identify the most suitable features. Upon selecting the optimal RCMs, various ML models are generated by optimising their hyperparameters using cross-validation. Subsequently, MMEs of tmax, tmin and pr are generated using the developed ML algorithms. These MMEs were subjected to an statistical performance evaluation.

Fig. 1 Outline of the stages of the data analysis process followed in the work
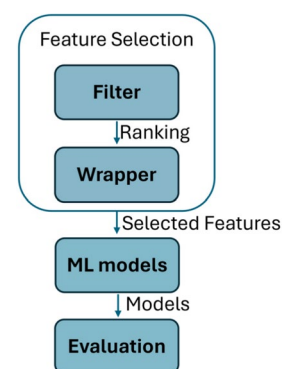
Table 1 Individual ranks of RCMs for each variable and season (DJF, MAM, JJA and SON) based on TSS and overall ranks of RCMs based on RM values according to their ability to simulate CLIMPY monthly precipitation (pr), monthly average of daily maximum (tmax) and minimum temperature (tmin) over the study area over the period 1980–2015

	RCM Names	pr				tmax				tmin				RM	Rank
		DJF	SON	JJA	MAM	DJF	SON	JJA	MAM	DJF	SON	JJA	MAM		
56	MPI-M-MPI-ESM-LR_SMHI-RCA4	42	41	2	42	3	1	26	11	9	14	51	3	0.7124	1
34	MIROC-MIROC5_CCLM4-8-17	7	11	33	49	6	46	2	16	48	13	25	4	0.6948	2
58	MPI-M-MPI-ESM-LR_ETH-COSMO-crCLIM-v1-1	2	12	5	12	35	43	50	58	2	4	47	13	0.6678	3
49	MPI-M-MPI-ESM-LR_CNRM-ALADIN63	33	20	18	16	20	3	33	30	11	21	55	31	0.6585	4
55	MPI-M-MPI-ESM-LR_MPI-CSC-REMO2009	45	27	19	41	13	13	36	37	3	7	52	5	0.6502	5
10	CNRM-CERFACS_SMHI-RCA4	32	53	21	39	38	2	1	27	55	15	12	17	0.6338	6
24	ICHEC-EC-EARTH_SMHI-RCA4	29	10	32	8	27	32	6	47	24	32	21	46	0.6315	7
52	MPI-M-MPI-ESM-LR_IPSL-WRF381P	38	23	29	24	17	16	60	24	6	6	58	14	0.6303	8
54	MPI-M-MPI-ESM-LR_MOHC-HadREM3-GA7-05	31	34	4	51	18	5	39	5	8	39	62	20	0.6291	9
51	MPI-M-MPI-ESM-LR_ICTP-RegCM4-6	46	15	53	58	25	9	42	2	13	11	45	6	0.6185	10
19	ICHEC-EC-EARTH_SMHI-RCA4	37	2	35	61	36	12	38	39	30	12	8	19	0.6138	11
53	MPI-M-MPI-ESM-LR_KNMI-RACMO22E	41	31	8	34	33	11	61	54	5	17	37	2	0.6080	12
1	CCCma-CanESM2_CCLM4-8-17	5	65	56	7	1	42	44	10	1	30	49	27	0.6045	13
57	MPI-M-MPI-ESM-LR_UHOH-WRF361H	50	46	20	38	12	14	52	20	7	33	48	9	0.5904	14
9	CNRM-CERFACS_MOHC-HadREM3-GA7-05	25	54	50	56	57	4	16	25	49	2	3	8	0.5904	15
29	IPSL-IPSL-CM5A-MR_DMI-HIRHAM5	69	66	15	2	42	27	37	4	31	46	2	10	0.5880	16
59	MPI-M-MPI-ESM-LR_MPI-CSC-REMO2009	3	8	28	44	30	35	58	62	4	23	61	7	0.5739	17
32	IPSL-IPSL-CM5A-MR_KNMI-RACMO22E	65	60	10	3	54	30	29	3	59	18	15	18	0.5728	18
2	CCCma-CanESM2_GERICS-REMO2015	10	62	6	35	26	44	21	34	16	38	33	42	0.5692	19
33	IPSL-IPSL-CM5A-MR_SMHI-RCA4	68	58	1	1	46	15	34	1	54	50	23	16	0.5692	20
25	ICHEC-EC-EARTH_ETH-COSMO-crCLIM-v1-1	62	55	49	9	31	62	10	19	18	8	9	40	0.5634	21
35	MIROC-MIROC5_GERICS-REMO2015	23	14	51	43	11	57	17	48	37	19	29	33	0.5516	22
21	ICHEC-EC-EARTH_ETH-COSMO-crCLIM-v1-1	19	9	39	15	39	37	24	61	39	25	28	49	0.5493	23
62	MPI-M-MPI-ESM-LR_SMHI-RCA4	12	38	52	29	21	22	35	9	22	47	46	53	0.5469	24
70	NCC-NorESM1-M_MOHC-HadREM3-GA7-05	30	3	9	10	41	28	59	7	47	60	39	55	0.5446	25
23	ICHEC-EC-EARTH_KNMI-RACMO22E	17	4	38	26	24	31	19	60	29	45	32	63	0.5446	26
5	CNRM-CERFACS_CNRM-ALADIN63	44	35	63	52	50	6	14	45	52	3	6	22	0.5399	27
31	IPSL-IPSL-CM5A-MR_IPSL-WRF381P	49	50	61	4	60	49	4	6	42	24	35	11	0.5364	28
18	ICHEC-EC-EARTH_MOHC-HadREM3-GA7-05	53	48	41	59	9	34	12	12	21	52	31	23	0.5364	29
11	ICHEC-EC-EARTH_CCLM4-8-17	8	25	45	48	7	45	30	51	57	27	10	43	0.5352	30
16	ICHEC-EC-EARTH_IPSL-WRF381P	64	29	24	54	2	29	28	32	19	56	34	29	0.5305	31
64	NCC-NorESM1-M_CNRM-ALADIN63	34	37	22	6	59	61	27	13	25	42	41	34	0.5293	32
61	IPSL-IPSL-CM5A-MR_GERICS-REMO2015	61	59	12	25	58	21	8	22	36	49	4	50	0.5246	33
30	CNRM-CERFACS_ETH-COSMO-crCLIM-..	16	43	43	67	48	19	13	41	44	28	18	26	0.5235	34

Table 1 (continued)

	RCM Names	pr				tmax				tmin				RM	Rank
		DJF	SON	JJA	MAM	DJF	SON	JJA	MAM	DJF	SON	JJA	MAM		
12	ICHEC-EC-EARTH_ETH-COSMO-crCLIM-v1-1	57	18	55	64	5	53	15	55	40	1	42	1	0.5235	35
15	ICHEC-EC-EARTH_ICTP-RegCM4-6	56	32	30	69	8	20	45	33	32	31	30	21	0.5223	36
17	ICHEC-EC-EARTH_KNMI-RACMO22E	39	28	37	60	19	36	25	56	35	22	26	28	0.5176	37
3	CNRM-CERFACS_CCLM4-8-17	11	51	27	53	56	25	3	35	51	40	13	48	0.5153	38
20	ICHEC-EC-EARTH_UHOH-WRF361H	40	17	47	57	10	50	51	50	23	44	5	24	0.5094	39
63	NCC-NorESM1-M_ETH-COSMO-crCLIM-v1-1	20	1	42	17	47	54	54	46	15	34	44	44	0.5094	40
40	MOHC-HadGEM2-ES_GERICS-REMO2015	4	16	64	20	64	63	68	67	26	5	22	12	0.4941	41
28	ICHEC-EC-EARTH_SMHI-RCA4	60	68	25	11	43	48	11	14	43	61	19	30	0.4918	42
50	MPI-M-MPI-ESM-LR_DMI-HIRHAM5	66	42	7	46	22	10	55	38	10	37	64	38	0.4894	43
38	MOHC-HadGEM2-ES_CNRM-ALADIN63	24	52	26	13	29	7	5	8	70	70	70	74	0.4742	44
22	ICHEC-EC-EARTH_DMI-HIRHAM5	43	21	62	27	34	33	23	57	14	53	40	45	0.4695	45
14	ICHEC-EC-EARTH_GERICS-REMO2015	59	44	23	65	16	51	32	23	20	54	16	54	0.4636	46
61	MPI-M-MPI-ESM-LR_GERICS-REMO2015	6	56	34	5	49	39	41	18	45	55	50	61	0.4613	47
26	ICHEC-EC-EARTH_DMI-HIRHAM5	67	69	36	30	37	55	18	36	17	29	7	62	0.4566	48
6	CNRM-CERFACS_DMI-HIRHAM5	58	61	48	68	52	8	7	43	58	10	14	37	0.4554	49
66	NCC-NorESM1-M_GERICS-REMO2015	14	13	13	21	40	59	47	21	61	58	57	60	0.4554	50
48	MPI-M-MPI-ESM-LR_ETH-COSMO-crCLIM-v1-1	26	19	3	23	4	24	56	40	68	66	67	69	0.4542	51
8	CNRM-CERFACS_KNMI-RACMO22E	21	64	57	63	55	18	22	53	60	9	27	32	0.4354	52
13	ICHEC-EC-EARTH_DMI-HIRHAM5	52	49	44	50	15	41	40	42	28	48	24	52	0.4308	53
47	MPI-M-MPI-ESM-LR_CCLM4-8-17	47	33	17	18	14	26	43	17	69	68	66	68	0.4296	54
69	NCC-NorESM1-M_KNMI-RACMO22E	54	26	40	33	51	17	9	28	46	64	59	59	0.4296	55
67	NCC-NorESM1-M_ICTP-RegCM4-6	48	7	14	66	62	52	31	15	64	57	36	47	0.4143	56
42	MOHC-HadGEM2-ES_IPSL-WRF381P	1	47	31	31	23	23	48	31	66	67	69	73	0.4014	57
65	NCC-NorESM1-M_DMI-HIRHAM5	63	45	46	37	53	47	57	44	12	35	38	35	0.3991	58
7	CNRM-CERFACS_GERICS-REMO2015	27	57	58	62	45	56	53	52	63	16	11	15	0.3955	59
27	ICHEC-EC-EARTH_KNMI-RACMO22E	51	67	60	32	28	60	20	49	34	65	20	39	0.3838	60
71	NCC-NorESM1-M_SMHI-RCA4	55	30	11	14	44	38	49	29	65	62	65	65	0.3815	61
60	MPI-M-MPI-ESM-LR_ETH-COSMO-crCLIM-v1-1	18	63	54	28	32	40	46	26	56	59	53	57	0.3756	62
44	MOHC-HadGEM2-ES_MOHC-HadREM3-GA7-05	13	5	68	19	71	71	70	65	27	36	60	36	0.3650	63
68	NCC-NorESM1-M_IPSL-WRF381P	35	36	16	36	61	58	62	59	53	41	54	58	0.3322	64
46	MOHC-HadGEM2-ES_UHOH-WRF361H	9	6	65	45	65	64	63	64	72	72	72	70	0.2171	65
36	MOHC-HadGEM2-ES_CCLM4-8-17	70	70	70	70	66	68	67	71	41	51	1	25	0.2136	66
43	MOHC-HadGEM2-ES_KNMI-RACMO22E	15	24	67	40	67	66	69	70	67	71	68	66	0.1901	67
45	MOHC-HadGEM2-ES_SMHI-RCA4	22	39	66	22	69	67	66	63	71	69	74	67	0.1843	68
37	MOHC-HadGEM2-ES_ETH-COSMO-crCLIM-v1-1	71	71	71	71	70	70	71	68	38	63	17	41	0.1526	69

