

Intercomparison of a High-Resolution Regional Climate Model Ensemble for Catchment-Scale Water Cycle Processes under Human Influence

Jane Leonor Roque¹ | Francis Da Silva Lopes² | Julian Alberto Giles¹ | Benjamin D. Gutknecht² | Bernd Schalge³ | Yikui Zhang⁴ | Marco Ferro⁵ | Petra Friederichs¹ | Klaus Goergen⁴ | Stefan Poll⁶ | Arianna Valmassoi³

¹Institute of Geosciences, Meteorology Section, University of Bonn, Germany

²Institute of Geodesy and Geoinformation, University of Bonn, Germany

³Deutscher Wetterdienst, Offenbach am Main, Germany

⁴Institute for Bio- and Geosciences (Agrosphere, IBG-3), Forschungszentrum Jülich, Jülich, Germany

⁵Institute for Food and Resource Economics, University of Bonn, Germany

⁶CASA Simulation and Data Laboratory Terrestrial Systems, Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Jülich, Germany

Correspondence

Jane Leonor Roque, Meteorology Section, Institute of Geosciences, University of Bonn, Bonn, North Rhine-Westphalia, 53121, Germany
Email: jroquema@uni-bonn.de

Funding information

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1502/1–2022 – Projektnummer: 450058266.

Understanding regional hydroclimatic variability and its drivers is essential for anticipating the impacts of climate change on water resources and sustainability. Yet, considerable uncertainty remains in the simulation of the coupled land–atmosphere water and energy cycles, largely due to structural model limitations, simplified process representations, and insufficient spatial resolution. Within the framework of the Collaborative Research Center 1502 – DETECT, this study presents a coordinated intercomparison of regional climate model simulations designed for water cycle process analysis over Europe. We analyze the performance of simulations using the ICON and TSMP1 model systems and covering the period 1990–2020, comparing against reference datasets (E-OBS, GPCC, and GLEAM). We focus on 2 m air temperature, precipitation and evapotranspiration over four representative basins – the Ebro, Po, Rhine, and Tisa – within the EURO-CORDEX domain.

Our analysis reveals systematic cold biases across all basins and seasons, with ICON generally outperforming TSMP1.

Precipitation biases exhibit substantial spread, particularly in summer, reflecting the persistent challenge of accurately simulating precipitation. ICON tends to underestimate evapotranspiration, while TSMP1 performs better some seasons. Sensitivity experiments further indicate that the inclusion of irrigation improves simulation performance in the Po basin, which is intensively irrigated, and that higher-resolution sea surface temperature forcing data improves the overall precipitation representation. This baseline evaluation provides a first assessment of DETECT's multi-model ensemble and highlights key structural differences influencing model skill across hydroclimatic regimes.

KEYWORDS

Terrestrial system modeling, ICON, EURO-CORDEX, model intercomparison, sea surface temperature, irrigation, water cycle

1 | INTRODUCTION

Anthropogenic changes of land use and land cover affect biophysical processes in the energy and water cycle (Pielke Sr. et al., 2011; Mahmood et al., 2014), from the local to regional scale and remotely. Accurate climate simulations are needed to provide decision makers with reliable information to develop effective climate change adaptation and mitigation strategies (Gutowski et al., 2020). In this context, regional drying and wetting trends emerged in recent years and pose a challenge in terms of process representation and understanding. While the presence of large-scale drying and wetting patterns has been described extensively (Masson-Delmotte et al., 2021), both satellite observations and in-situ hydroclimatic trends cannot be explained by simple paradigms such as the “dry-gets-drier, wet-gets-wetter” or “global aridification” hypotheses, due to their complex time-varying spatial heterogeneous footprint (Zaitchik et al., 2023). Further comparative studies of global and regional climate simulations have revealed that even the sign of the terrestrial water storage trends differs between models and observations (Greve et al., 2014; Gudmundsson et al., 2021; Jensen et al., 2024; Hirschi et al., 2025), highlighting the need for an improved representation of the coupled water-energy cycles representation across different Earth System compartments.

The underlying modeling uncertainty can be attributed to a number of factors, e.g., model structural deficits, unrepresented processes -such as groundwater dynamics-, or to the low resolution of the current modeling systems. In fact, global models typically operate at grid resolutions of 50-100 km while many regional climate models run on resolutions of around 10-25 km, relying on the parameterization of many water cycle related processes. Deep and shallow convection at these resolutions must be parametrized and many key land-surface and hydrologic processes are only represented in a simplified manner (Schär et al., 2020; Lucas-Picher et al., 2021). In recent decades, the development and evaluation of regional climate models have revealed substantial improvements in the simulation of regional climate patterns, but also persistent biases and spread among models, particularly for precipitation and land-surface fluxes (Christensen and Christensen, 2007; Jacob et al., 2020; Prein et al., 2016).

The models differ in their physical parameterizations, soil and vegetation schemes, treatment of lateral boundary

conditions, etc., which strongly affect their representation of surface fluxes and the terrestrial water balance (Kotlarski et al., 2014; Poshyvailo-Strube et al., 2024; Silwimba et al., 2025). Furthermore, the general lack of explicit coupling to subsurface or groundwater components limits their ability to capture the full spectrum of land-atmosphere interactions. In addition, most model systems do not account for lateral subsurface flow, i.e., 3D hydrodynamics, but feature only 1D gravitational subsurface flow with open lower boundary conditions. Therefore, subsurface-land-atmosphere feedbacks do not account for groundwater dynamics, which has an impact on land-atmosphere coupling and on soil moisture-temperature or -precipitation feedback loops (Keune et al., 2016). Also, human interventions, such as irrigation or groundwater abstraction, which have a substantial impact on the terrestrial water cycle (Siebert et al., 2010; Wada et al., 2010), strongly affect the surface energy partitioning and hydrometeorology but are either over-simplified or absent in models (Hirsch et al., 2018; Seneviratne et al., 2018; Valmassoi and Keller, 2022). Such limitations may lead models to produce contrasting water balance representations and underestimate, or even obscure, the anthropogenic impact on the hydroclimatic variability (Winkler et al., 2023; Casirati et al., 2025).

In recent years, partly due to the increase in computational resources, there is a tendency towards the use of convection-permitting regional climate models at grid resolutions below 4 km, which overcome some of the aforementioned limitations (Prein et al., 2015; Schär et al., 2020). At these resolutions, atmospheric models show improved representations of key processes (e.g., atmospheric convection, mesoscale circulations, small-scale weather events), leading to a more accurate representation of precipitation intensity, spatial distribution, and timing (e.g., Ban et al., 2021; Fosser et al., 2020). Barlage et al. (2021) further showed that the representation of terrestrial hydrologic processes is scale-dependent and that an enhanced treatment of groundwater-land interactions can influence hydrometeorology at kilometer scales in the WRF model. However, the representation of hydrologic processes in convection-permitting models often remains simplified and human interventions are usually not represented.

To address these limitations, modeling efforts are increasingly moving towards fully coupled regional Earth or climate system models. These models explicitly link the subsurface, land surface and atmosphere. The Terrestrial System Modelling Platform (TSMP) is such a model (Shrestha et al., 2014; Gasper et al., 2014). In its first version, TSMP1, the COSMO atmospheric model is coupled to the Community Land Model and the ParFlow integrated hydrologic model via the OASIS3-MCT coupler (Shrestha et al., 2014). TSMP1 thus enables a physically consistent representation of water and energy fluxes from the groundwater to the atmosphere, advancing the simulation of regional hydroclimate dynamics.

The Collaborative Research Center 1502 – DETECT – (Siegismund and Kusche, 2024), funded by the German Research Foundation, builds upon these recent advances in regional climate modeling. In light of the intensification of the hydrologic cycle under climate change—with major implications for water resources, sustainability, and water security (Huntington, 2006; Wada and Bierkens, 2014)—there are still significant gaps in our understanding of how feedbacks within the coupled water-energy system respond to and amplify changes in the system state. These uncertainties are reflected in the often-contradictory patterns of observed hydrologic change (Jensen et al., 2024). The central hypothesis driving the CRC 1502 DETECT is that, in addition to greenhouse gas forcing and natural variability, human water use as well as land-use and land-cover change have induced persistent modifications of the coupled land-atmosphere water and energy cycles, leading to multiple local and non-local hydroclimatic effects and contributing to observed regional- to continental-scale water storage trends (Ferro et al., 2025).

To answer the central hypothesis, coordinated regional simulation experiments with a state-of-the-art multi-physics multi-model ensemble over Europe were designed and performed within DETECT. Similar to studies such as Kotlarski et al. (2014), a model intercomparison and evaluation to observational reference data is a pivotal step to assess simulation performances, e.g., with respect to terrestrial water cycle processes and how the models represent key aspects of land-atmosphere interactions.

This study analyzes the regional climate model ensemble for water cycle and water resource research from the first funding phase of DETECT. We rely on the framework introduced by Seneviratne et al. (2018) to compare ten simulations from different models for a 30 year time span. We use 2-meter air temperature, precipitation, and evapotranspiration over land as key indicators of the energy and water cycle, and we assess their performance through an inter-model comparison and against external reference datasets. We focus on four European basins that are representative of the continent's different bioregions, i.e., where climatic conditions can vary considerably: the Rhine, the Tisa, the Ebro, and the Po river basins. Throughout our analysis, we address the following research questions to guide our model intercomparison: (1) What are the key similarities and differences in the representation of the water cycle across model systems? (2) What do these reveal about model behavior at the basin scale? (3) What are the capabilities of the simulations to represent the human impact? In addition, the study introduces some of the datasets produced during the first funding phase of DETECT.

2 | METHODOLOGY

2.1 | Simulation data

Ten simulations are available for comparison, all covering the EURO-CORDEX model domain (see Figure 1). Table 1 summarizes some key characteristics of the model simulations, and each experiment acronym refers to the model name, horizontal grid spacing and experiment type.

Seven of the ten experiments use the ICOSahedral Nonhydrostatic model (ICON v2.6.4, except v2.6.6-nwp3c for the ICON_3_irr/ctrl), developed by the German Weather Service and the Max Planck Institute for Meteorology (Zängl et al., 2015; Heinze et al., 2017; Hohenegger et al., 2023; Gassmann and Herzog, 2008) and operationally used for numerical weather prediction as well as climate simulations. The other three experiments use a fully integrated regional climate modelling framework, the TSMP1, designed to explicitly resolve land-atmosphere-subsurface interactions. The modelling system couples three component models via the OASIS3-MCT coupler (Valcke, 2013): (i) the Consortium for Small Scale MOdelling (COSMO) atmosphere model v5.01 (Baldauf et al., 2011; Doms et al., 2011), which is the predecessor of the ICON model; (ii) the Community Land Model (CLM v3.5) developed at the National Center for Atmospheric Research (Oleson et al., 2008); and (iii) the ParFlow v3.12.0 surface-subsurface flow integrated hydrologic model (Kuffour et al., 2020; Jones and Woodward, 2001; Kollet and Maxwell, 2006; Maxwell, 2013).

As can be seen from Table 1, most of the simulations are forced with the ERA5 reanalysis (Hersbach et al., 2020), a global reanalysis with around 31 km of horizontal resolution and 137 vertical levels, which is widely used as initial and boundary conditions for regional climate model studies. It is the recommended dataset to drive CORDEX evaluation simulations (Katragkou et al., 2024). The original resolution was remapped to 12 km resolution using routines from the model's runtime environment (Rockel and Geyer, 2022), and, if needed, using sea surface temperature (SST) as the prescribed lower boundary condition over the EURO-CORDEX domain.

The two convection-permitting ICON experiments (ICON_3_irr and ctrl) use instead the newly produced ICON-DREAM reanalysis (Valmassoi et al., 2024) as boundary conditions, since it provides a higher spatial resolution (approximately 13 km, compared to 31 km of ERA5), which is preferable for the target resolution. To ensure consistency, we used the same ICON-nwp version as ICON-DREAM reanalysis. The 3 km ICON reanalysis (ICON_3_rea) uses the ERA5 and its 10 ensemble members, which are at a coarser resolution (62 km), as ICON-DREAM was not yet available at the start of the DETECT simulation experiment.