Table 1 (continued)

	RCM Names	pr				tmax				tmin				RM	Rank
		DJF	SON	JJA	MAM	DJF	SON	JJA	MAM	DJF	SON	JJA	MAM		
41	MOHC-HadGEM2-ES_ICTP-RegCM4-6	36	40	59	47	63	65	64	66	73	73	71	71	0.1455	70
39	MOHC-HadGEM2-ES_DMI-HIRHAM5	28	22	69	55	68	69	65	69	74	74	73	72	0.1338	71

3.1 Ranking of RCMs

Within intelligent data analysis, one of the first phases is data pre–processing. In this instance, a selection of characteristics was applied to create an RCM ranking and to select those with the most relevant information for the attainment of a reliable predictive model. The procedure followed entails filter-wrapper processing, which consists of two parts: the filter part and the wrapper part. Initially, a ranking was created using a quantitative measure (filter part), and subsequently, the most relevant ones were selected (wrapper part– Sect. 3.3). Thus, the following procedure was applied to rank the RCMs according to their performance based on the observational data: The time series of pr, tmax, and tmin were divided into the four seasons representative of the Atlantic climate of the study area, namely, winter (DJF), spring (MAM), summer (JJA) y autumn (SON). For each variable and season the TSS (Taylor Skill Score, Taylor 2001) was calculated (filter index). The TSS provides a quantitative measure of the ability of each RCM to simulate the variables pr, tmax, and tmin. It is based on the correlation and the ratio of the standard deviation of the RCMs against the observations of a given climate variable:

$$TSS = \frac{4(1 + R)^4}{(\sigma_f + 1/\sigma_f)^2(1 + R_0)^4} , \tag{1}$$

where σ_f refers to the ratio of the standard deviation of the RCMs versus the observations, R refers to the Pearson correlation coefficient, and R_0 represents the maximum value of the correlation, namely 1. TSS ranges from 0 to 1. A higher value indicates better simulation performance, while a lower value indicates worse performance. Based on the TSS results, twelve rankings were obtained, one per variable and season, which were taken into account to calculate the metric value rating RM (Ahmed et al. 2020):

$$RM = 1 - \frac{1}{nm} \sum_{i=1}^{n} rank_i , \tag{2}$$

where n and m represent the number of RCMs and seasons, respectively, while $rank_i$ refers to the number of the ranking corresponding to the member at the ith season. Finally, the RCM members were ordered according to the RM. As a result, we obtained a ranking of the RCM models ordered from best to worst according to their performance in relation to observational data in the studied basin.

3.2 Development of SEM and ML–MME Algorithms

After developing the ranking of the RCMs, the MMEs structure and characteristics have been designed. In the

first place, when formulating ML–MME algorithms, it is crucial to account for the seasonal dynamics inherent in the variables. This consideration enhances the algorithms' ability to discern patterns of variability. Due to the evident interannual temperature dynamics in out mid-latitude region, we have opted to consider the seasons independently, specifically for tmin and tmax when constructing the ML–MME algorithms (Morales-García et al. 2023; Ahmed et al. 2020). Conversely, with precipitation, we have pursued an alternative strategy: Given the complex nature of this variable and the alterations observed in the annual cycle over recent decades in European mid-latitudes (Christidis and Stott 2022; Paluš et al. 2005), establishing clear seasonal patterns becomes a more intricate task. Designing ML algorithms solely based on the seasons might prove misguided, potentially hindering the algorithms from accurately capturing the variable's behaviour.

To address this complexity and unbalance of the data, we have chosen to consider monthly precipitation events categorising them into two subgroups Chao et al. (2018): those exceeding the 80th percentile and those below it according to observational data. Through the separation of precipitation into two distinct databases, the range of the variable was reduced, leading to increased accuracy in the results obtained by the ML models. Following this rationale, each ML–MME technique has resulted in four algorithms for tmax and tmin, corresponding to each season. Additionally, two algorithms have been generated for precipitation: one for events within the 0–80 percentile interval and another for events in the 80–100 percentile interval.

Different methods were applied to construct the MME on a monthly scale, including, on the one hand, SEM, and on the other hand, three ML techniques: RF, GB and LR. The first MME development technique is the SEM, commonly and widely used for MME calculation (Clark 2017). The remaining three techniques are more elaborate and are based on ML regression models. These three techniques are detailed below:

- Random Forest (RF). RF is a machine learning technique whose basis is a combination of predictor trees such that each tree depends on the values of a random vector tested independently and with the same distribution for each of them. It is a substantial modification of bagging that builds a large collection of uncorrelated trees and then averages them. The algorithm for inducing a random forest was developed by Breiman (2001). Bagging is the ensemble learning method typically used to reduce the variance within a noisy data set. The RF method combines the idea of bagging and random attribute selection to build a collection of decision trees with controlled variation. The selection of a random subset of attributes is an example of the random subspace method, a way to perform stochastic discrimination (Breiman 2001).

- Gradient Boosting (GB). GB is a machine learning technique for regression analysis and statistical classification problems based on boosting. Boosting consists of combining the results of several weak classifiers to obtain a robust classifier. When these weak classifiers are added, they are added in such a way that they have different weights depending on the accuracy of their predictions. After a weak classifier is added, the data changes its weight structure: cases that are misclassified gain weight and those that are correctly classified lose weight. Thus, the strong classifiers focus more strongly on the cases that were misclassified by the weak classifiers. The GB technique creates a predictive model based on weak prediction models, usually decision trees. The GB is an ensemble that provides a set of prediction models, which conclude a satisfactory prediction outperforming in some cases the random forest ensemble (Bentéjac et al. 2021).

- Linear Regression (LR). LR is a supervised learning algorithm used in machine learning and statistics. In its simplest version, it calculates a line that will indicate the trend of a continuous data set. LR can be defined as an approach to model the relationship between a dependent scalar variable and one or more explanatory variables. The LR technique should minimise the cost of a quadratic error function and those coefficients will correspond to the optimal line. There are several methods to minimise the cost. The most common is to use a vector version and the so-called Normal Equation which will give a direct result (Weisberg 2005).

For the selection of the hyperparameters of the machine learning techniques, a Grid has been used by means of cross-validation to sweep through all the parameters and thus select the most optimal ones.

3.3 Selection of RCMs

After the RCM ranking was completed and the MMEs characteristics defined, the process of selecting the optimal number of RCMs to be considered when creating the MMEs for each variable (tmax, tmin and pr) was initiated. This process is the wrapper part of feature selection presented in Fig. 1. The MMEs were developed considering the RM-based rank of RCMs from 1 to 40 (Table 1). Initially, only the outputs of the RCM with a rank of 1 were used to provide inputs to the MME. Subsequently, the outputs of the RCM with a rank of 2 were added to the input set, followed by the incremental introduction of RCMs with overall ranks 3, 4, 5... 40 into the input set, one RCM at a time. This approach, known as the top-ranked approach (Ahmed et al. 2020), started with the best-performing RCM (rank 1) and

progressed with subsequent RCMs in ascending order of their RM-based rank.

The evaluation of the performance of the MME outputs, generated with varying numbers of RCMs, has been conducted on the reconstructed time series. This reconstruction of the results obtained by the MME has been carried out by transforming the data divided into seasons (tmax, tmin) or percentile intervals (pr) described in Sect. 3.2. into a time series.

The evaluation metric was the Modified Index of Agreement (*md*, (3)), which was initially proposed by Willmott (1981) and has been later widely applied (Ahmed et al. 2020). It ranges from 0 to 1, with higher values indicating a better fit of the model

$$md = 1 - \frac{\sum_{i=1}^{n}(x_{\text{obs},i} - x_{\text{sim},i})^j}{\sum_{i=1}^{n}(|x_{\text{sim},i} - \bar{x}_{\text{obs}}| + |x_{\text{obs},i} - \bar{x}_{\text{obs}}|)^j} , \quad (3)$$

where $x_{\text{sim},i}$ and $x_{\text{obs},i}$ are the ith data point in the simulated RCM and the observed data series of a climate variable, respectively. It has been calculated for the four grid cells considered in this study.

With this procedure all RCMs are incoporated into the MMEs. Then the cut-off point is made just at the RCMs that start to worsen the *md* metric or when an overfitting issue is observed. This indicates that from that RCM onwards, the information provided by the other RCMs is more noisy than beneficial.

3.4 Evaluation of SEM and ML–MME Algorthims

Once the selection phase was completed and the definitive MMEs were built, the evaluation was carried out. The data was divided in the training and testing phases, representing 80 % and 20 % of the data, respectively, divided chronologically. Therefore, the training phase covered the period of 1980–2006 while the test phase covered the period of 2007–2015. Notably, data from all four points in the mesh have been incorporated to feed the algorithms. Moreover, the evaluation was carried out with three additional metrics commonly used in the characterisation of time series similarities: the coefficient of determination (R^2), the root-mean-square error (RMSE), and the root mean square percentage error (RMSEPE).