The first seven simulations, which are convection-parameterized, have a 12 km horizontal grid spacing and 50/60 vertical levels, TSMP1 and ICON respectively (for more information: Zhang et al., 2024; da Silva Lopes et al., 2025),

TABLE 1 Model characteristics. Please note: If no information is provided in the SST column, then the SST is prescribed jointly with the boundary condition data and stems form the same dataset; likewise for the irrigation methods column, no information means the model is run without irrigation.

Acronym	Model	Boundary conditions	Full Period	Horizontal resolution	No. of vertical levels	Radiation scheme	Land surface scheme	Land use external data	Irrigation method	SST
TSMP1_12_ctrl	TSMP1	ERA5	1979–2022	12 km	50	RRTM	CLM 3.5	GLC-2000	-	-
TSMP1_12_irri_w	TSMP1	ERA5	1984–2022	12 km	50	RRTM	CLM 3.5	GLC-2000	HWU1 ¹	-
TSMP1_12_irri_s	TSMP1	ERA5	1990–2022	12 km	50	RRTM	CLM 3.5	GLC-2000	HWU2 ²	-
ICON_12_ctrl	ICON	ERA5	2002–2020	12 km	60	ecRad	TERRA	GlobCover2009	-	-
ICON_12_sst_era5	ICON	ERA5	2002–2011	12 km	60	ecRad	TERRA	GlobCover2009	-	ERA5 ^{SST3}
ICON_12_sst_mur	ICON	ERA5	2002–2011	12 km	60	ecRad	TERRA	GlobCover2009	-	MURSST ⁴
ICON_12_sst_ostia	ICON	ERA5	2002–2011	12 km	60	ecRad	TERRA	GlobCover2009	-	OSTIA ⁵
ICON_3_ctrl	ICON	ICON-DREAM	2010–2022	3 km	75	ecRad	TERRA	GlobCover2009	-	-
ICON_3_irr	ICON	ICON-DREAM	2010–2022	3 km	75	ecRad	TERRA	GlobCover2009	RAW ⁶	-
ICON_3_rea	ICON	ERA5	2018–2021	3 km	75	ecRad	TERRA	GlobCover2009	-	-

¹ Human water use 1 (Wada et al., 2016)

² Human water use 2 (Siebert and Döll, 2010)

³ ERA5^{SST} (Yang et al., 2021)

⁴ MURSST (Chin et al., 2017)

⁵ OSTIA (Donlon et al., 2012)

⁶ Readily available water

extending up to 22 km and 23.5 km, respectively. The last three convection-permitting simulations have a horizontal resolution of 3 km with 75 vertical levels extending up to 22 km (ICON_3_irr/ctrl) and 30 km (ICON_3_rea).

All experiments use the single-moment cloud microphysics and precipitation scheme (Doms et al., 2011; Seifert, 2008), and the prognostic turbulent kinetic energy scheme for vertical mixing (Raschendorfer, 2001). The Tiedtke bulk mass flux scheme (Tiedtke, 1989) is used for shallow convection for all experiments, but disabled for deep and mid-level convection (Lucia et al., 2022) in the 3 km simulations.

TSMP1, based on COSMO and with coupled CLM and ParFlow constitutes a different model system than ICON. Nevertheless, the atmospheric model COSMO and ICON share some common parametrizations and physics packages and diagnostics, e.g., for cloud cover. Notable difference in parameterizations between the TSMP1 and ICON-experiments are, e.g., the radiation scheme and the land surface model components. TSMP1 uses the RTTM radiation parameterization available to the COSMO predecessor, all ICON experiments use the more advanced ecRad radiation scheme, which brings improvements in process representation (Hogan and Bozzo, 2018). While the ICON-experiments use the TERRA land surface model (Schrodin and Heise, 2001; Schulz et al., 2016), which was also default in the COSMO model, TSMP1 replaces it with the CLM3.5.

Now each set of experiments is presented to further contextualize the respective experiment designs, depending on each simulation's objectives.

2.1.1 | Terrestrial System Modeling Platform (TSMP) simulations

The Terrestrial System Modelling Platform (TSMP) is a fully integrated regional climate modelling framework designed to explicitly resolve subsurface-land-atmosphere interactions. In contrast to regional climate models that employ simplified hydrologic schemes, TSMP incorporates a full three-dimensional representation of subsurface hydrodynamics, including groundwater flow. This capability enables a more realistic simulation of soil moisture states, energy fluxes, and associated land-atmosphere feedbacks, thereby improving the representation of heatwave intensity and reducing temperature biases, as demonstrated for both the 2003 European heatwave and multi-decadal climatologies (Keune et al., 2018; Furusho-Percot et al., 2022; Poshyvailo-Strube et al., 2024). This bedrock-to-atmosphere coupling allows TSMP to capture surface-subsurface exchange processes and groundwater dynamics with high fidelity, leading to improved soil moisture profiles and evapotranspiration behaviour (Zipper et al., 2019; Naz et al., 2022). Previous work by (Hartick et al., 2022) has shown that TSMP's explicit groundwater representation enhances the persistence of drought events through groundwater memory effects.

In this study, we conducted three TSMP1 experiments covering the period 1990–2022: one control simulation without irrigation (CTL) and two with irrigation (HWU-W and HWU-S). The irrigation forcing is based on two monthly datasets spanning 1984–2022: HWU-W from (Wada et al., 2016) and HWU-S from (Siebert and Döll, 2010). The irrigation forcing is implemented in ParFlow through source-sink terms applied to the relevant model layers. Keune et al. (2018) gives an overview on the use of TSMP1 including human water use.

2.1.2 | Sea surface temperature experiments

The SST product choice as a prescribed lower boundary in limited-area climate simulations remains a critical issue, especially when the model domains partly cover the ocean (cf. Katragkou et al., 2015; Cassola et al., 2016; Liu et al., 2024). Different SST products, ranging from global model outputs (Donlon et al., 2002) to high-resolution satellite-based datasets, can lead to substantial variability in the lower boundary forcing. This is particularly important for water cycle studies, where SST patterns influence the low-level atmospheric dynamics (e.g., Zhou et al., 2019), ocean latent

heat flux, and thus the moisture transport to land (Donlon et al., 2002), with an impact on precipitation (Goergen and Kollet, 2021). The DETECT SST sensitivity experiments aim to assess the sensitivity of the simulated water budget components to SST forcing over the EURO-CORDEX (Jacob et al., 2020) domain ($\sim 27^{\circ}\text{N}$ – 72°N , $\sim 22^{\circ}\text{W}$ – 45°E). The setup, parametrizations and experimental design for ICON 12 km SST experiments are described in detail by da Silva Lopes et al. (2025).

The SST experiments cover a time span from June 2002 to May 2011, each using the same ICON-CLM configuration, and they differ in the imposed SST dataset. Changing only the lower boundary condition (SST) while retaining the lateral boundary forcing can introduce inconsistencies in the regional energy and moisture budgets. Nonetheless, da Silva Lopes et al. (2025) showed that these effects were negligible, with basin-scale water budget offsets (see their Table 2), indicating no significant imbalance across experiments.

The four SST products tested in our regional climate model simulations are: (1) hourly ERA5 skin temperature (ERA5^{SKT}), (2) daily ERA5 SST (ERA5^{SST} , Yang et al., 2021), (3) daily Multi-scale Ultra-high Resolution SST (MURSST, Chin et al., 2017), and (4) daily gap-free maps of the Operational Sea Surface Temperature and Ice Analysis foundation SST (OSTIA, Donlon et al., 2012). ERA5^{SKT} reflects the radiative equilibrium temperature of an infinitesimally thin surface layer and includes diurnal variability (Donlon et al., 2002; Minnett et al., 2019; ECMWF, 2020), based on model-data synthesis (Zeng et al., 2005; Luo and Minnett, 2020). ERA5^{SST} , MURSST and OSTIA, by contrast, provide foundation SST (Minnett et al., 2019), a bulk near-surface temperature unaffected by diurnal warming, derived from observation-based datasets (Titchner and Rayner, 2014; Donlon et al., 2012; Chin et al., 2017; Hirahara et al., 2016; Hersbach et al., 2020). MURSST's higher resolution (0.01°) enables finer representation of mesoscale ocean features (e.g., Lebeaupin et al., 2015; Seo et al., 2023). However, its availability only since June 2002 constrained the temporal extent of the simulations. For preprocessing, we use Climate Data Operators (Schulzweida, 2023), interpolating SSTs to the ICON-CLM 0.11° grid. Skin temperature over land and ice, as well as the land-sea mask, remain unchanged. ERA5^{SKT} serves as our reference, consistent with its use in atmospheric data assimilation (Akella et al., 2017; Zeng et al., 2005), weather prediction (Scanlon et al., 2024), and air-sea interaction studies (Fairall et al., 1996).

2.1.3 | Irrigation experiment with ICON-nwp

The main purpose of irrigation is to fulfill crop water requirements, leading to higher soil moisture values. This increase in soil moisture initiates a range of feedbacks and interactions with various components and processes of the Earth system (McDermid et al., 2023). However, the representation of irrigation in weather and climate models remains a significant limitation. Omitting irrigation introduces significant biases even in reanalysis datasets (Tuinenburg and de Vries, 2017), as they miss displaying heterogeneity at sub-grid scale (Mangan et al., 2023). The cooling effect is one of the most recognized irrigation effects, yet the specific extent of this impact remains uncertain locally and globally (Valmassoi and Keller, 2022). As shown by Keune et al. (2018) irrigation in combination with groundwater abstraction may not only have local effects but can also lead to remote impact due to atmospheric moisture transport.

In this study, we performed one control and one irrigation experiment with the model ICON-nwp in Limited Area Mode (LAM) to quantify the long-term impact of irrigation on surface and atmospheric variables.

The control experiment covers the time span 2010-01-01/2022-12-31. The irrigation experiment has a parametrization located in the interface between ICON and the land surface scheme. The amount of irrigation is fixed, equal to the readily available water (RAW) (Allen et al., 1998). The irrigation experiment started from a restart on 2010-04-11 00UTC and covers the same period as the control. In each year, the irrigation event starts on May 1, at 5 UTC and runs for 5 hours. It repeats every 12 days until Aug 17, with a total of 10 irrigation events. We considered 2010 the spin-up time for both experiments.

2.1.4 | Sparse Input Reanalysis (2018–2021)

For the sparse-input reanalysis, the ICON model is run in conjunction with the KENDA (Schraff et al., 2016) data assimilation framework. Unlike all other experiments, the reanalysis is not a free forecast but runs in an assimilation loop where every three hours observations are assimilated. The observation types assimilated here include synoptic stations, buoys, aircraft data, pilot observations and radiosondes, but critically do not include any radar or satellite data, except atmospheric motion vectors. This setup is therefore considered a sparse-input reanalysis, which is a compromise between computational demand and accuracy of the analysis.

The data assimilation system KENDA uses the localized ensemble transform Kalman filter (LETKF) and therefore needs an ensemble in addition to the main deterministic run. In operational systems (which mostly use 3/4DVar or similar variational methods), the ensemble is usually run at a lower resolution than the deterministic run to reduce the computational costs while still adequately sampling uncertainty. In our case, this approach would have meant to run the ensemble in lower resolution, non-convection-permitting mode, leading to different precipitation and cloud statistics. Consequently, it was decided to keep the resolution at 3 km for the ensemble as well.

As mentioned before, ERA5 was chosen as boundary data. Not only is it a reanalysis itself, it also features 10 ensemble members which is ideal, since boundary data is also needed for the ensemble. However, this comes with two constraints. First, this limits the ensemble to just 10 members, which is rather low for such data assimilation methods. Several tests have been conducted and results have shown that indeed adding ensemble members does improve the quality of the analysis further, but we have also noted that there are no adverse effects to using just 10 members, the system works in a stable and predictable way. Secondly, ERA5 data is only available at a rather coarse resolution so the transition to the 3 km model domain is rather extreme. This was tested and only a very minor improvement can be achieved by using operational ICON fields (~13 km) instead.

With January 2018 being considered a spin-up phase, currently the reanalysis covers the time February 2018 to December 2021, with an extension until the end of 2023 underway. Unlike other simulations, evapotranspiration is not directly available as an output variable but is instead derived from surface latent heat flux.

2.2 | Reference datasets

For comparison of simulated 2 m air temperature, precipitation, and evapotranspiration against external references, we selected several gridded observational and calibrated reanalysis products (Table 2) as follows.