3.5 Application of ML–MME data to Temez Hydrological Model

The Temez model (Témez 1977), extensively applied in Spanish watersheds (Pérez-Sánchez et al. 2019; Escriva-Bou et al. 2017; Chavez-Jimenez et al. 2013; García-Barrón et al. 2015; Jódar et al. 2017; Marcos-Garcia et al. 2017; Senent-Aparicio et al. 2018), falls within the category of aggregated

watershed simulation models (Estrela 1992). Operating from the onset of rainfall to the initiation of runoff and subsequent discharge into rivers, the Temez model manages moisture balances across interconnected processes within a hydrological system. Input variables for the Temez model encompass the spatial average monthly precipitation for the entire basin and Potential Evapotranspiration (ETP). In line with the current investigation's focus on monthly climate data, ETP was determined using the Thornthwaite method (Thornthwaite 1948).

We assessed the hydrological model's outcomes based on four widely adopted evaluation criteria in hydrological research (Jimeno-Sáez et al. 2018). These criteria include the Nash–Sutcliffe Efficiency coefficient (NSE), the percent bias (PBIAS), the Pearson correlation coefficient (*r*), and the Kling–Gupta Efficiency coefficient (KGE).

After the evaluation of the four proposed ML–MME techniques, the algorithms were applied to future climate projections for the RCP8.5 emission scenario for long-term future and were utilized as input data for simulating future streamflow.

4 Results and Discussion

4.1 Ranking of RCMs

Table 1 presents the RCM rankings based on TSS across the DJF, MAM, JJA, and SON seasons for the variables tmin, tmax, and pr. Notably, substantial variations emerge among seasons and variables. In certain instances, an RCM that excels in simulating one variable and season finds itself at the lower end of the ranking when compared to other variables and seasons. A case in point is IPSL–RCA4 (Code 33), which stands out as the top performer in simulating precipitation during SON and JJA, as well as maximum temperature in SON. However, it exhibits inefficiencies in comparison to other RCM members when simulating precipitation in DJF and MAM (Kotlarski et al. 2014).

A notable observation is the high contribution of the GCM driver on the ranking position, which is in line with what is stated by Vautard et al. (2021), who established that some variables are conditioned by large-scale boundary conditions defined by the GCMs. For instance, RCM members driven by the MPI–ESM–LR GCM consistently achieve the highest RM values (Table 1), indicating superior overall performance. This aligns with findings from Brands et al. (2013), underscoring the GCM's excellent ability to simulate precipitation over European mid-latitudes. A poor RCM performance, however, can also have a significant impact on the simulation, as in the case of the 60 and 48 models which, despite having the MPI as driver, occupy poor positions in the ranking. In the same way, RCMs with CNRM–CM5 as

driver also rank high, because they are able to adequately characterise the temperatures (McSweeney et al. 2015). Conversely, a GCM with deficiencies in simulating climate conditions adversely affects the ranking of RCMs that are driven by it. Such is the case with MOHC–HadGEM2, which exhibits notable biases in climate variables representation. Consequently, MOHC–HadGEM2 attains lower positions across all variables and seasons.

4.2 Selection of the Optimal Number of RCMs

To extract meaningful insights for determining the optimal number of RCMs to include in further analyses, we conducted an examination of the ML–MME learning curve. All machine learning techniques described previously have been used to select the number of RCMs. As depicted in Fig. 2, the *md* values, relative to observations, are plotted against the number of RCMs utilised to construct the SEM and the ML–MMEs. The incorporation order of RCMs follows a top-ranked approach (Ahmed et al. 2020). Notably, for fewer than three RCMs, the *md* values exhibit a substantial increase initially, stabilising thereafter to an asymptotic trend for most ML techniques across all variables and periods. An exception is observed with GB, where, beyond
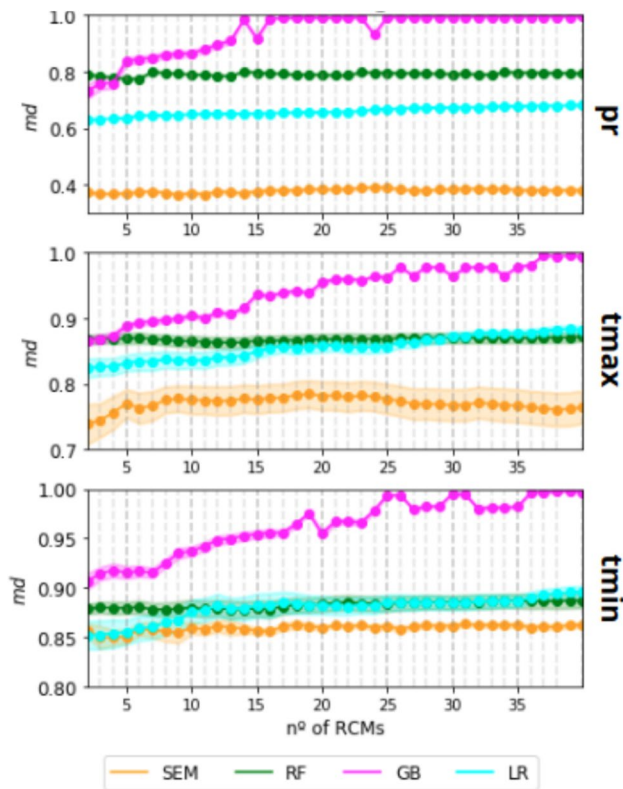
Fig. 2 *md* vs. the number of RCMs for precipitation (pr), maximum temperature (tmax) and minimum temperature (tmin). The shaded area represents the standard deviation of the four grids of the mesh

a certain quantity of RCMs (for pr 16, for tmax 35 and for tmin 25), the *md* values approach 1. This indicates overfitting (Ying 2019; Dietterich 1995).

Upon closer examination of individual variables, precipitation stands out with notable differences between SEM and ML–MME. SEM records *md* values near 0.4, while ML–MME techniques yield values ranging from 0.6 to 0.8 (excluding the overfitting case of GB). For temperature variables, the initial *md* is higher, approximately 0.6, indicating that RCMs exhibit a greater capacity to replicate monthly temperature patterns compared to precipitation. This difference arises due to the higher complexity inherent in the dynamics of precipitation, which poses challenges for numerical models to simulate accurately (Perkins et al. 2007; Aghakhani Afshar et al. 2017), specifically affecting RCMs (Vautard et al. 2021; Herrera et al. 2020; Kotlarski et al. 2014). While improvements are observed in temperature variables with ML–MME, the contrast in *md* values is less pronounced, particularly for minimum temperature.

After reviewing the evolution of result improvements concerning the number of RCMs, and recognising a plateau after the initial progress, we opted to include a total of seven RCMs. This decision is motivated also to avoid instances of overfitting, as observed with GB for tmin variable, while maintaining a balance between model complexity and predictive performance. The number of models utilized aligns with the findings of Dey et al. (2022), who, following a preselection process, incorporated 5 models into their analysis. Likewise, Ahmed et al. (2020) achieved comparable results in their precipitation analysis, drawing from data generated by 7-10 models exhibiting high performance.

4.3 Evaluation of SEM and ML–MMEs

Figures 3, 4, and 5 offer an assessment of the SEM and ML–MME results relative to CLIMPY observations for the variables pr, tmax, and tmin. To enhance result clarity, we focused on evaluating the spatial average of pr, tmax, and tmin within the study area. Notably, in the first column of Fig. 3, the Taylor diagram for precipitation during both the training and test periods indicates substantial enhancements resulting from ML–MME application compared to SEM. Among the ML–MME methods, RF and LR yield comparable outcomes, while GB achieves the most favourable results at the annual scale for both training and test periods.

Concerning the spatial average of temperatures, Taylor diagrams do not reveal appreciable improvements. Both SEMs of tmin and tmax already exhibit statistics indicative of a robust representation of monthly temperatures in the study area, attributed to the high-quality simulations of the pre-selected RCMs (Table 1). The exceptional starting point of RCMs' simulation quality may limit the potential enhancement capacity that ML–MMEs could offer.
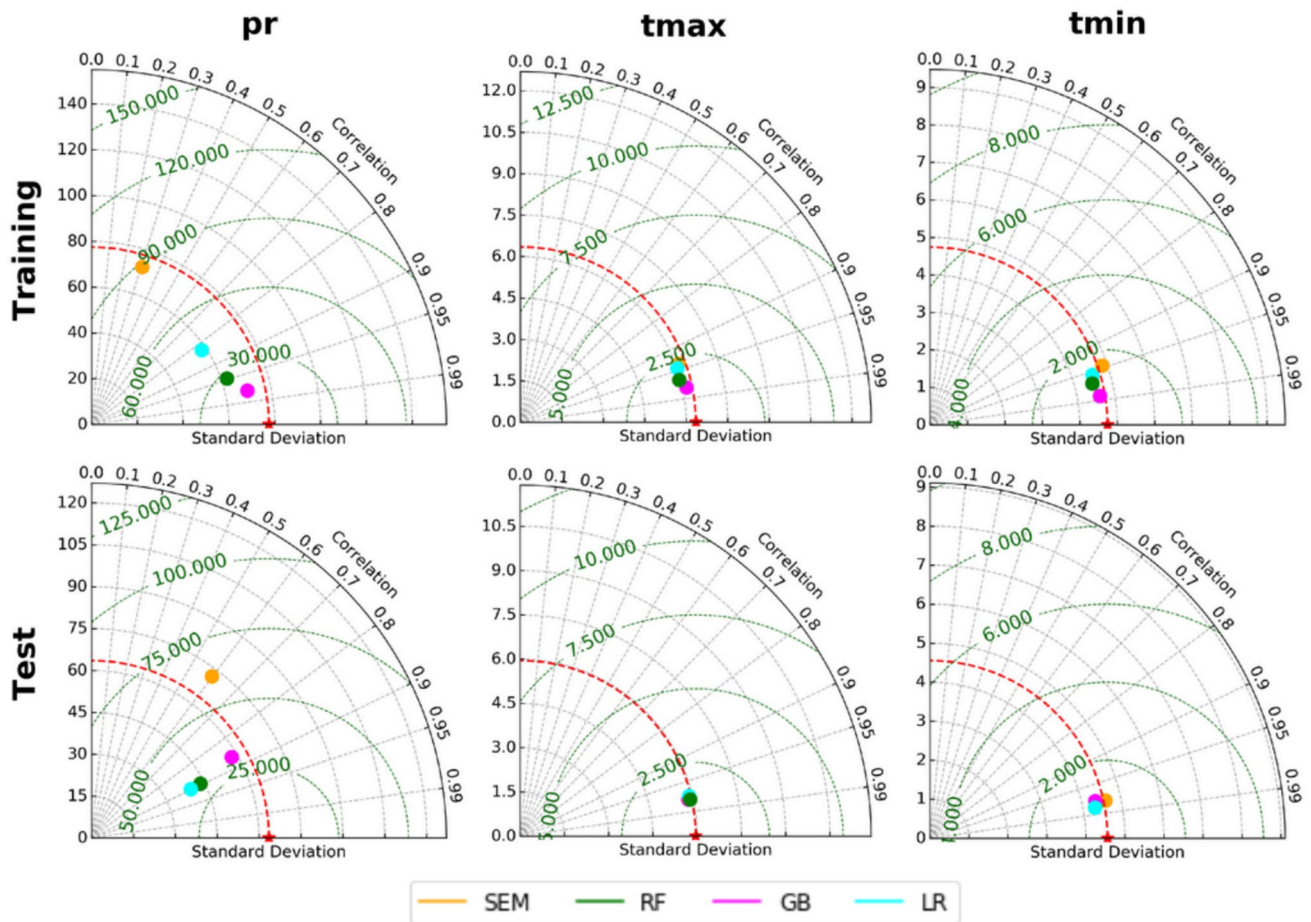
Fig. 3 Taylor diagrams of the spatial average of the variables precipitation (pr), maximum temperature (tmax) and minimum temperature (tmin) for the training (1980–2006) and test (2007–2015) periods