(a) E-OBS

We used daily mean air temperature data from the E-OBS gridded dataset, developed by the European Climate Assessment & Dataset (ECA&D) and the Copernicus Climate Change Service (Haylock et al., 2008; Cornes et al., 2018). E-OBS provides daily meteorological variables across Europe based on quality-controlled station observations, interpolated onto regular latitude–longitude grids at 0.1° and 0.25° resolution. The dataset spans from 1950 to 2024 (v31.0e) and covers the European land domain from 25°N–71.5°N and 25°W–45°E. E-OBS mean temperature usually describes the temperature measured at 2 meters above the surface.

E-OBS is based on a 20-member ensemble to account for interpolation uncertainty, with the ensemble mean typically used as a best estimate. The gridded fields are produced using conditional simulation methods that preserve spatial correlations while providing uncertainty information. While the dataset is widely used as a high-resolution reference for climate studies in Europe, its accuracy depends on the underlying station network, and uncertainties are larger in regions with sparse observational coverage. In this study, we used the ensemble-mean

TABLE 2 Reference data sets

Model	Version	Variable	Gridded native resolution
E-OBS	v31.0e	Monthly mean of daily mean air temperature	1/4°
GPCC	v2022	Monthly precipitation (full data monthly)	1/4°
GLEAM	v4.2a	Monthly Actual Evaporation	1/10°

daily-mean temperature at 0.25° and computed the monthly means.

(b) GPCC

We compare simulated land precipitation with monthly precipitation sums from the Global Precipitation Climatology Centre (GPCC) 'Full Data Monthly' product version 2022 (Schneider et al., 2022). Operated by the German Meteorological Service (DWD) under the auspices of the World Meteorological Organization (WMO), GPCC provides globally gridded monthly observations of land-surface precipitation. These are based on quality-controlled data from rain gauges and are optimized for spatial coverage. Station density varies with time and region, but it is particularly high across Europe and is generally of the order of 1–2 gauges per 0.25° cell.

(c) GLEAM

We utilized evaporation estimates from the GLEAM (Global Land Evaporation Amsterdam Model) dataset (gleam.eu). GLEAM provides daily evaporation (or evapotranspiration) and its component fluxes (transpiration, soil evaporation, interception loss, open-water evaporation, snow sublimation) over land surfaces, along with auxiliary variables such as potential evaporation, root-zone and surface soil moisture, evaporative stress, and sensible heat flux.

The latest version (GLEAM 4.2a, Miralles et al. (2025)) offers continuous global coverage from 1980 to 2023 at 0.1° spatial resolution, using a hybrid machine-learning and physical model approach. The model computes potential evaporation based on surface net radiation, temperature, wind speed, vapor pressure deficit, and vegetation properties, and then applies evaporative stress constraints (derived from remote sensing of soil moisture, vegetation optical depth, etc.) to partition into actual evaporation components. As with any modeled dataset, uncertainties arise from forcing errors, remote sensing uncertainties, and limitations in representing spatial heterogeneity, particularly in regions with sparse sensor coverage or complex terrain. Here we use GLEAM monthly Actual Evaporation averaged over the selected basins as a reference.

2.3 | Statistical validation

To compare the model performance of the simulations described above, we focused on three key variables: 2 m air temperature (T2m), precipitation (P) and evapotranspiration (ET). The corresponding observational datasets used for validation, as stated in the previous section, are E-OBS (T2m), GPCC (P) and GLEAM (ET). Initially, due to the long temporal coverage, we evaluate the TSMP1 simulations against observations based on the climatological mean of each variable at grid-point scale. Then, given our interest in the water cycle representation, we aggregate data from all simulations and observations at the watershed level. We chose four representative watersheds, two located in mid-Europe and two located in southern Europe: Rhine, Tisa (sub-watershed from the Danube), Ebro and Po (Figure 1). Therefore, at a watershed level, we work with timeseries of spatially aggregated monthly means (T2m) and monthly sums (P and ET) .

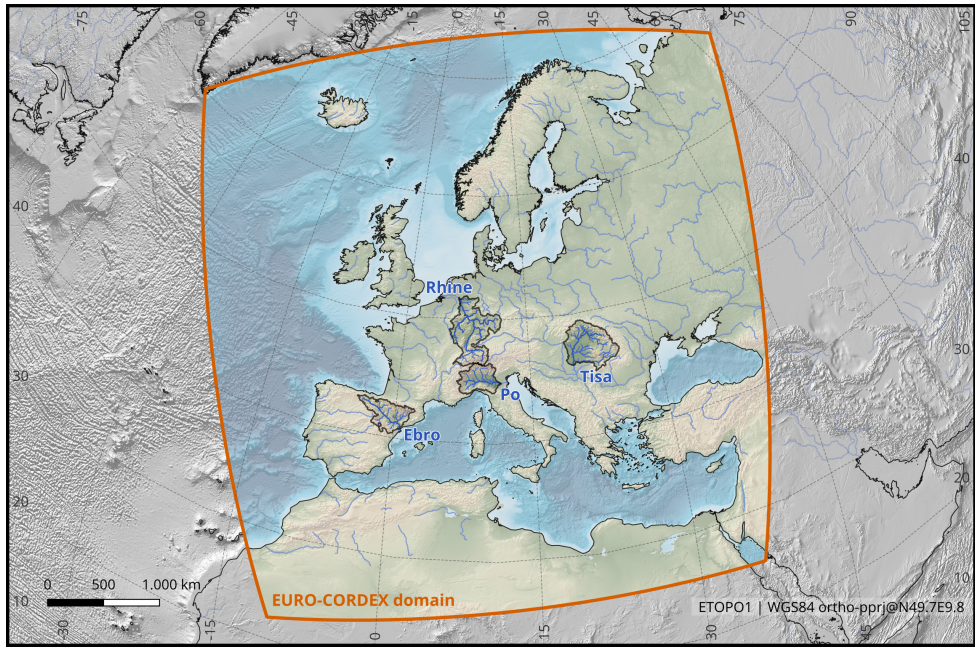


FIGURE 1 Overview map of our study area in the EURO-CORDEX domain. Highlighted are the four selected river catchments used in this study for area-averaged analysis: Rhine River (Rees station), Rio Ebro (Tortosa station), Fiume Po (Pontelagoscuro station), and Tisa River (Senta station, tributary to Danube River). The named stations are GRDC-registered locations of river gauges and thus define the upstream portion of the drainage basins, respectively (GRDC, 2023).

The validation consists of five statistical metrics: The bias, standard deviation (STD), root mean square error (RMSE), mean absolute error (MAE) and the Pearson correlation coefficient. We calculate these metrics for an inter-annual seasonal and yearly evaluation with data separated to two simulation periods. We opt to separate the data to account for overlapping simulation years across experiments. The first period corresponds to 2003–2010 (January 1, 2003 - December 31, 2010). The second period corresponds to 2011–2020 (January 1, 2011 - December 31, 2020).

Due to horizontal grid resolution differences between the simulations and the reference datasets, we remap the modeled data to the grids of the reference datasets for the comparisons on a grid-point scale. T2m is remapped using bilinear interpolation and P and ET are remapped using conservative methods to preserve the total amount of each field (Taylor, 2024). For this task we use the xESMF Python package (Zhuang et al., 2025).

3 | RESULTS AND DISCUSSION

3.1 | Climatological biases of TSMP1 CTRL

The TSMP1 CTRL climatological biases (i.e., differences of long-term means, simulation minus reference) of T2m, P and ET for the period 1990–2020 are shown in Figure 2. Regarding T2m, Figure 2a shows that TSMP1 CTRL has a negative T2m bias over all land areas. The bias is more negative (stronger) over areas with complex topography.

Figure 2b shows that a positive precipitation bias that covers almost all of Europe, with some scattered areas hav-

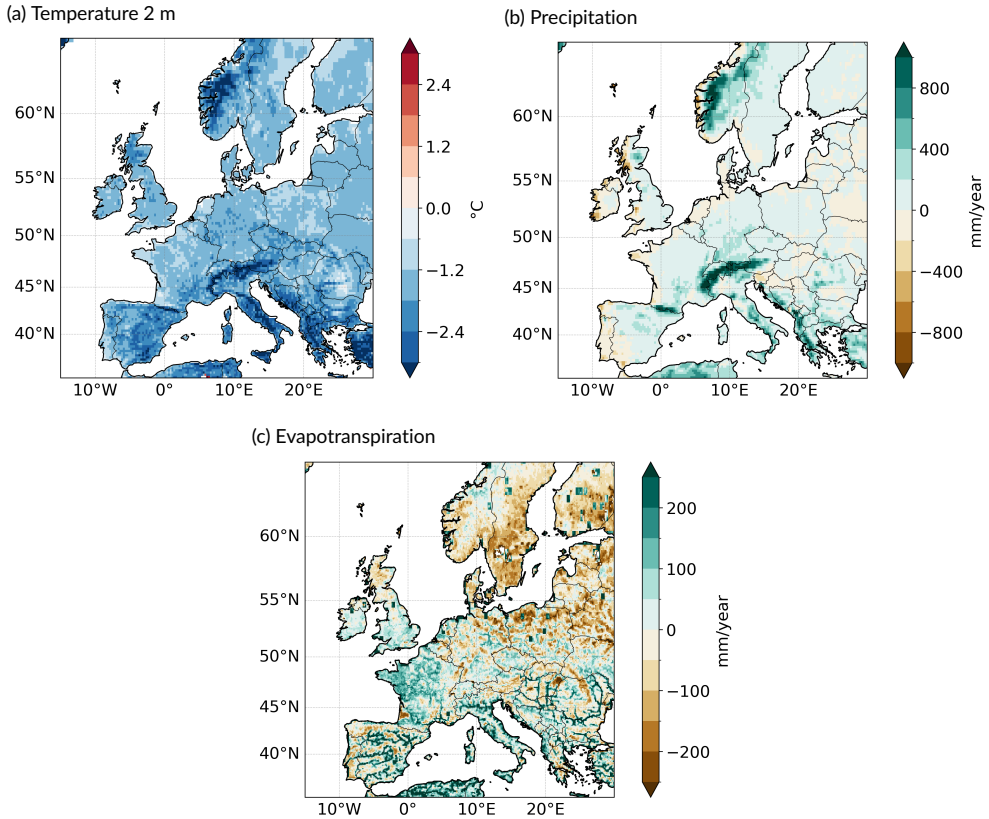


FIGURE 2 TSMP1 CTRL (TSMP1_12_ctrl) 1990-2020 mean biases of T2m, P and ET with respect to E-OBS, GPCC and GLEAM respectively.

ing a slight negative bias. Regions of high positive bias are directly associated with elevated and complex topography, like over the Alps, the Pyrenees, Norway and the Balkans. Finally, ET, Figure 2c shows positive biases mostly over Italy, the Balkans, France, Spain, Ireland and southern Great Britain. Meanwhile, negative biases are visible mostly over northern and eastern Europe. Interestingly, high positive biases are seen following many rivers, most noticeably in the Iberian peninsula, Italy and the Balkans. This effect is attributed to how the rivers are represented in ParFlow. ParFlow does not use a river routing scheme. Overland flow water, simulated with a 2D kinematic wave equation, is flowing to local depressions and then along river corridors, as convergence zones. At the same time, these convergence zones are characterized by shallow water tables, aquifer-streamflow interaction occurs. Due to ParFlow's 3D subsurface hydrodynamics, lateral subsurface flow along topographic gradients is also directed to the river valleys. In the DETECT setup Parflow uses a hydrofacies distribution where the alluvia in the riverbeds have high permeability and porosity values assigned to mimic river bed properties. In addition, ParFlow as an integrated model treats overland flow and surface hydrology and subsurface 3D hydrodynamics in a continuum approach, which means that saturation and infiltration excess overland flow can be simulated.