For a more detailed analysis of precipitation performance, Fig. 4 presents monthly time series plots of the spatial average results for SEM and ML–MME. The improvement across all ML–MMEs in comparison to SEM is evident. Whereas SEM exhibited a fit close to zero, high RMSE, and *md* below 0.5 in both periods, all ML–MME techniques demonstrate significantly improved performance, indicating their superior ability to simulate monthly precipitation patterns. Notably, GB achieves the best *md* results, with values of 0.88 and 0.75 for the training and test periods, respectively. RF, however, is not far behind, boasting an R^2 in the test period of 0.80, surpassing GB's 0.75. Despite LR showing higher RMSE values (around 44 mm/month) and a lower capacity to detect precipitation minima and maxima, the ML–MME based on LR markedly improves the representation of the study area's precipitation compared to SEM. These results are in line with those obtained in several studies (Acharya et al. 2014; Salman et al. 2018; Li et al. 2021). For instance, Dey et al. (2022) developed ML-based MME approaches for CMIP6 in an Indian River basin

obtaining that the RF-based ML–MME showed improved performance compared to SEM. In the same vein, Jose et al. (2022) proposed RF as the best suitable ML model over India for creating MME and simulating the past observed climate variables, in a tropical river basin. In addition to studies conducted at basin scales, ML–MME approaches have also been applied at broader spatial scales. This is the case of Wang et al. (2018) who applied SEM, BMA, RF, and SVM with CMIP5 data over Australia, concluding that RF and SVM could generate better-performing MMEs compared to SEM and BMA.

Figure 5 provides a thorough evaluation of SEM, ML–MMEs, and the seven individual RCMs, both at the annual and seasonal scales during the test period. Notably, when comparing SEM with the ML–MME techniques, a widespread enhancement is observed, particularly in precipitation. For instance, the DJF season, which records the lowest *md* values (around 0.2) for individual RCMs, sees substantial improvement with ML–MME techniques, elevating *md* to approximately 0.55 for RF and LR, and
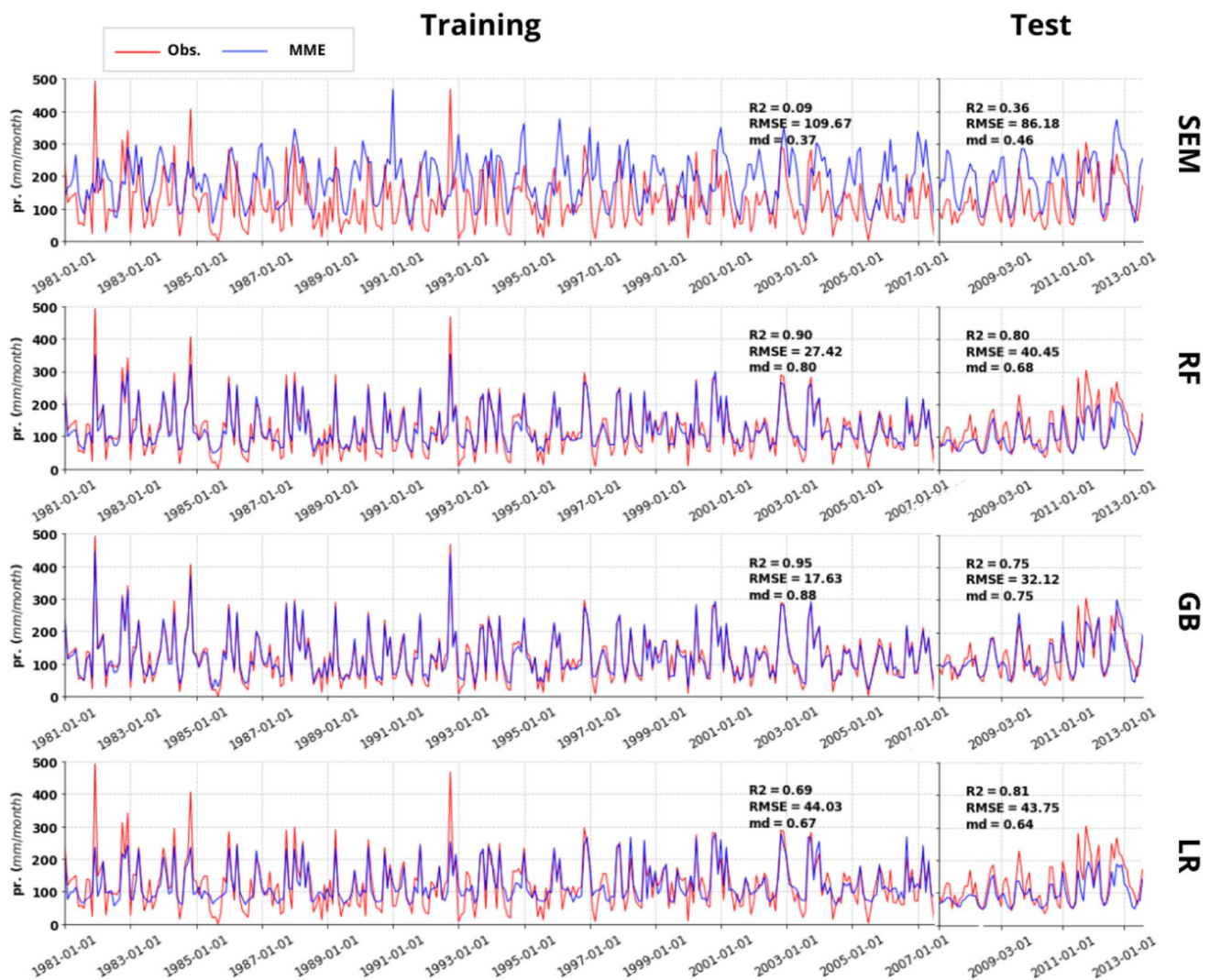
Fig. 4 Spatially averaged observed precipitation and simulated precipitation time series and evaluation metrics (SEM and ML–MME) for the training (1980–2006) and test (2007–2015) periods

surpassing 0.70 for GB. This improvement is consistent across all seasons and holds true for annual data as well. Similarly, R^2 and RMSE exhibit substantial enhancements across the board for precipitation. The coefficient R^2, which occasionally dips to 0 for certain RCMs and seasons, now consistently remains above 0.6 for all seasons and ML–MME techniques, reaching annual values of 0.8. The RMSEPE, expressed as a fraction, which exceeds 3 in some individual RCMs, is consistently below 1 for all ML–MME cases. This noticeable and significant improvement in the characterization of precipitation at both seasonal and annual levels, as evidenced by the three metrics analyzed in the study region, represents a significant qualitative advantage offered by ML-MMEs compared to the results obtained from individual RCM members. This enhancement could potentially yield significant benefits

for regional planning, including water and agricultural management, as well as climate risk preparedness, among others.