Two primary factors may contribute to the cold bias of T2m and the positive bias of ET: First, the ParFlow model in TSMP1_12_ctrl tends to simulate a relatively shallow groundwater table across most of the European continent,

except in parts of Eastern Europe. This may result from a revised spinup in combination with a revised hydrogeological parametrisation that leads to a higher water table depth in the initial conditions. The shallower water table likely enhances soil moisture availability near the surface, leading to higher root water uptake, capillary rise and eventually ET in regions such as France, the Mediterranean region, and Central Europe, thereby contributing to surface cooling and the cold bias. Second, the cold bias may also be linked to the SST dataset used as the lower boundary condition in the COSMO model. The dataset, appears to introduce cold anomalies not only over land but also over adjacent ocean areas, which may hint at a potentially large-scale forcing issue. Both factors are currently under investigation and require further validation.

3.2 | Model inter-comparison with Reference datasets

3.2.1 | Seasonal interannual variability (Period 2003–2010)

The seasonal T2m biases averaged per basin are presented as boxplots showing the interannual data spread across each simulation (Figure 3). The biases for this variable exhibit larger differences between the medians of TSMP1 and ICON experiments. Overall, all experiments systematically underestimate T2m across seasons, with the exception of the Tisa basin, where only ICON shows a warm bias in most seasons.

In northern hemispheric winter (DJF) and spring (MAM), ICON median biases in the Ebro, Po, and Rhine basins remain close to -0.5 K, whereas TSMP1 biases reach approximately -2 K and up to -3 K in DJF for the Po basin. The bias spread, similar to other variables, is larger in TSMP1 experiments. In the Tisa basin, ICON maintains a warm bias in most of the experiments, with a maximum of 0.36 K for ICON_12_sst_mur, and for MAM these are almost distributed around zero. Moreover, TSMP1 consistently shows a cold bias exceeding -1.5 K.

In northern hemispheric summer (JJA), ICON achieves its best performance, with median biases near zero in the Ebro, Po, and Rhine basins. In contrast to this, compared to the preceding seasons, biases in the Tisa basin increase for the ICON model and reduce for the TSMP1, being 0.85 K the highest mean bias for ICON_12_sst_mur and -1.97 K the lowest for TSMP1_12_irri_s. In all cases, TSMP1 maintains the cold bias in all basins.

In autumn (SON), both models show cold biases in the Ebro, Po, and Rhine basins, while ICON continues to display a warm bias in the Tisa basin. Overall, ICON biases are consistently closer to zero than those of TSMP1, indicating better performance in representing T2m.

In this regard, our findings are consistent with Pham et al. (2021), who reported smaller biases for ICON-CLM and COSMO-CLM in winter and autumn and larger biases in spring and summer. The persistent cold bias across all seasons in TSMP1, therefore, cannot be attributed solely to the performance of COSMO. As discussed in section 3.1, the systematic cold bias in TSMP1 is largely associated with the groundwater configuration in ParFlow and possible biases in the prescribed SST lower boundary conditions, rather than with the atmospheric component COSMO.

Compared to T2m results, seasonal precipitation biases averaged per basin also show differences between ICON and TSMP1 simulations, especially for the Ebro and Rhine river basins during all seasons (Figure 4).

In DJF, the model performance is similar in the Ebro and Rhine basins. Here, ICON experiments exhibit medians near zero and smaller fluctuations around this quantile. In contrast, all TSMP1 experiments overestimate precipitation, with medians around $\pm 20 \text{ mm month}^{-1}$ and a larger bias spread in the Ebro basin. In the other basins, the Po and Tisa, there is a consistent wet bias, with medians between 10 and 20 mm month^{-1} in the Po basin. In the Tisa basin, ICON has a median bias lower than 10 mm month^{-1} while TSMP1's median is closer to 20 mm month^{-1} .

In MAM, most experiments exhibit positive median biases, indicating a tendency toward wet conditions. The precipitation biases of ICON and TSMP1 differ markedly in the Ebro and Rhine basins. ICON shows a wet median

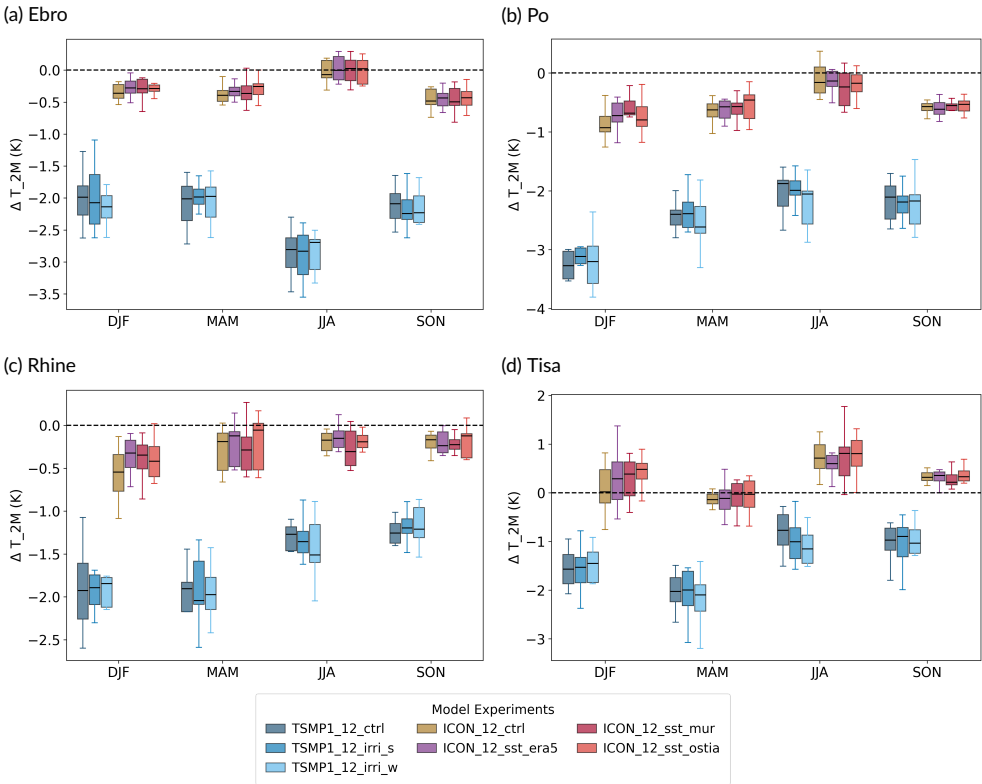


FIGURE 3 Boxplots of seasonal interannual T2m biases for different watersheds for the period 2003–2010.

bias of approximately $\pm 10 \text{ mm month}^{-1}$, whereas TSMP1 exhibits a larger bias of about $\pm 20 \text{ mm month}^{-1}$. By contrast, differences between models and experiments are less pronounced in the Po and Tisa basins.

In JJA, the model behavior varies across basins, but in general two groups can be distinguished again: the Ebro and Rhine basins on one side, and the Po and Tisa basins on the other. In the Ebro and Rhine basins, TSMP1 experiments maintain their characteristic wet bias, while ICON shifts to a dry bias. The spread among TSMP1 experiments is substantially larger than for ICON, and ICON's median biases remain much closer to zero, being the $-5.10 \text{ mm month}^{-1}$ (ICON_12_sst_mur) the closest to zero in the Ebro. In the Po basin, most experiments continue to show a wide spread. However, ICON median biases remain near zero, while TSMP1 continues to overestimate precipitation. In the Tisa basin, all experiments exhibit a dry bias, with median values between ± 20 and $\pm 30 \text{ mm month}^{-1}$.

In SON, ICON experiments persist with dry biases but with lower values, whereas TSMP1 maintains the wet biases observed in summer for the Ebro and Rhine basins. In the Po basin, both the spread and median biases remain similar to those of other seasons. In the Tisa basin, most experiments display a wet bias much closer to zero than in the other seasons. Overall, ICON biases are lower in all seasons and basins, however, both models still display a wide spread of biases, which is to be expected. Pham et al. (2021) also reported large ranges of precipitation biases for both ICON-CLM and COSMO-CLM, reflecting the high spatial variability of precipitation. Although they did not identify a clear advantage of one model over the other, our results indicate that coupling COSMO-CLM with different land surface and hydrologic models amplifies the differences in simulated precipitation.

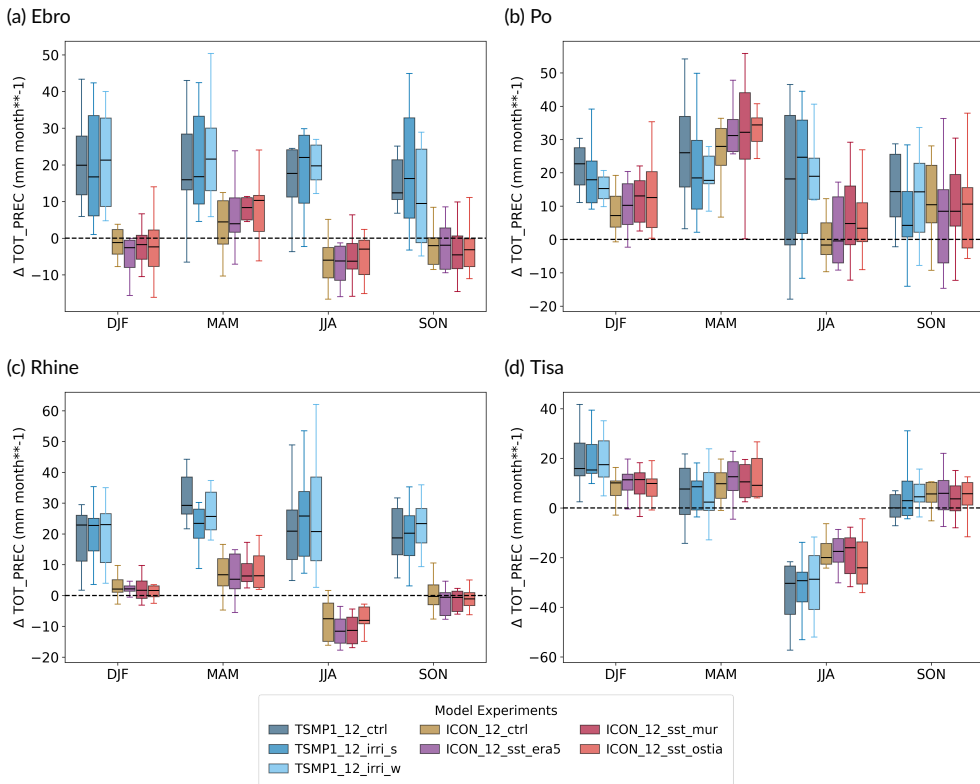


FIGURE 4 Boxplots of seasonal interannual precipitation biases for different watersheds for the period 2003–2010.

Similar to the seasonal distribution of T2m and precipitation biases, the seasonal distribution of ET biases reveals contrasts between the experiments, especially in MAM and JJA (Figure 5).

In DJF, biases are generally small across all simulations and basins, with medians close to zero and limited variability. However, an exception is the Po river basin, as the medians of all experiments are close to 5 mm month^{-1} , showing an overestimation.

In MAM, most experiments overestimate ET, with ICON simulations showing particularly consistent positive biases across basins. Although the TSMP1 experiments also overestimate ET in the Ebro and the Po basins, their medians are closer to zero compared to the ICON experiments. Moreover, the overall spread of TSMP1 results in the Ebro basin is smaller than that of the ICON experiments. In contrast, TSMP1 experiments slightly underestimate ET in the Rhine and Tisa river basins, often with medians lower than 5 mm month^{-1} .

In JJA, the models diverge considerably, as both the medians and the spread among experiments vary strongly. Overall, ICON simulations systematically overestimate ET, with median biases well above zero for the Po, Rhine, and Tisa watersheds. In addition, ICON experiments exhibit a substantially larger spread in the Po, and Tisa basins, particularly in Tisa, where values range from about -15 to 20 mm month^{-1} . In contrast, the spread of the bias in the TSMP1 experiments is much less in these basins. In the Ebro basin, ICON experiments perform better than TSMP1 as their median biases underestimate ET but remain closer to zero, whereas TSMP1 consistently overestimates ET, often exceeding 20 mm month^{-1} . This overestimation may be linked to the hydrologic representation in TSMP1. Naz et al. (2022) conducted ParFlow-CLM simulations over the CORDEX European domain at 3 km resolution and reported an ET overestimation in dry regions such as the Iberian Peninsula. Similarly, Poshyvailo-Strube et al. (2024) showed that TSMP simulations are particularly sensitive to groundwater representation in the Iberian Peninsula, highlighting this region as susceptible to such biases.

SON shows relatively moderate biases in most models and basins, with distributions centered close to zero. However, simulations in the Po and Rhine watersheds exhibit a larger positive bias, with medians around 5 mm month^{-1} . The TSMP1 results are consistent with the findings of Naz et al. (2022), who reported smaller discrepancies between ParFlow-CLM-simulated evapotranspiration and FLUXNET observations during winter, spring, and autumn, but larger differences in summer across most stations. They further suggested that the positive ET bias in summer could be caused by enhanced simulated water availability in ParFlow-CLM, a feature we also observed, particularly in the Ebro and Po basins. Overall, the boxplots highlight a strong seasonality in model performance, with ICON mostly favoring overestimations across different experiments and water basins, while TSMP1 shows variable behavior that depends on the basin and season.

3.2.2 | Yearly interannual variability (Period 2003–2010)

To provide a concise overview of model performance, Figures S5, S3 and S1 present both the seasonal biases and the yearly Mean Absolute Error (MAE) for T2m, precipitation, and ET across all analyzed basins for the current period. The results indicate that within the TSMP1 experiment group, the control simulation (TSMP1_12_ctrl) exhibits the lowest MAE for all three variables in the Ebro and Tisa basins, demonstrating the best overall mean agreement with the reference datasets.

Regarding the interannual variability in the Ebro basin, the Taylor diagram indicates that the setup TSMP1_12_irri_w most accurately reproduces the observed interannual precipitation variability, whereas the control simulation continues to perform best for ET and T2m (Figure S7a). In the Tisa basin, the control simulation attains again the most favorable combination of RMS, variance and correlation across all variables, confirming its overall superior performance (Figure S7d). In the Po basin, TSMP1_12_ctrl maintains the lowest MAE for ET, whereas the TSMP1_12_irri_s

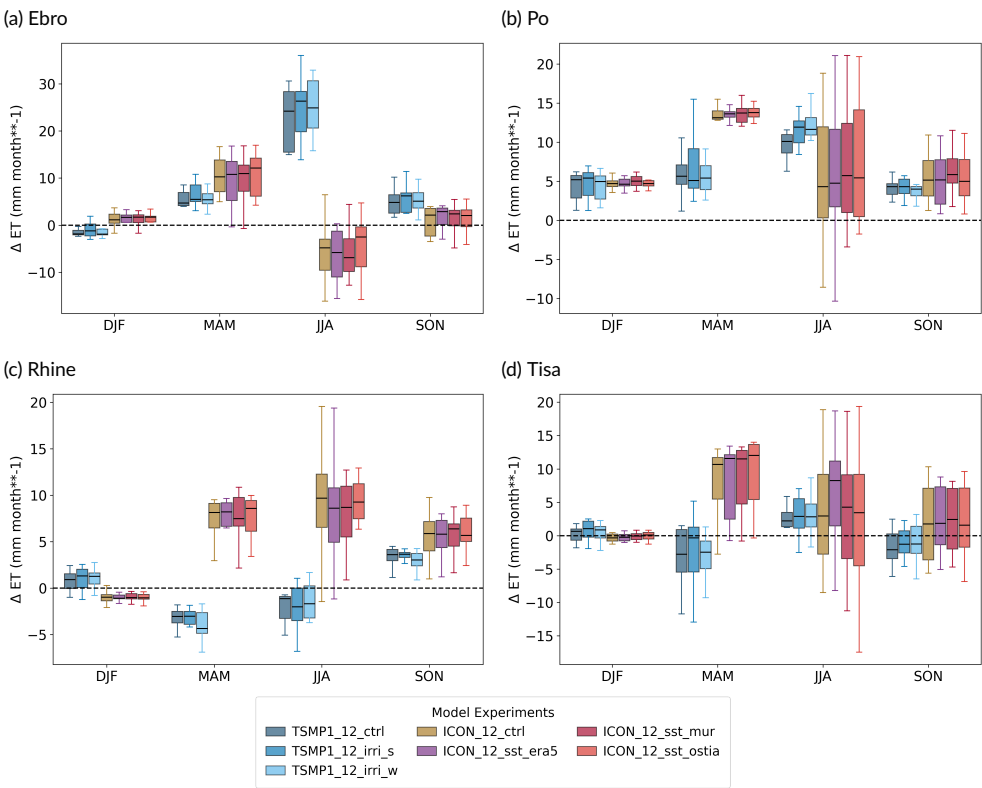


FIGURE 5 Boxplots of seasonal interannual ET biases for different watersheds for the period 2003–2010.

experiment performs best for precipitation and T2m. This improved performance in the irrigation experiment highlights the importance of representing irrigation processes in this basin, which contains one of the largest irrigated areas in Europe. Similarly, in the Rhine basin, the TSMP1 simulations that include irrigation achieve the best performance. Specifically, TSMP1_12_irri_w produces the lowest MAE for ET, while TSMP1_12_irri_s performs best for precipitation and T2m. Interannual Taylor diagrams further support these findings. In the Po basin, TSMP1_12_irri_w outperforms other configurations for ET and precipitation, whereas TSMP1_12_irri_s provides the best representation of T2m (Figure S7b). In the Rhine basin, TSMP1_12_irri_w achieves the best agreement for ET and T2m, while the control simulation (TSMP1_12_ctrl) performs best for precipitation (Figure S7c). Although irrigation is less extensive in the Rhine basin compared to the Po, these results suggest that even median to small irrigated areas can influence the surface–atmosphere exchanges.

The evaluation of ICON_12 simulations across the analyzed basins reveals different sensitivities to SST boundary conditions. In the Ebro basin, the control simulation (ICON_12_ctrl) achieves the lowest MAE for ET, while ICON_12_sst_mur and ICON_12_sst_era5/ICON_12_sst_ostia perform best for precipitation and T2m respectively. This pattern indicates that SST boundary forcing might play a relevant role in modulating precipitation and temperature fields, emphasizing the importance of using observation-based SST datasets as lower boundary in free simulations. Similarly, in the Po basin, ICON_12_sst_era5 obtains the lowest MAE for ET and T2m, whereas the ICON_12_ctrl simulation performs best for precipitation. Although ICON_12_ctrl and ICON_12_sst_era5 best capture the mean state biases in the Ebro and Po basins for different variables, the interannual variability illustrated in the Taylor diagrams indicates that no single experiment consistently outperforms the others. For instance, ICON_12_sst_mur shows the highest agreement for ET in the Ebro basin (Figure S7a), while ICON_12_sst_era5 provides the best representation of precipitation. In the Po basin, although ICON_12_sst_mur obtains the lowest RMS for ET, the correlation is negative (Figure S7b). In the Rhine basin, all SST experiments outperform the control simulation, with ICON_12_sst_era5 achieving the lowest MAE for ET and ICON_12_sst_ostia for precipitation and T2m. This suggests that, in a basin characterized by strong synoptic influence, the representation of SST conditions significantly influences the simulated surface fluxes and temperature. In the Tisa basin, results are partly consistent with those of the Ebro and Po basins since the ICON_12_ctrl simulation shows the best performance for ET and T2m, while ICON_12_sst_mur exhibits the lowest MAE for precipitation. Taylor diagram analysis indicates that in the Rhine, no single configuration consistently outperforms the others across all variables (Figure S7c). However, in the Tisa basin ICON_12_sst_mur exhibits the highest skill for precipitation and T2m, while the control run best reproduces ET, matching with the MAE results. The consistent better representation of precipitation by ICON_12_sst_mur in this basin might be related to the high resolution of MURSST dataset (1 km), however, this hypothesis still needs some testing. The outcomes indicate that, while SST boundary conditions affect atmospheric variables to varying degrees, their influence is region-specific.

3.2.3 | Seasonal interannual variability (Period 2011–2020)

For the period 2011–2020, we also compare ICON and TSMP1 experiments, however, ICON simulations are only available at 3 km resolution. Similar to the period 2003–2010, the bias distribution for T2m shows pronounced differences between TSMP1 and ICON experiments for all basins and seasons (Figure 6). In DJF, both models exhibit a cold bias in the Ebro, Po, and Rhine basins. On average, ICON experiments show relatively small biases, with a median cold bias of less than 1 K, whereas TSMP1 simulations are systematically colder, with 50% of the data below -1.5 K. In the Tisa basin, ICON cold bias is distributed around zero, while TSMP1 retains a median cold bias below -1 K. In MAM, both models maintain a cold bias as well, although ICON medians remain consistently closer to zero than TSMP1. Similarly, JJA is the season when the differences between models increase. In the Ebro basin, ICON simulations perform

notably better, specifically ICON_3_irr experiment which has distributions centered near zero, while TSMP1 biases are grouped around -2.5 K. The data spread is comparable for both models. For the Po, Rhine, and Tisa basins, ICON exhibits a warm bias in about 50% of its data, with median values of $+1.5$ K, $+0.5$ K, and $+2$ K, respectively. Iriza-Burca et al. (2024) also found that ICON-LAM and COSMO-LAM at 2.8 km resolution tend to overestimate summer forecasted values in Romania compared to observations, but only during nighttime. While our analysis focused on monthly means, this discrepancy may be related. In contrast, TSMP1 continues to show cold biases with medians around -2 K, -1.5 K, and -1.5 K, and generally smaller spreads than ICON. In SON, ICON also outperforms TSMP1, as its bias distribution is centered around zero in the Po and Rhine, while in the Ebro the bias is centered around -0.5 K. The Tisa basin remains an exception, where ICON continues to show a warm bias of about $+0.5$ K. Here, TSMP1 biases are consistently colder with a bias spread comparable to ICON. In general, ICON experiments show smaller and more balanced seasonal biases, while TSMP1 consistently maintains a stronger cold bias across all basins and seasons. These results are consistent with Rieger et al. (2022), who found that ICON-LAM outperforms COSMO-LAM in representing surface variables such as T2m and temperature profiles across different regional configurations (e.g., ICON-IT vs. COSMO-IT, ICON-RO vs. COSMO-RO).

While the seasonal bias analysis (2011–2020) excludes the reanalysis (ICON_3_rea) dataset due to its shorter simulation period (2018–2020), insights from the corresponding time series analysis (Figure S9 and Table S2, Appendix) provide additional context for ICON simulations. ICON_3_rea results show consistently lower RMSE values than ICON_3 and TSMP1 experiments in the Po, Rhine, and Tisa basins, further supporting its better short-term performance in reproducing T2m patterns. Therefore, assimilating T2m and relative humidity through the DA routine allows to improve these scores.

Unlike T2m, the seasonal bias distribution for precipitation shows mainly a moderate difference between ICON and TSMP1 simulations (Figure 7), as the bias distribution is comparable across most basins and seasons. During DJF, all experiments exhibit a wet bias in approximately 50% of the data, with the strongest overestimation found in the Rhine and Tisa basins for the TSMP1 simulations. In both the Rhine and Tisa basins, ICON outperforms TSMP1, showing narrower bias spreads and medians below 20 mm month^{-1} . Other high-resolution simulations (2.8 km) over Romania (within the Tisa basin) with ICON-LAM and COSMO-LAM also indicate that both models overestimate the magnitude and spatial extent of intense winter precipitation (Iriza-Burca et al., 2024). Similarly, in MAM, all experiments systematically overestimate precipitation, with consistent results in all basins. The largest differences between TSMP1 and ICON occur in JJA, particularly in the Ebro, Po, and Rhine basins. In Ebro and Po, ICON biases are centered close to zero, indicating good performance, while TSMP1 strongly overestimates precipitation, with median biases exceeding 10 mm month^{-1} . In contrast, the Rhine shows opposite tendencies between models as TSMP1 simulations are predominantly wet-biased and ICON simulations are predominantly dry-biased. The generally better representation of precipitation by ICON in this season and across these basins may be linked to its treatment of shallow and deep convection. For instance, Lucia et al. (2022) showed that ICON simulations at 2.5 km resolution over Italy during summer performed better when deep convection was switched off compared to runs where it was enabled. In the Tisa basin, the bias distribution is similar across all experiments, with about half of the data consistently showing a dry bias. In SON, the bias distribution is again comparable between experiments in the Ebro, Po, and Tisa basins. The Tisa basin has generally the best representation, as 50% of the distribution of all experiments centers around zero. In the Rhine, ICON performs better, with median biases closer to zero and smaller spreads than TSMP1. Overall, ICON tends to reproduce precipitation more accurately in summer and autumn, whereas both models systematically overestimate precipitation in winter and spring.