For temperatures, while no notable seasonal improvement is evident in r and md, annual values display enhancement for both tmax and tmin. However, the improvement in simulation quality, even at the seasonal scale, is manifested as a decrease in the RMSE values. Individual RCMs exhibit RMSE values ranging from 2.0 °C to 5.2 °C for tmax. Post-application of ML–MME techniques, RMSE is drastically reduced, with values between 0.8 °C and 3 °C. A parallel behaviour is observed for tmin. This improvement in temperature representation is of particular interest in an area like the analyzed study region, where the presence of snow and snowmelt processes are key factors directly dependent on temperatures, greatly influencing regional management.
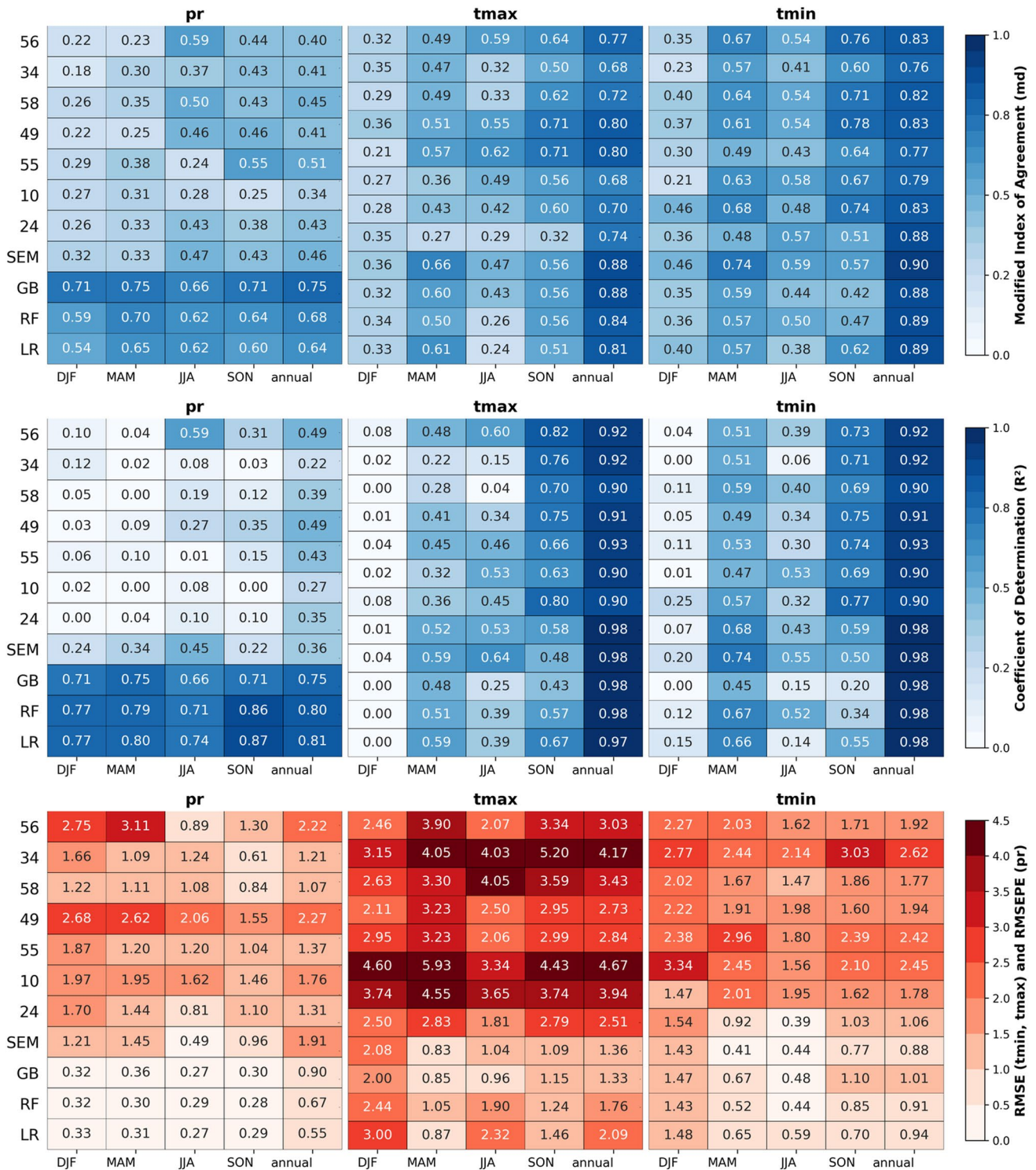
Fig. 5 Heat maps representing the *md*, R^2, RMSE (tmax, tmin) and RMSEPE (pr) obtained from the comparison of the observations versus the SEM, the ML–MMEs and the individual RCMs for the test (2007–2015) period

In each examined case, MMEs consistently outperform individual members, even when represented by the least effective MME, SEM. This observation is supported by numerous studies that emphasise the MME's ability to

enhance individual member performance and reduce climate output uncertainties. Notable analyses include regions such as India (Gusain et al. 2019), the USA (Srivastava et al. 2020), China (Zhuang et al. 2016), and Europe (Evin et al.

2021). Additionally, our results indicate that ML–MME exhibits superior performance to SEM, particularly for precipitation, as depicted in Figs. 4 and 5. This finding underscores the ML–MME's relevance at the catchment scale. The enhanced performance of ML–MME over SEM may be attributed to ML approaches' capacity to address nonlinear, high-dimensional correlations between climate model outputs and observational datasets (Dey et al. 2022). Moreover, as highlighted by Li et al. (2021), ML–MME algorithms could be able to capture detailed information at local scales due to the incorporation of high-resolution observations on the construction of ML–MME algorithms.

In this study, we successfully integrate the EURO-CORDEX RCMs, the climatic simulations with higher spatial resolution for the study area, with the strengths of ML mathematical algorithms. This combination holds promise for reducing uncertainty in basin-scale climate projections. In the following section (Sect. 4.4), we utilise the outputs of the ML–MME algorithms to feed a hydrological model within the Esca River basin.

4.4 Application of SEM and ML–MME Climate Data to Temez Hydrological Model

4.4.1 Temez Model Setup

For the model setup development, the simulation period was divided into two distinct phases: the calibration period, spanning from 1981 to 2000, and the subsequent validation period, covering 2001 to 2014. A warm-up year was introduced to attain a stable state for the Temez model. Calibration focused on adjusting four key parameters: H_{max} (maximum soil storage capacity), C (surplus starting coefficient), I_{max} (maximum infiltration) and α (groundwater contribution coefficient). The first two parameters govern soil storage regulation, the third distinguishes surface runoff from groundwater runoff, and the fourth modulates subsurface drainage (Murillo and Navarro 2011). Table 2 presents the metrics described in Sect. 3.4 for the comprehensive assessment of hydrological simulation.

According to what was established by Moriasi et al. (2007) and Brighenti et al. (2019), the performance of the model both in the calibration and validation period is satisfactory since the results of NSE and KGE exceed 0.5 and the PBIAS reaches its maximum in the calibration period with −12.76 %, remaining below the ±25 %.

4.4.2 Evaluation of Streamflow for SEM and ML–MME Input Data

Starting from the calibrated and validated Temez model, the simulations described bellow have been carried out in order to evaluate the impact of the climate-corrected data, which have been analysed in detail in section 4.3, on the characterisation of the flow variable. First, the monthly flow simulation has been developed by feeding the Temez model with data from precipitation observations and with the ETP derived from the tmax and tmin observations, denoted as $Q_{sim\text{-}OBS}$. Following the same approach, four additional flow simulations, subsequently identified as $Q_{sim\text{-}SEM}$, $Q_{sim\text{-}GB}$, $Q_{sim\text{-}LR}$ and $Q_{sim\text{-}RF}$, were developed. Each simulation incorporated input data derived from MME techniques: SEM, GB, LR, and RF, respectively. To facilitate the explanation, another term has been incorporated that refers to the group formed by the simulated flows using the climate data derived from the ML–MME, $Q_{sim\text{-}ML\text{-}MME}$.

Table 3 presents the statistics of the described simulations for the training period (1980–2006) and test period (2007–2015) of the ML–MME algorithms. The choice of these specific periods aligns with the study's focus on improving climate representation through ML–MME

Table 2 Calibration (1981–2000) and validation (2001–2014) results for the Temez hydrological model

	NSE	r	RMSE	KGE	PBIAS
Calibration	0.63	0.85	13.27	0.78	− 12.76
Validation	0.67	0.83	13.08	0.82	7.21

The presented statistics are the Nash–Sutcliffe Efficiency coefficient (NSE), the Pearson correlation coefficient (r), the Root Mean Square Error (RMSE), the Kling–Gupta Efficiency coefficient (KGE), and the Percent Bias (PBIAS)

Table 3 Statistics of simulated vs. observed streamflows for the training (1980–2006) and test (2007–2015) periods

	Training				Test			
	NSE	r	RMSE	KGE	NSE	r	RMSE	KGE
$Q_{sim\text{-}OBS}$	0.67	0.85	12.55	0.81	0.60	0.82	15.00	0.78
$Q_{sim\text{-}SEM}$	− 1.84	0.59	36.59	− 0.27	− 1.97	0.58	40.95	− 0.36
$Q_{sim\text{-}GB}$	0.69	0.85	12.08	0.81	0.48	0.74	17.13	0.73
$Q_{sim\text{-}LR}$	0.56	0.77	14.48	0.69	0.52	0.74	16.42	0.61
$Q_{sim\text{-}RF}$	0.66	0.83	12.59	0.76	0.61	0.80	14.86	0.63

The presented statistics are the Nash–Sutcliffe Efficiency coefficient (NSE), the Pearson correlation coefficient (r), the Root Mean Square Error (RMSE), and the Kling–Gupta Efficiency coefficient (KGE)

techniques and assessing the extent to which these improvements influence streamflow representation. The congruence in analysis periods for both climate variables and flow variables enhances the study's coherence. From the analysis of the statistics in Table 3 it is concluded that while the $Q_{\text{sim-SEM}}$ obtains unsatisfactory results for both periods, the ML–MMEs manages to enhance the representation of the flow significantly. Notably, both $Q_{\text{sim-RF}}$ and $Q_{\text{sim-GB}}$ exhibit statistics comparable to $Q_{\text{sim-OBS}}$, with NSE remaining above 0.60 for the training period and r achieving values exceeding 0.74 in both periods. The $Q_{\text{sim-LR}}$ simulation, although satisfactory, yields inferior results with higher PBIAS and lower NSE and KGE values. These outcomes signify that the improvements in climate variable representation by ML–MMEs propagate and significantly enhance flow characterisation in both the training and test periods.

To further assess the performance of the hydrological simulations, the annual cycle for the four $Q_{\text{sim-ML–MME}}$ together with the $Q_{\text{sim-OBS}}$ and Q_{OBS} have been depicted in Fig. 6. The latter refers to the observed flow rates. It is observed how in the training period (1980–2006) the annual cycle of the streamflow consists of two maxima in January and May and a minimum recorded in August and September. This interannual dynamics is captured by the calibrated and validated Temez model for the $Q_{\text{sim-OBS}}$ simulation. If we pay attention to the $Q_{\text{sim-MME}}$, we observe that while Q_{SEM} fails to characterise the annual cycle with a generalised overestimation of the flow that extends over most of the year, the other

$Q_{\text{sim-MME}}$ accurately reproduce the hydrological cycle of the Esca river. The annual cycle of the test period (2007–2015) presents differences with respect to the training period, especially in the spring maximum, which is more accentuated and reaches 60 Hm^3. The Temez model with input data from climate observations ($Q_{\text{sim-OBS}}$) has more difficulty in simulating the hydrological cycle for this period, although it roughly succeeds in characterising it. The $Q_{\text{sim-ML–MME}}$ simulations accurately reproduce the $Q_{\text{sim-OBS}}$ cycle, especially $Q_{\text{sim-GB}}$, while $Q_{\text{sim-SEM}}$ demonstrates poor performance. In essence, the $Q_{\text{sim-ML–MME}}$ reproduce the interannual dynamics captured by the Temez model in the $Q_{\text{sim-OBS}}$ simulation, thus demonstrating that the improvements achieved in the climate representation derived from the application of ML–MME techniques have a positive impact on the characterisation of the hydrological cycle. On the other hand, it is important to highlight that the differences derived from the flow observations (Q_{OBS}) and the simulations are attributed to the errors provided by the Temez model, probably related to the misrepresentation of snow accumulation and melting processes by the hydrological model (Jimeno-Sáez et al. 2020).