Additional insights from the reanalysis dataset for the period 2018–2020 (Figure S10 and Table S2) confirm ICON_3 tendencies. ICON_3_rea simulation achieved lower RMSE values than both ICON_3 and TSMP1 across all

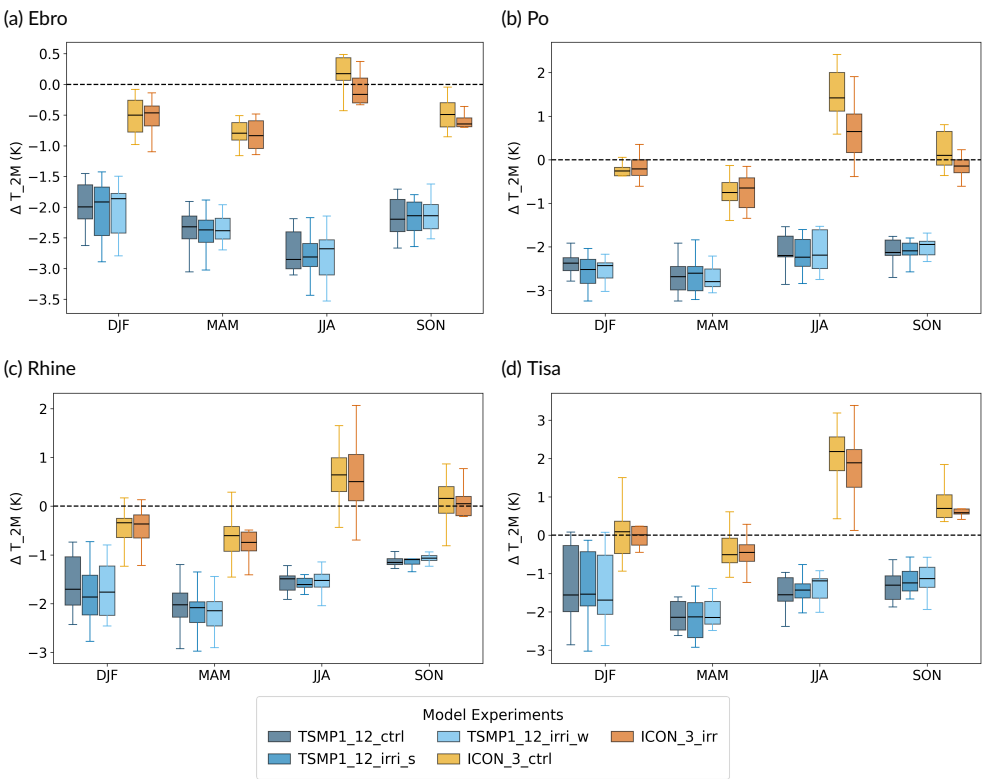


FIGURE 6 Boxplots of seasonal interannual T2m biases for different watersheds for the period 2011–2020.

basins, indicating again improved short-term skill in reproducing precipitation variability.

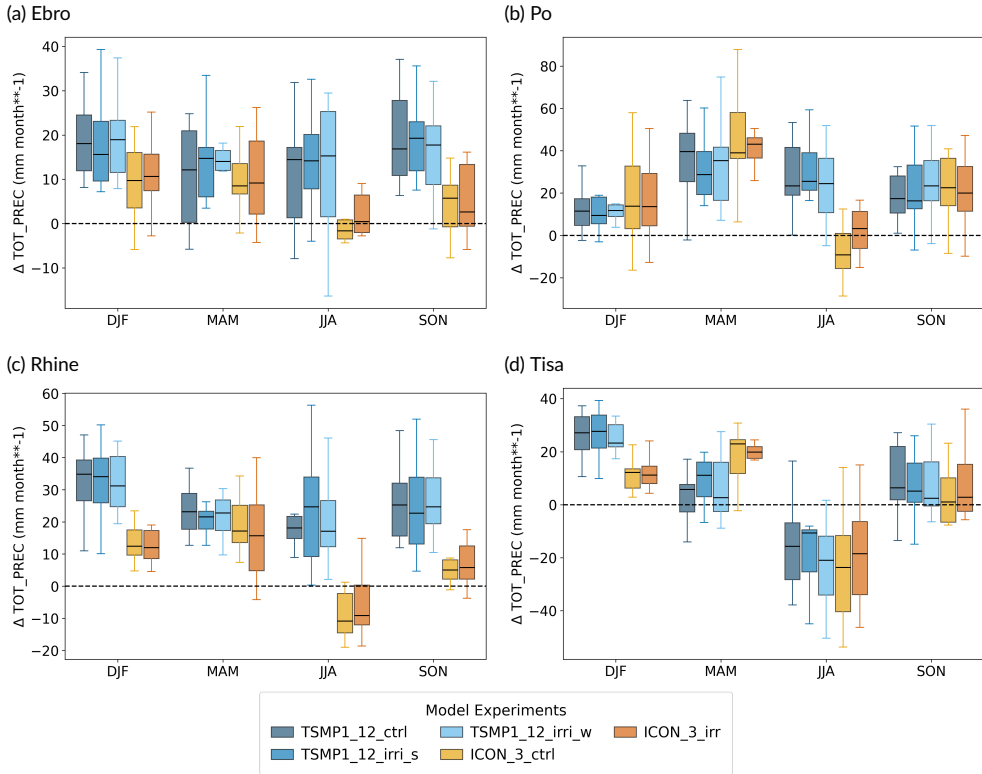


FIGURE 7 Boxplots of seasonal interannual precipitation biases for different watersheds.

Figure 8 presents the distribution of seasonal ET biases averaged per basin. DJF generally shows small biases, which are close to zero for most experiments and basins. However, Po river basin is an exception, as the medians are located around 5 mm month^{-1} . In MAM, the bias spread increases across all experiments, although the direction of the bias varies per basin. In the Ebro and Po basins, both models overestimate ET, with TSMP1 experiments medians nearer to zero than ICON. In the case of the Rhine and Tisa basins, ICON continues overestimating ET while TSMP1 underestimates it. In general, in both basins, the biases are of similar magnitude in opposite directions. During JJA, ICON_3_ctrl exhibits median biases closer to zero in the Ebro and Po basins compared to TSMP1. In the Rhine and Tisa basins, both the spread and median biases of TSMP1 are lower compared to ICON. SON shows a moderate bias in most experiments and basins, generally reflecting ET overestimation. In the Ebro and Tisa basins, bias distributions are closer to zero for both models, whereas in the Po and Rhine basins ICON experiments show larger spreads. In general, the boxplots show that the TSMP1 experiments generally exhibit smaller spreads and biases than ICON, particularly in the Rhine and Tisa basins, while ICON performs better in reducing median biases in the Ebro basin.

Complementary results of ICON_3_rea, period 2018–2020, (Figure S11 and Table S3) provide additional context for ICON simulations. During this overlapping period, the reanalysis exhibits lower RMSE values than the ICON_3_ctrl and ICON_3_irr simulations in the Ebro, Rhine and Tisa basins, suggesting a better monthly representation of evapotranspiration. So, even though evapotranspiration is not assimilated, its representation improves. However, in the Po

basin, the highest bias and RMSE correspond to ICON_3_rea. These results might indicate a poor representation of the observed variability on a monthly scale.

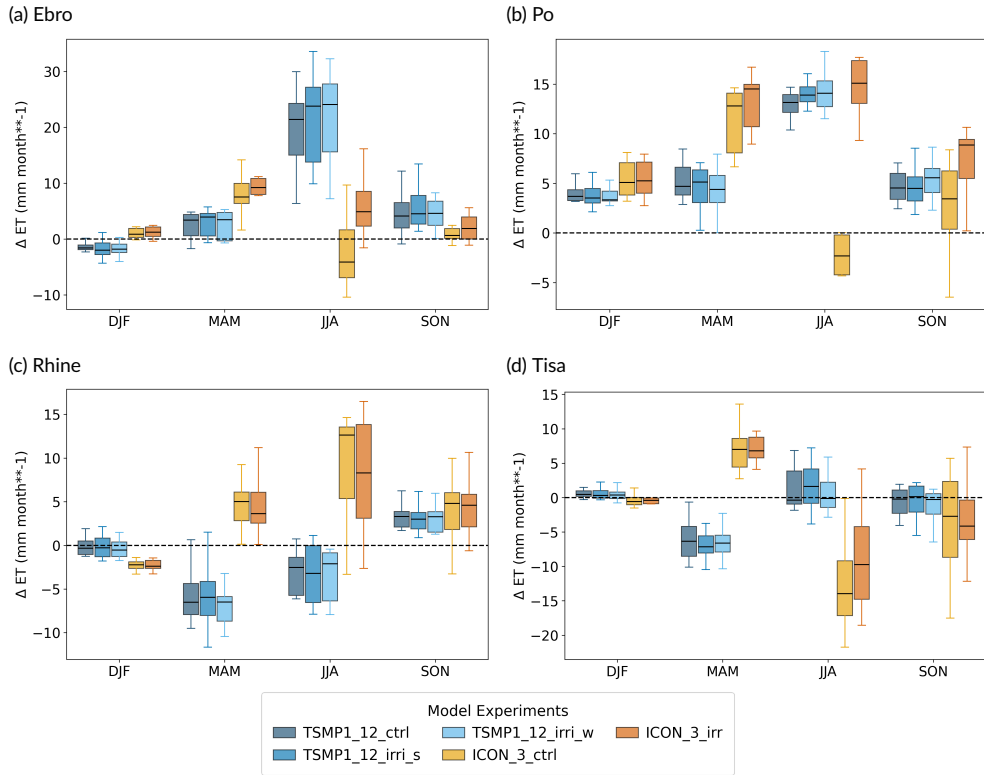


FIGURE 8 Boxplots of seasonal interannual ET biases for different watersheds for the period 2011–2020.

3.2.4 | Yearly interannual variability (Period 2011–2020)

The mean absolute error (MAE) for the 2011–2020 period is shown in Figures S2, S4 and S6. They present seasonal biases and the yearly Mean Absolute Error (MAE) for ET, precipitation, and T2m across all analyzed basins. Here we focus on the comparison within two modeling groups, TSMP1 and ICON. Similar to the previous period, results indicate that within the TSMP1 experiment group, the control simulation (TSMP1_12_ctrl) exhibits the lowest MAE for ET and T2m in the Ebro basin, demonstrating the best overall agreement with the reference datasets. Taylor diagrams support the fact that TSMP1_12_ctrl achieves the best results for ET and T2m, while TSMP1_12_irri_s obtains the lowest RMS and the highest correlation for precipitation (Figure S8a). In the Po basin, TSMP1_12_irri_s again achieve the best performance in precipitation with the lowest MAE for ET, while TSMP1_12_ctrl performs best for ET. T2m has TSMP1_12_ctrl and TSMP1_12_irri_w very close. Interannual Taylor diagrams further support that TSMP1 irrigation experiments always outperform in all variables, specifically TSMP1_12_irri_w (Figure S8b). This consistent improvement in the irrigation experiments highlights the continued importance of representing irrigation processes in this basin, which is among the most heavily irrigated regions in Europe. For the Rhine basin, both the

control simulation and TSMP1_12_irri_s achieve the lowest MAE for ET, indicating comparable skill in reproducing surface fluxes. Precipitation is best represented by TSMP1_12_irri_w, whereas the control run performs best for T2m. Likewise, the interannual variability of ET is best captured by TSMP1_12_ctrl, while the representation of interannual variability in precipitation and T2m is superior in TSMP1_12_irri_w (Figure S8c). This differs from the previous period, where irrigation configurations consistently outperformed the control run. These results might suggest that model performance in this basin is highly sensitive to interannual climatic variability. In the Tisa basin, the results are again mixed, with different experiments performing best for different variables. TSMP1_12_irri_s shows the lowest MAE for ET, the control simulation the lowest for precipitation, and TSMP1_12_irri_w for T2m. When representing the interannual variability in this basin, TSMP1_12_irri_w outperforms the other experiments in ET and T2m, however, the control run obtains better results for precipitation (Figure S8d). These findings emphasize that the model skill at 12 km for TSMP1 varies by variable and basin, reflecting differences in land-surface processes, and even in the influence of irrigation on surface energy and water fluxes.