4.5 Future Projections of Climate and Hydrological Variables

Thus far, it has been demonstrated that the utilization of ML-MME techniques has not only enhanced the representation
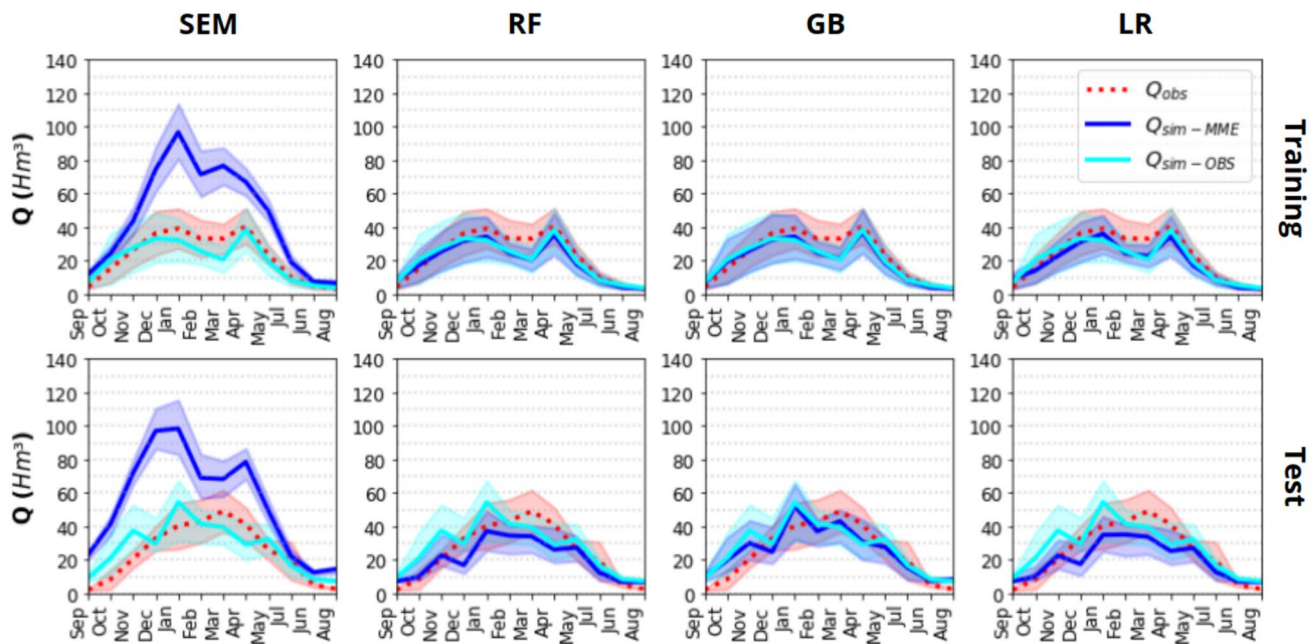

Fig. 6 Annual cycle of streamflow for the training (1980–2006) and test (2007–2015) periods. Results are shown for observational flow data (Q_{OBS}), Temez-simulated flow with input data from CLIMPY climate observations ($Q_{\text{sim-OBS}}$) and Temez-simulated flow with input data from SEM and ML–MMEs ($Q_{\text{sim-MME}}$). The shaded area represents the annual variability of the streamflow results

of climate variables but has also significantly improved the accuracy of hydrological characterization during the historical period in the study area. Extending this methodology to future scenarios under the RCP8.5 emission scenario suggests that projections from trained ML–MME models may offer more realistic data than those from the SEM (Liang et al. 2008).

Figure 7 illustrates the annual cycles of the analysed variables-pr, tmax, tmin, and Q-for two distinct periods: historical (1986–2015), and long-term future (2065–2095). This figure juxtaposes simulation data from the ML–MME techniques with observational data from the historical period. A comparative analysis reveals that ML–MME techniques better characterise climatic patterns compared to the SEM. Specifically, while the SEM tends to overestimate precipitation during DJF and MAM, the ML–MME captures the interannual dynamics more accurately, manifesting two peaks in April and November and a minimum that extends from June to August (Lemus-Canovas et al. 2019). Similarly, ML–MME techniques more precisely replicate interannual
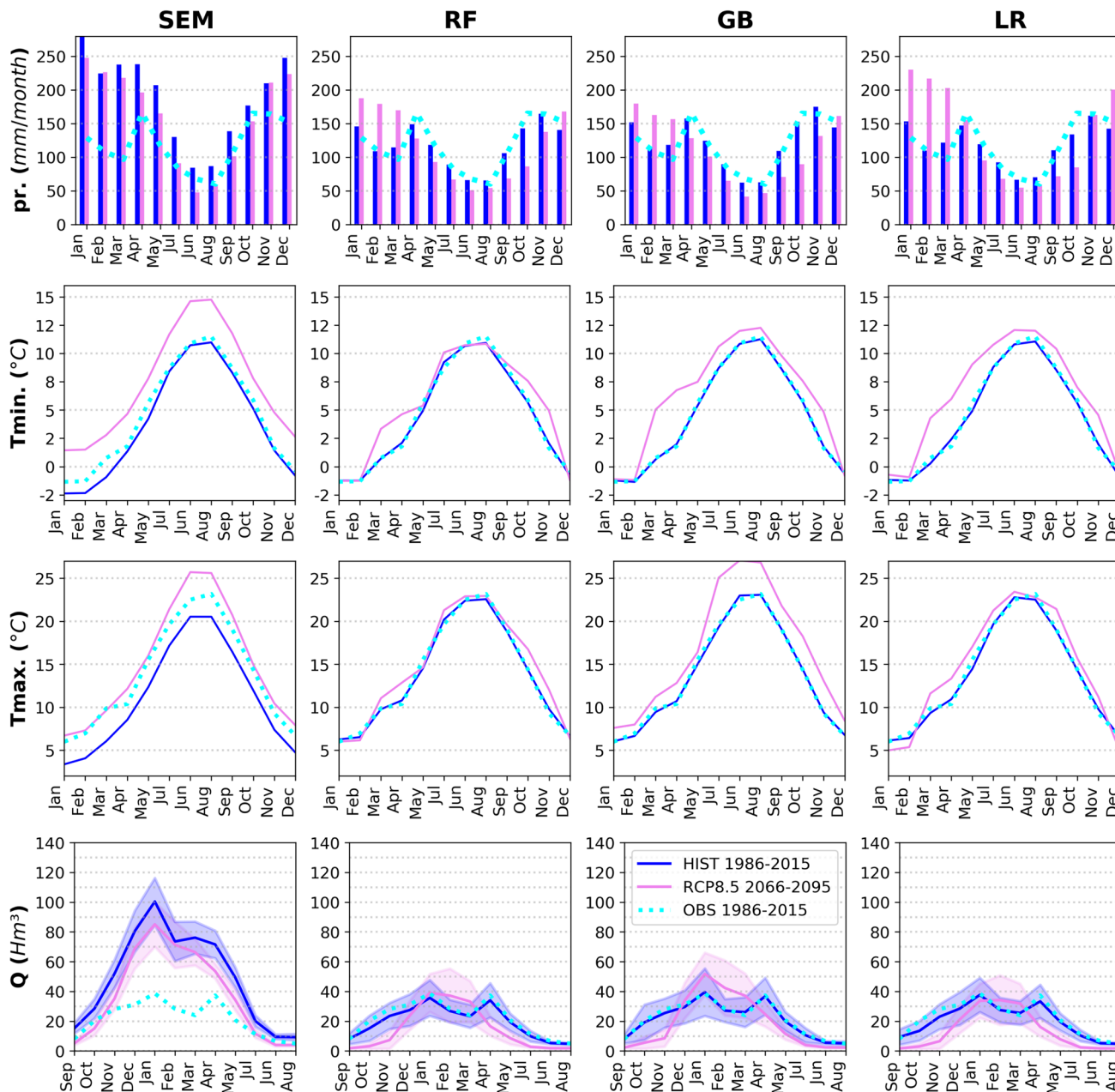
Fig. 7 Annual cycles of pr, tmin, tmax and Q for historical and long-term future (RCP8.5 emission scenario) covering 1986–2015, 2066–2095 respectively. The shaded area for *Q* variable represents the annual variability of the results

temperature variations. Further, the ML–MME techniques positively influence the streamflow annual cycle representation by the Temez model in the study area. Indeed, simulations driven by the SEM consistently exhibit overestimations, as discussed in Sect. 4.4, whereas RF–MME, GB–MME, and LR–MME demonstrate markedly superior performance (Fig. 7).

These results and those analysed in sects. 4.3. and 4.4. indicate that the ML–MME techniques provide more realistic information than SEM, also for the projections of the RCP8.5 emission scenario. If we focus on RF and GB we see that according to these projections, precipitation will decrease throughout the year except for DJF and MAM where will increase, thus modifying the interannual precipitation patterns. Concurrently, temperatures are expected to rise consistently (Amblar-Francés et al. 2020; Lemus-Canovas and Lopez-Bustins 2021), with minimum temperatures notably increasing in March and April. These shifts in interannual dynamics will likely reshape the hydrological cycle, resulting in a more pronounced summer minima and intensified, albeit shorter-duration, maxima in February and March, as projected by RF and GB and in line with the results obtained in numerous Pyrenean Rivers (López-Moreno et al. 2014; García Ruiz et al. 2001; Stahl et al. 2010; Zabaleta et al. 2017; Boé et al. 2009; OPCC-CTP 2018). While the simplicity of this hydrological modeling approach, coupled with monthly-scale analysis, limits our conclusions to informative insights, it also highlights the potential of integrating ML-MME techniques with more intricate hydrological models on a daily scale thus paving the way for the development of projections that can facilitate more precise resource-planning and adaptation strategies in the context of climate change.

5 Conclusions

In this study, we effectively implemented machine learning algorithms to develop Multi Model Ensembles (MMEs) based on Regional Climate Models (RCMs) within the Esca River basin, situated in the high mountain region of the Pyrenees. A comprehensive ranking of the RCMs was established, revealing substantial variability in performance across individual variables and seasons, with MPI-driven RCMs consistently outperforming others.

To determine the optimal number of RCMs for MME construction, a top-ranked approach was adopted. Seven RCMs were selected based on performance curves analysis, forming the definitive MMEs.

Noteworthy enhancements were observed in precipitation representation on both annual and seasonal scales by the Machine–Learning (ML) based MMEs. Although the results obtained for temperatures using ML-based MMEs are more subtle at seasonal scale, a relevant improvement is observed in the annual RMSE values. Hydrological simulations employing MMES of climate variables based on Random Forest, Linear Regression and Gradient Boosting yielded outcomes comparable to those fed by climate observations, significantly outperforming simulations based on single RCMs and SEM. Our results showcase two key findings. Firstly, they highlight the potential of machine learning techniques in constructing MMEs to enhance the characterization of climate variables. Secondly, they underscore the advantages of utilizing these ML-MMEs as input data for hydrological models.

Additionally, our methodology showcased versatility by applying algorithms to climate projections under the RCP8.5 scenario, providing more realistic information than traditional methods and thereby offering opportunities for reducing uncertainty in climate outputs for adaptation planning and basin-scale impact analyses in the context of climate change. This contribution holds particular significance and novelty in a region characterized by complex topography, such as the high mountain region of the Pyrenees, where predicting future changes is not only a complex task but also essential for the climate change adaptation of the region.

Declarations

References

Acharya N, Shrivastava NA, Panigrahi BK et al (2014) Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Clim Dyn 43(5–6):1303–1310. https://doi.org/10.1007/s00382-013-1942-2

Aghakhani Afshar A, Hasanzadeh Y, Besalatpour AA et al (2017) Climate change forecasting in a mountainous data scarce watershed using CMIP5 models under representative concentration pathways. Theoret Appl Climatol 129(1–2):683–699. https://doi.org/10.1007/s00704-016-1908-5

Ahmed K, Sachindra DA, Shahid S et al (2020) Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. Atmos Res 236(2019):104806. https://doi.org/10.1016/j.atmosres.2019.104806

Amblar-Francés MP, Ramos-Calzado P, Sanchis-Lladó J et al (2020) High resolution climate change projections for the Pyrenees region. Adv Sci Res 17:191–208. https://doi.org/10.5194/asr-17-191-2020

Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. Artif Intell Rev 54:1937–1967

Berthold MR, Borgelt C, Höppner F et al (2010) Guide to intelligent data analysis: how to intelligently make sense of real data. Springer Science & Business Media

Boé J, Terray L, Martin E et al (2009) Projected changes in components of the hydrological cycle in French river basins during the 21st century. Water Resour Res 45(8):1–15. https://doi.org/10.1029/2008WR007437

Brands S, Herrera S, Fernández J et al (2013) How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa?: A performance comparison for the downscaling community. Clim Dyn 41(3–4):803–817. https://doi.org/10.1007/s00382-013-1742-8

Breiman L (2001) Random forests. Mach Learn 45:5–32

Brighenti T, Bonuma N, Grison F et al (2019) Two calibration methods for modeling streamflow and suspended sediment with the swat model. Ecol Eng 127:103–113. https://doi.org/10.1016/j.ecoleng.2018.11.007

Calì Quaglia F, Terzago S, von Hardenberg J (2022) Temperature and precipitation seasonal forecasts over the Mediterranean region: added value compared to simple forecasting methods. Clim Dyn 58(7–8):2167–2191. https://doi.org/10.1007/s00382-021-05895-6

Careto JAM, Soares PMM, Cardoso RM et al (2022) Added value of EURO-CORDEX high-resolution downscaling over the Iberian Peninsula revisited - Part 2: Max and min temperature. Geosci Model Dev 15(6):2653–2671. https://doi.org/10.5194/gmd-15-2653-2022

Careto JAM, Soares PMM, Cardoso RM et al (2022) Added value of EURO-CORDEX high-resolution downscaling over the Iberian Peninsula revisited - Part 1: precipitation. Geosci Model Dev 15(6):2635–2652. https://doi.org/10.5194/gmd-15-2635-2022

Chao Z, Pu F, Yin Y et al (2018) Research on real-time local rainfall prediction based on MEMS sensors. J Sens 2018:1–9. https://doi.org/10.1155/2018/6184713

Chavez-Jimenez A, Lama B, Garrote L et al (2013) Characterisation of the sensitivity of water resources systems to climate change. Water Resour Manag 27(12):4237–4258. https://doi.org/10.1007/s11269-013-0404-2

Chen W, Jiang Z, Li L (2011) Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. J Clim 24(17):4741–4756. https://doi.org/10.1175/2011JCLI4102.1

Christidis N, Stott PA (2022) Human influence on seasonal precipitation in Europe. J Clim 35(15):5215–5231. https://doi.org/10.1175/JCLI-D-21-0637.1

Ciarlo JM, Coppola E, Fantini A et al (2021) A new spatially distributed added value index for regional climate models: the EURO-CORDEX and the CORDEX-CORE highest resolution ensembles. Clim Dyn 57(5–6):1403–1424. https://doi.org/10.1007/s00382-020-05400-5

Clark AJ (2017) Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. Weather Forecast 32(4):1569–1583

Colombo AF, Etkin D, Karney BW (1999) Climate variability and the frequency of extreme temperature events for nine sites across Canada: implications for power usage. J Clim 12(8):2490–2502

Coppola E, Raffaele F, Giorgi F et al (2021) Climate hazard indices projections based on CORDEX-CORE, CMIP5 and CMIP6 ensemble. Clim Dyn. https://doi.org/10.1007/s00382-021-05640-z

Crawford J, Venkataraman K, Booth J (2019) Developing climate model ensembles: a comparative case study. J Hydrol 568:160–173. https://doi.org/10.1016/j.jhydrol.2018.10.054

Cuadrat JM, Serrano-Notivoli R, Tejedor E et al (2020) Climpy: Climate of the pyrenees (1.0) [data set]. Zenodo https://doi.org/10.5281/zenodo.3611127

Dey A, Sahoo DP, Kumar R et al (2022) A multimodel ensemble machine learning approach for CMIP6 climate model projections in an Indian River basin. Int J Climatol. https://doi.org/10.1002/joc.7813

Dietterich T (1995) Overfitting and undercomputing in machine learning. ACM Comput Surv (CSUR) 27(3):326–327

Dobor L, Hlásny T (2019) Choice of reference climate conditions matters in impact studies: case of bias-corrected CORDEX data set. Int J Climatol 39(4):2022–2040. https://doi.org/10.1002/joc.5930

Escriva-Bou A, Pulido-Velazquez M, Pulido-Velazquez D (2017) Economic value of climate change adaptation strategies for water management in spain's jucar basin. J Water Res Plan ASCE 2:143

Estrela T (1992) Modelos matemáticos para la evaluación de recursos hídricos. Centro de Estudios Hidrográficos y Expón de Obras Públicas CEDEX 2:55

Evin G, Somot S, Hingray B (2021) Balanced estimate and uncertainty assessment of European climate change using the large EURO-CORDEX regional climate model ensemble. Earth Syst Dyn 12(4):1543–1569. https://doi.org/10.5194/esd-12-1543-2021

Fatichi S, Vivoni ER, Ogden FL et al (2016) An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J Hydrol 537:45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

García Ruiz J, Beguería S, López-Moreno J et al (2001) Los recursos hídricos superficiales del pirineo aragonés y su evolución reciente. Geoforma Ed 192

García-Barrón L, Camarillo JM, Morales J et al (2015) Temporal analysis (1940–2010) of rainfall aggressiveness in the Iberian Peninsula basins. J Hydrol 525:747–759. https://doi.org/10.1016/j.jhydrol.2015.04.036

Giorgi F, Jones C, Asrar G (2009) Addressing climate information needs at the regional level: the CORDEX framework. Organ WMO Bull 58:175–183

Gusain A, Ghosh S, Karmakar S (2019) Added value of cmip6 over cmip5 models in simulating indian summer monsoon rainfall. Atmos Res 232:104680. https://doi.org/10.1016/j.atmosres.2019.104680

Herger N, Abramowitz G, Knutti R et al (2018) Selecting a climate model subset to optimise key ensemble properties. Earth Syst Dyn 9(1):135–151. https://doi.org/10.5194/esd-9-135-2018

Herrera S, Soares PM, Cardoso RM et al (2020) Evaluation of the EURO-CORDEX Regional Climate Models Over the Iberian

Peninsula: Observational Uncertainty Analysis. J Geophys Res: Atmos 125(12):1–16. https://doi.org/10.1029/2020JD032880

IPCC (2022) Climate change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009325844

Jacob D, Petersen J, Eggert B et al (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg Environ Change 14(2):563–578. https://doi.org/10.1007/s10113-013-0499-2

Jacob D, Teichmann C, Sobolowski S et al (2020) Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community. Reg Environ Change. https://doi.org/10.1007/s10113-020-01606-9

Jimeno-Sáez P, Senent-Aparicio J, Pérez-Sánchez J et al (2018) A comparison of SWAT and ANN models for daily runoff simulation in different climatic zones of peninsular Spain. Water (Switzerland). https://doi.org/10.3390/w10020192

Jimeno-Sáez P, Pulido-Velazquez D, Collados-Lara AJ et al (2020) A preliminary assessment of the undercatching and the precipitation pattern in an alpine basin. Water. https://doi.org/10.3390/w12041061

Jódar J, Cabrera JA, Martos-Rosillo S et al (2017) Groundwater discharge in high-mountain watersheds: a valuable resource for downstream semi-arid zones. the case of the Bérchules River in Sierra Nevada (Southern Spain). Sci Total Environ 593–594:760–772. https://doi.org/10.1016/j.scitotenv.2017.03.190

Jose DM, Vincent AM, Dwarakish GS (2022) Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques. Sci Rep 12(1):1–25. https://doi.org/10.1038/s41598-022-08786-w

Keller L, Zischg AP, Mosimann M et al (2019) Large ensemble flood loss modelling and uncertainty assessment for future climate conditions for a Swiss pre-alpine catchment. Sci Total Environ 693:133400. https://doi.org/10.1016/j.scitotenv.2019.07.206

Knutti R, Allen MR, Friedlingstein P et al (2008) A review of uncertainties in global temperature projections over the twenty-first century. J Clim 21(11):2651–2663. https://doi.org/10.1175/2007JCLI2119.1