The ICON experiments ICON_3_ctrl and ICON_3_irr focus on the impact of irrigation for the period 2011–2020. In the Ebro and Tisa basins, the control simulation achieves the lowest MAE for ET, precipitation and T2m. Figure S8a supports these results since ICON_3_ctrl RMSE is the lowest for all three variables in the Ebro basin and for precipitation and ET in the Tisa basin. In the case of T2m, ICON_3_irr obtained a slightly lower RMSE than the control. The better performance of ICON_3_ctrl for ET might be attributed to the TERRA land surface scheme, which is already well-tuned within the ICON modeling framework. In the Po Basin, ICON_3_ctrl shows the lowest annual MAE for ET and T2m, while ICON_3_irr performs best for precipitation. Although ICON_3_ctrl performs slightly better annually (0.23 vs. 0.25 MAE), ICON_3_irr improves summer results, reducing the biases from 1.48 K to 0.65 K. In addition, ICON_3_irr obtains a marginally lower RMSE than the control for T2m when representing the interannual variability (Figure S8b). This experiment also obtains the lowest RMS errors for ET, although it exhibits a negative correlation. The same T2m trend holds for the Rhine in relation to the MAE, since ICON_3_ctrl shows only a slightly better improvement for T2m (0.26 vs. 0.28 MAE). In the seasonal comparison, ICON_3_irr exhibits the best agreement in JJA. This irrigation experiment also improves the MAE for ET. The interannual variability for the Rhine shows that ICON_3_irr provides the best representation of T2m, characterized by positive correlations and the lowest RMSE. This experiment also achieves the lowest RMSE with a negative correlation for ET. ICON_3_ctrl outperforms for precipitation with lower RMSE, higher correlation and similar STD than GPCC. These findings imply that irrigation is indeed an anthropogenic factor that influences T2m, an effect widely recognized within the Earth system community (Valmassoi and Keller, 2022). In the current experiments, irrigation also appears to improve the representation of precipitation in the Po basin, a variable that remains subject to considerable uncertainties regarding its interaction with irrigation (McDermid et al., 2023). In any case, more analysis is needed to confirm this improvement. Overall, the results highlight again the importance of accounting for irrigation in weather and climate simulations, especially for the summer season. However, further parameter tuning and reduction of irrigation related uncertainties remain necessary for the rest of the year.

4 | SUMMARY AND CONCLUSIONS

The present study presents a multi-model regional climate model ensemble dataset of the CRC 1502 - DETECT, designed for water cycle process analyses. We compare the performance of the atmospheric simulations in representing key aspects of the regional climate over the period 1990–2020 across four representative basins within the EURO-CORDEX domain. Our analysis focuses on the variables 2 m air temperature (T2m), precipitation (P) and evapotran-

spiration (ET), evaluated using monthly values spatially-averaged over the Ebro, Po, Rhine, and Tisa basins. To assess the model performance against the reference observational datasets GLEAM, GPCC, and E-OBS, we employ common statistical metrics as well as Taylor diagrams to analyze the interannual variability of seasonal and yearly averages or accumulations. This study provides a baseline analysis of the simulated weather and climate in the DETECT project pan-European analysis domain and may inform ensuing analyses on water cycle processes, such as land-atmosphere interactions, and can be used to evaluate future model and simulation experiment improvements. The dataset will become publicly available also within the course of the DETECT project.

Our study shows that all experiments systematically underestimate T2m across most seasons and basins, although the magnitude of the bias differs between ICON and TSMP1. In general, TSMP1 shows higher variability and stronger cold biases often exceeding -2 K, whereas ICON performs particularly better, with generally smaller biases. Moreover, the largest model differences for this variable are observed in summer. Regarding P and ET, we found that all simulations across both periods (2003–2010, 2011–2020) exhibit a comparable bias distribution in winter; while the largest differences between model performance occur in spring and, most prominently, in summer. For instance, substantial contrasts between ICON and TSMP1 simulations are evident for P, particularly over the Ebro and Rhine river basins in all seasons, with the largest differences occurring in summer. In contrast, the Tisa basin shows the smallest inter-model differences across all experiments. Overall, ICON exhibits smaller biases across all seasons and basins. However, both models display considerable spread in their P bias, which is expected given the inherent difficulty of simulating P (Giorgi, 2019; Tapiador et al., 2019). In the case of ET, ICON simulations systematically overestimate this variable and show a larger spread, whereas the TSMP1 experiments exhibit a narrower bias distribution. In general, TSMP1 outperforms ICON in the Rhine and Tisa basins, whereas ICON performs better in the Ebro basin. These results highlight that model performance varies substantially by basin and variable and may be clearly attributed to model structural differences.

When analyzing the interannual variability across all variables and basins per simulation group, we found that both TSMP1 and ICON_3 irrigation experiments tend to perform better in the Po basin, mainly for T2m and P. ET also shows a yearly improved representation in TSMP1 simulations. These results are supported by lower MAE and RMS error values compared with their respective control simulation. Given that the Po basin is among the most intensively irrigated regions in Europe, these findings emphasize the importance of accurately representing irrigation processes to improve the model performance in this basin.

The SST sensitivity experiments conducted with the ICON_12 simulations reveal that SST influences T2m and P simulations within the EURO-CORDEX domain. This is especially the case when analyzing results from the Rhine basin, where all SST experiments outperformed the control run across all assessed variables. Among the configurations, ICON_12_sst_era5 presented better results in simulated T2m for three of the four analyzed basins, while ICON_12_sst_mur stood out in two of them. Using foundation SSTs instead of skin SSTs, da Silva Lopes et al. (2025) found an increase of about 2.4% in annual continental precipitation and a corresponding rise in surface freshwater flux, at finer spatial scales, precipitation differences were locally up to five times larger than the domain average. These results highlight the importance of accurate SST specification in constraining regional climate model boundary conditions and improving simulated land-atmosphere interactions.

This study provides a comprehensive, general intercomparison of regional atmospheric simulations for essential climate variables, with a focus on water cycle investigations. Despite the dedicated ensemble, we acknowledge some limitations. First, the simulation period analyzed was relatively short, especially when divided into two sub-periods. Second, all simulations have different specific goals, and their model configuration reflects those distinct objectives. Third, as mentioned in the Methods section, the reference datasets contain inherent uncertainties, particularly in regions with sparse station coverage, such as E-OBS. Finally, we focused on monthly averages for specific basins,

omitting daily, diurnal, and spatial variability. In the future, more specific events that require shorter timescales can be analyzed, such as heatwaves or heavy precipitation events. Nonetheless, this study lays the basis for future intercomparisons and integrated analyses across the diverse simulation packages and modeling components developed within CRC 1502 - DETECT and can be used as a baseline for future model and simulation experiment improvements.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1502/1–2022 - Projektnummer: 450058266.

We would like to thank members of the CRC 1502 - DETECT for their supervision, discussion, and their support and contributions to model development, data generation, data provisioning, as well as input to the set-up and operation of simulations: Stefan Siebert (B05), Stefan Kollet and Michael Schindelegger (D02); support was also provided by Niklas Wagner.

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS supercomputer JUWELS (Jülich Supercomputing Centre, 2021b) and on the supercomputer JURECA (Jülich Supercomputing Centre, 2021a) under grant no. cjjsc39, detectrea, both at the Jülich Supercomputing Centre (JSC) at Forschungszentrum Jülich.

The authors thankfully acknowledge the Deutsche Wetterdienst (DWD - German Weather Service) for providing computing resources in the operational High Performance Computing system.

We thank the research data infrastructure and services provided by the Jülich Supercomputing Centre, Germany, as part of data project 'detectdata'.

We thank the "Global Runoff Data Centre at the Federal Institute of Hydrology (BfG), 56068 Koblenz, Germany" for providing HydroSHEDS-based watershed boundaries.

We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (<http://www.uerra.eu>) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (<https://www.ecad.eu>).

conflict of interest

The authors declare no conflict of interest.

references

- Akella, S., Todling, R. and Suarez, M. (2017) Assimilation for skin SST in the NASA GEOS atmospheric data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **143**, 1032–1046.
- Allen, R., Pereira, L., Raes, D. and Smith, M. (1998) Crop evapotranspiration: Guidelines for computing crop requirements. *Tech. rep.*, FAO. URL: <https://www.fao.org/4/x0490e/x0490e00.htm#Contents>.
- Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Weather Rev.*, **139**, 3887–3905.
- Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., Ahrens, B., Alias, A., Anders, I., Bastin, S., Belušić, D., Berthou, S., Brisson, E., Cardoso, R. M., Chan, S. C., Christensen, O. B., Fernández, J., Fita, L., Frisius, T., Gašparac, G., Giorgi, F., Goergen, K., Haugen, J. E., Hodnebrog, Ø., Kartsios, S., Katragkou, E., Kendon, E. J., Keuler, K., Lavin-Gullon, A., Lenderink, G., Leutwyler, D., Lorenz, T., Maraun, D., Mercogliano, P., Milovac, J., Panitz, H.-J., Raffa, M., Remedio, A. R., Schär, C., Soares, P. M. M., Srncic, L., Steensen, B. M., Stocchi, P., Tölle, M. H., Truhetz, H., Vergara-Temprado, J.,

- de Vries, H., Warrach-Sagi, K., Wulfmeyer, V. and Zander, M. J. (2021) The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Climate Dynamics*, **57**, 275–302. URL: <https://doi.org/10.1007/s00382-021-05708-w>.
- Barlage, M., Chen, F., Rasmussen, R., Zhang, Z. and Miguez-Macho, G. (2021) The Importance of Scale-Dependent Groundwater Processes in Land-Atmosphere Interactions Over the Central United States. *Geophysical Research Letters*, **48**, e2020GL092171. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL092171>.
- Casirati, S., D'Odorico, P., Reichle, R. H., Rouf, T., Stookey, M. and Giroto, M. (2025) A model-based evaluation of the effects of irrigation expansion on regional and global land surface climate. *Earth's Future*, **13**, e2025EF006271. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2025EF006271>. E2025EF006271 2025EF006271.
- Cassola, F., Ferrari, F., Mazzino, A. and Miglietta, M. M. (2016) The role of the sea on the flash floods events over Liguria (northwestern Italy). *Geophys. Res. Lett.*, **43**, 3534–3542.
- Chin, T. M., Vazquez-Cuervo, J. and Armstrong, E. M. (2017) A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sens. Environ.*, **200**, 154–169.
- Christensen, J. H. and Christensen, O. B. (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change*, **81**, 7–30. URL: <https://doi.org/10.1007/s10584-006-9210-7>.
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. and Jones, P. D. (2018) An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, **123**, 9391–9409.
- Doms, G., Förstner, J., Heise, E., Herzog, H. J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J. P. and Vogel, G. (2011) A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization. *Deutscher Wetterdienst, Offenbach*.
- Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E. and Wimmer, W. (2012) The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sens. Environ.*, **116**, 140–158.
- Donlon, C. J., Minnett, P. J., J., C. G. T., Nightingale, Barton, I. J., Ward, B. and Murray, M. J. (2002) Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369.
- ECMWF (2020) IFS Documentation CY47R1–Part IV: Physical Processes. In *IFS Documentation CY47R1*. ECMWF.
- Fairall, C. W., Bradley, E. F., Godfrey, J. S., Wick, G. A., Edson, J. B. and Young, G. S. (1996) Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res. Oceans*, **101**, 1295–1308.
- Ferro, M., Dutta, T., Hüttel, S., Lindner, M., Poll, S. and Börner, J. (2025) Climate-relevant land cover composition and configuration trajectories in europe. *Environmental Research Letters*, **20**, 054018. URL: <https://doi.org/10.1088/1748-9326/adc8be>.
- Fosser, G., Kendon, E. J., Stephenson, D. and Tucker, S. (2020) Convection-permitting models offer promise of more certain extreme rainfall projections. *Geophysical Research Letters*, **47**, e2020GL088151. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088151>. E2020GL088151 2020GL088151.
- Furusho-Percot, C., Goergen, K., Hartick, C., Poshyvailo-Strube, L. and Kollet, S. (2022) Groundwater model impacts multianual simulations of heat waves. *Geophys. Res. Lett.*, **49**, e2021GL096781.
- Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M. and Kollet, S. (2014) Implementation and scaling of the fully coupled terrestrial systems modeling platform (TerrSysMP v1.0) in a massively parallel supercomputing environment – a case study on JUQUEEN (IBM blue gene/q). *Geosci. Model Dev.*, **7**, 2531–2543.
- Gassmann, A. and Herzog, H. J. (2008) Towards a consistent numerical compressible non-hydrostatic model using generalized Hamiltonian tools. *Quart. J. Roy. Meteor. Soc.*, **134**, 1597–1613.