Kotlarski S, Keuler K, Christensen OB et al (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7(4):1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Krishnamurti TN, Kishtawal CM, LaRow TE et al (1999) Improved weather and seasonal climate forecasts from multimodel superensemble. Science 285(5433):1548–1550. https://doi.org/10.1126/science.285.5433.1548

Krishnamurti TN, Kishtawal CM, Zhang Z et al (2000) Multimodel ensemble forecasts for weather and seasonal climate. J Clim 13(23):4196–4216. https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2

Lambert SJ, Boer GJ (2001) CMIP1 evaluation and intercomparison of coupled climate models. Clim Dyn 17(2–3):83–106. https://doi.org/10.1007/pl00013736

Lemus-Canovas M, Lopez-Bustins JA (2021) Assessing internal changes in the future structure of dry-hot compound events: the case of the Pyrenees. Nat Hazard 21(6):1721–1738. https://doi.org/10.5194/nhess-21-1721-2021

Lemus-Canovas M, Lopez-Bustins JA, Trapero L et al (2019) Combining circulation weather types and daily precipitation modelling to derive climatic precipitation regions in the Pyrenees. Atmos Res 220(January):181–193. https://doi.org/10.1016/j.atmosres.2019.01.018

Li T, Jiang Z, Le Treut H et al (2021) Machine learning to optimize climate projection over China with multi-model ensemble simulations. Environ Res Lett. https://doi.org/10.1088/1748-9326/ac1d0c

Liang XZ, Kunkel KE, Meehl GA et al (2008) Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections. Geophys Res Lett 35(8):1–5. https://doi.org/10.1029/2007GL032849

López-Moreno JI, Vicente-Serrano SM, Zabalza J et al (2014) Respuesta hidrológica del pirineo central al cambio ambiental proyectado para el siglo xxi. Pirineos 169:160. https://doi.org/10.3989/Pirineos.2014.169004

Lutz AF, ter Maat HW, Biemans H et al (2016) Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. Int J Climatol 36(12):3988–4005. https://doi.org/10.1002/joc.4608

Marcos-Garcia P, Lopez-Nicolas A, Pulido-Velazquez M (2017) Combined use of relative drought indices to analyze climate change impact on meteorological and hydrological droughts in a Mediterranean basin. J Hydrol 554:292–305. https://doi.org/10.1016/j.jhydrol.2017.09.028

McSweeney CF, Jones RG, Lee RW et al (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. Clim Dyn 44(11–12):3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Meehl GA, Zwiers F, Evans J et al (2000) Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. Bull Am Meteor Soc 81(3):427–436. https://doi.org/10.1175/1520-0477(2000)081<0427:TIEWAC>2.3.CO;2

Morales-García J, Bueno-Crespo A, Terroso-Sáenz F et al (2023) Evaluation of synthetic data generation for intelligent climate control in greenhouses. Appl Intell 53(21):24765–24781. https://doi.org/10.1007/s10489-023-04783-2

Moriasi D, Arnold J, Van Liew M et al (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 10(13031/2013):23153

Murillo J, Navarro J (2011) Aplicación del modelo de témez a la determinación de la aportación superficial y subterránea del sistema hidrológico cornisa-vega de granada para su implementación en un modelo de uso conjunto. Bol Geol Minero 122:363–388

Oh SG, Suh MS (2017) Comparison of projection skills of deterministic ensemble methods using pseudo-simulation data generated from multivariate Gaussian distribution. Theoret Appl Climatol 129(1–2):243–262. https://doi.org/10.1007/s00704-016-1782-1

OPCC-CTP (2018) Climate change in the Pyrenees: Impacts, vulnerabilities and adaptation Bases of knowledge for the future climate change adaptation strategy in the Pyrenees. OPCC-CTP

Paluš M, Novotná D, Tichavský P (2005) Shifts of seasons at the European mid-latitudes: Natural fluctuations correlated with the North Atlantic Oscillation. Geophys Res Lett 32(12):1–4. https://doi.org/10.1029/2005GL022838

Pavan V, Doblas-Reyes FJ (2000) Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. Clim Dyn 16(8):611–625. https://doi.org/10.1007/s003820000063

Pérez-Sánchez J, Senent-Aparicio J, Segura-Méndez F et al (2019) Evaluating hydrological models for deriving water resources in peninsular Spain. Sustainability (Switzerland) 11(10):1–36. https://doi.org/10.3390/su11102872

Perkins SE, Pitman AJ, Holbrook NJ et al (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20(17):4356–4376. https://doi.org/10.1175/JCLI4253.1

Piani C, Weedon GP, Best M et al (2010) Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J Hydrol 395(3–4):199–215. https://doi.org/10.1016/j.jhydrol.2010.10.024

Reder A, Raffa M, Montesarchio M et al (2020) Performance evaluation of regional climate model simulations at different spatial and temporal scales over the complex orography area of the Alpine region. Nat Hazards 102(1):151–177. https://doi.org/10.1007/s11069-020-03916-x

Ruane A, McDermid S (2017) Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment. Earth Persp. https://doi.org/10.1186/s40322-017-0036-4

Sa'adi Z, Shahid S, Chung ES et al (2017) Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models. Atmos Res 197:446–460. https://doi.org/10.1016/j.atmosres.2017.08.002

Sachindra DA, Ahmed K, Rashid MM et al (2018) Statistical downscaling of precipitation using machine learning techniques. Atmos Res 212:240–258. https://doi.org/10.1016/j.atmosres.2018.05.022

Salman SA, Shahid S, Ismail T et al (2018) Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. Atmos Res 213:509–522. https://doi.org/10.1016/j.atmosres.2018.07.008

Sand M, Skeie RB, Sandstad M et al (2023) A multi-model assessment of the global warming potential of hydrogen. Commun Earth Environ 4(1):203

Sanderson BM, Knutti R, Caldwell P (2015) Addressing interdependency in a multimodel ensemble by interpolation of model properties. J Clim 28(13):5150–5170. https://doi.org/10.1175/JCLI-D-14-00361.1

Semenov MA, Stratonovitch P (2010) Use of multi-model ensembles from global climate models for assessment of climate change impacts. Clim Res 41(1):1–14

Senent-Aparicio J, López-Ballesteros A, Pérez-Sánchez J et al (2018) Using multiple monthly water balance models to evaluate gridded precipitation products over peninsular spain. Remote Sens. https://doi.org/10.3390/rs10060922

Serrano-Notivoli R, Beguería S, Saz MÁ et al (2017) Spread: a high-resolution daily gridded precipitation dataset for spain-an extreme events frequency and intensity overview. Earth Syst Sci Data 9(2), 721–738. https://doi.org/10.20350/digitalCSIC/7393

Smiatek G, Kunstmann H, Senatore A (2016) EURO-CORDEX regional climate model analysis for the Greater Alpine Region: Performance and expected future change. J Geophys Res 121(13):7710–7728. https://doi.org/10.1002/2015JD024727

Srivastava A, Grotjahn R, Ullrich P (2020) Evaluation of historical cmip6 model simulations of extreme precipitation over contiguous us regions. Weather Clim Extrem 29:100268. https://doi.org/10.1016/j.wace.2020.100268

Stahl K, Hisdal H, Hannaford J et al (2010) Streamflow trends in Europe: evidence from a dataset of near-natural catchments. Hydrol Earth Syst Sci 14(12):2367–2382. https://doi.org/10.5194/hess-14-2367-2010

Taylor KE (2001) In a single diagram. J Geophys Res 106:7183–7192

Témez J (1977) Modelo Matemático de Trasformación "Precipitación-Escorrentía". Asociación de Investigación Industrial Eléctrica (ASINEL) 2

Teng J, Potter NJ, Chiew FH et al (2015) How does bias correction of regional climate model precipitation affect modelled runoff?

Hydrol Earth Syst Sci 19(2):711–728. https://doi.org/10.5194/hess-19-711-2015

Thornthwaite CW (1948) An approach toward a rational classification of climate. Geogr Rev 38(1):55–94

Torma C, Giorgi F, Coppola E (2015) Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps. J Geophys Res: Atmosph 175(4449):238. https://doi.org/10.1038/175238c0

Vautard R, Kadygrov N, Iles C et al (2021) Evaluation of the large EURO-CORDEX regional climate model ensemble. J Geophys Res: Atmos 126(17):1–28. https://doi.org/10.1029/2019JD032344

Venkataraman K, Tummuri S, Medina A et al (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. https://doi.org/10.1016/j.jhydrol.2016.01.001

Wang B, Zheng L, Liu DL et al (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(13):4891–4902. https://doi.org/10.1002/joc.5705

Weisberg S (2005) Applied linear regression, vol 528. John Wiley & Sons

White RH, Toumi R (2013) The limitations of bias correcting regional climate model inputs. Geophys Res Lett 40(12):2907–2912. https://doi.org/10.1002/grl.50612

Willmott CJ (1981) On the validation of models. Phys Geogr 2(2):184–194. https://doi.org/10.1080/02723646.1981.10642213

Xu R, Chen N, Chen Y et al (2020) Downscaling and projection of multi-CMIP5 precipitation using machine learning methods in the upper Han River Basin. Adv Meteorol. https://doi.org/10.1155/2020/8680436

Ying X (2019) An overview of overfitting and its solutions. In: Journal of physics: Conference series, IOP Publishing, p 022022

Zabaleta A, Meaurio M, Uriarte J et al (2017) Hydric vulnerability: recent hydrologic trends in the bay of biscay. 2nd International colloquium on Climate Change in mountain areas PYRADAPT 2:72–76. https://doi.org/10.5194/hess-14-2367-2010, URL https://drive.google.com/file/d/1vBJ2pKzIvF-5X7E4Ij2gCxSekqkoiz-K/view

Zhu X, Zhang A, Wu P et al (2019) Uncertainty impacts of climate change and downscaling methods on future runoff projections in the Biliu River basin. Water (Switzerland) 11(10):1–17. https://doi.org/10.3390/w11102130

Zhu H, Jiang Z, Li L et al (2023) Intercomparison of multi-model ensemble-processing strategies within a consistent framework for climate projection in china. Sci China Earth Sci 66(9):2125–2141

Zhuang XW, Li YP, Huang GH et al (2016) Assessment of climate change impacts on watershed in cold-arid region: an integrated multi-GCM-based stochastic weather generator and stepwise cluster analysis method. Clim Dyn 47(1–2):191–209. https://doi.org/10.1007/s00382-015-2831-7