- Giorgi, F. (2019) Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, **124**, 5696–5723. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD030094>.
- Goergen, K. and Kollet, S. (2021) Boundary condition and oceanic impacts on the atmospheric water balance in limited area climate model ensembles. *Scientific Reports*, **11**, 6228. URL: <https://doi.org/10.1038/s41598-021-85744-y> <http://www.nature.com/articles/s41598-021-85744-y>.
- GRDC (2023) Watershed boundaries of grdc stations. URL: https://grdc.bafg.de/products/basin_layers/watershed_boundaries/.
- Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M. and Seneviratne, S. I. (2014) Global assessment of trends in wetting and drying over land. *Nature Geoscience*, **7**, 716–721. URL: <https://doi.org/10.1038/ngeo2247>.
- Gudmundsson, L., Boulange, J., Do Hong, X., Gosling Simon, N., Grillakis Manolis, G., Koutroulis Aristeidis, G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne Sonia, I., Satoh, Y., Thiery, W., Westra, S., Zhang, X. and Zhao, F. (2021) Globally observed trends in mean and extreme river flow attributed to climate change. *Science*, **371**, 1159–1162. URL: <https://doi.org/10.1126/science.aba3996>.
- Gutowski, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., Arritt, R. W., Bukovsky, M. S., Calvin, K. V., Feng, Z., Jones, A. D., Kooperman, G. J., Monier, E., Pritchard, M. S., Pryor, S. C., Qian, Y., Rhoades, A. M., Roberts, A. F., Sakaguchi, K., Urban, N. and Zarzycki, C. (2020) The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bulletin of the American Meteorological Society*, **101**, E664 – E683. URL: <https://journals.ametsoc.org/view/journals/bams/101/5/bams-d-19-0113.1.xml>.
- Hartick, C., Furusho-Percot, C., Clark, M. P. and Kollet, S. (2022) An interannual drought feedback loop affects the surface energy balance and cloud properties. *Geophys. Res. Lett.*, **49**, e2022GL100924.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D. and New, M. (2008) A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*, **113**.
- Heinze, R., Dipankar, A., Henken, C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Girolamo, P. D., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H. C., Hoose, C., Jahns, T., Kalthoff, N., Klocke, D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C. I., Muppa, S. K., Neggers, R. A. J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D. and Quaas, J. (2017) Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Quart. J. Roy. Meteor. Soc.*, **143**, 69–100.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Sabater, J. M., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D. and Thépaut, J. N. (2020) The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049.
- Hirahara, S., Balmaseda, M. A., Boisseson, E. D. and Hersbach, H. (2016) Sea surface temperature and sea ice concentration for ERA5. *ERA Report Series 26*, European Centre for Medium Range Weather Forecasts, Berkshire, UK. <https://www.ecmwf.int/sites/default/files/elibrary/2016/16555-sea-surface-temperature-and-sea-ice-concentration-era5.pdf>.
- Hirsch, A. L., Guillod, B. P., Seneviratne, S. I., Beyerle, U., Boysen, L. R., Brovkin, V., Davin, E. L., Doelman, J. C., Kim, H., Mitchell, D. M., Nitta, T., Shiogama, H., Sparrow, S., Stehfest, E., van Vuuren, D. P. and Wilson, S. (2018) Biogeophysical Impacts of Land-Use Change on Climate Extremes in Low-Emission Scenarios: Results From HAPPI-Land. *Earth's Future*, **6**, 396–409. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2017EF000744>. [_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017EF000744](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017EF000744).

- Hirschi, M., Stradiotti, P., Crezee, B., Dorigo, W. and Seneviratne, S. I. (2025) Potential of long-term satellite observations and reanalysis products for characterising soil drying: trends and drought events. *Hydrology and Earth System Sciences*, **29**, 397–425. URL: <https://hess.copernicus.org/articles/29/397/2025/>. Publisher: Copernicus GmbH.
- Hogan, R. K. and Bozzo, A. (2018) A flexible and efficient radiation scheme for the ECMWF model. *J. Adv. Model. Earth Syst.*, **10**, 1990–2008.
- Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behraves, M., Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datsis, G., Esch, M., George, G., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemann, F., Marotzke, J. and Stevens, B. (2023) ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. *Geosci. Model Dev.*, **16**, 779–811.
- Huntington, T. G. (2006) Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, **319**, 83–95. URL: <http://doi.org/10.1016/j.jhydrol.2005.07.003>.
- Iriza-Burca, A., Dumitrache, R. C., Maco, B. A., Huştiu, M., Fundel, F., Rieger, D. and Potthast, R. (2024) Comparison of cosmo and icon-lam high-resolution numerical forecast for romanian territory: Case studies and evaluation. *Atmósfera*, **38**, 421–451. URL: <https://www.revistascca.unam.mx/atm/index.php/atm/article/view/53305>.
- Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., E., Cardoso, R. M., Casanueva, A., Christensen, O. B., Christensen, J. H., Coppola, E., Cruz, L. D., Davin, E. L., Dobler, A., Domínguez, M., Fealy, R., Fernandez, J., Gaertner, M. A., García-Díez, M., Giorgi, F., Gobiet, A., Goergen, K., Gómez-Navarro, J. J., Alemán, J. J. G., Gutiérrez, C., Gutiérrez, J. M., Güttler, I., Haensler, A., Halenka, T., Jerez, S., Jiménez-Guerrero, P., Jones, R. G., Keuler, K., Kjellström, E., Knist, S., Kotlarski, S., Maraun, D., van Meijgaard, E., Mercogliano, P., Montávez, J. P., Navarra, A., Nikulin, G., de Noblet-Ducoudré, N., Panitz, H. J., Pfeifer, S., Piazza, M., Pichelli, E., Pietikäinen, J. P., Prein, A. F., Preuschmann, S., Rechid, D., Rockel, B., Romera, R., Sánchez, E., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Sørland, S. L., Termonia, P., Truhetz, H., Vautard, R., Warrach-Sagi, L. and Wulfmeyer, V. (2020) Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Reg. Environ. Change*, **20**, 51.
- Jensen, L., Gerdener, H., Eicker, A., Kusche, J. and Fiedler, S. (2024) Observations indicate regionally misleading wetting and drying trends in CMIP6. *npj Climate and Atmospheric Science*, **7**, 249. URL: <https://www.nature.com/articles/s41612-024-00788-x>. Publisher: Nature Publishing Group.
- Jones, J. E. and Woodward, C. S. (2001) Newton–krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Adv. Water Resour.*, **24**, 763–774.
- Jülich Supercomputing Centre (2021a) JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre. *Journal of large-scale research facilities*, **7**. URL: <http://dx.doi.org/10.17815/jlsrf-7-182>.
- (2021b) JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre. *Journal of large-scale research facilities*, **7**. URL: <http://dx.doi.org/10.17815/jlsrf-7-183>.
- Katragkou, E., García-Díez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P. M. M., Pytharoulis, I., Tegoulis, I., Tsikerdekis, A. and Jacob, D. (2015) Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. *Geosci. Model Dev.*, **8**, 603–618.
- Katragkou, E., Sobolowski, S. P., Teichmann, C., Solmon, F., Pavlidis, V., Rechid, D., Hoffmann, P., Fernandez, J., Nikulin, G. and Jacob, D. (2024) Delivering an Improved Framework for the New Generation of CMIP6-Driven EURO-CORDEX Regional Climate Simulations. *Bulletin of the American Meteorological Society*, **105**, E962–E974. URL: <https://doi.org/10.1175/BAMS-D-23-0131.1>.

- Keune, J., Gasper, F., Goergen, K., Hense, A., Shrestha, P., Sulis, M. and Kollet, S. (2016) Studying the influence of groundwater representations on land surface-atmosphere feedbacks during the European heat wave in 2003. *Journal of Geophysical Research: Atmospheres*, **121**, 13301–13325. URL: <https://dx.doi.org/10.1002/2016JD025426>.
- Keune, J., Sulis, M., Kollet, S., Siebert, S. and Wada, Y. (2018) Human water use impacts on the strength of the continental sink for atmospheric water. *Geophys. Res. Lett.*, **45**, 4068–4076.
- Kollet, S. J. and Maxwell, R. M. (2006) Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Adv. Water Resour.*, **29**, 945–958.
- Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V. (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, **7**, 1297–1333. URL: <http://doi.org/10.5194/gmd-7-1297-2014>.
- Kuffour, B. N. O., Engdahl, N. B., Woodward, C. S., Condon, L. E., Kollet, S. and Maxwell, R. M. (2020) Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. *Geosci. Model Dev.*, **13**, 1373–1397. URL: <https://gmd.copernicus.org/articles/13/1373/2020/>.
- Lebeaupin, B. C., Bastin, S., Béranger, K. and Drobinski, K. (2015) Regional mesoscale air–sea coupling impacts and extreme meteorological events role on the Mediterranean Sea water budget. *Climate Dyn.*, **44**, 1029–1051.
- Liu, Y. L., Alexander, L. V., Evans, J. P. and Thatcher, M. (2024) Sensitivity of Australian rainfall to driving SST data sets in a variable-resolution global atmospheric model. *J. Geophys. Res.: Atmos.*, **129**, 15.
- Lucas-Picher, P., Argüeso, D., Brisson, E., Trambay, Y., Berg, P., Lemonsu, A., Kotlarski, S. and Caillaud, C. (2021) Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, **12**, e731. URL: <https://doi.org/10.1002/wcc.731>.
- Lucia, C. D., Bucchignani, E., Mastellone, A., Adinolfi, M., Montesarchio, M., Cinquegrana, D., Mercogliano, P. and Schiano, P. (2022) A sensitivity study on high resolution nwp icon–lam model over italy. *Atmosphere*, **13**.
- Luo, B. and Minnett, P. J. (2020) Evaluation of the ERA5 sea surface skin temperature with remotely-sensed shipborne marine-atmospheric emitted radiance interferometer data. *Remote Sens.*, **12**, 1873.
- Mahmood, R., Pielke Sr., R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., McAlpine, C., Carleton, A. M., Hale, R., Gameda, S., Beltrán-Przekurat, A., Baker, B., McNider, R., Legates, D. R., Shepherd, M., Du, J., Blanken, P. D., Frauenfeld, O. W., Nair, U. and Fall, S. (2014) Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, **34**, 929–953. URL: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3736>.
- Mangan, M. R., Hartogensis, O., Boone, A., Branch, O., Canut, G., Cuxart, J., de Boer, H. J., Page, M. L., Martínez-Villagrasa, D., Miró, J. R., Price, J. and de Arellano, J. V.-G. (2023) The surface-boundary layer connection across spatial scales of irrigation-driven thermal heterogeneity: An integrated data and modeling study of the liaise field campaign. *Agricultural and Forest Meteorology*, **335**.
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R. and Zhou, B. (eds.) (2021) *Climate Change 2021: The Physical Science Basis*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. URL: <https://www.ipcc.ch/report/ar6/wg1/>. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
- Maxwell, R. M. (2013) A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Adv. Water Resour.*, **53**, 109–117.

- McDermid, S., Nocco, M., Lawston-Parker, P., Keune, J., Pokhrel, Y., Jain, M., Jägermeyr, J., Brocca, L., Massari, C., Jones, A. D., Vahmani, P., Thiery, W., Yao, Y., Bell, A., Chen, L., Dorigo, W., Hanasaki, N., Jasechko, S., Lo, M. H., Mahmood, R., Mishra, V., Mueller, N. D., Niyogi, D., Rabin, S. S., Sloat, L., Wada, Y., Zappa, L., Chen, F., Cook, B. I., Kim, H., Lombardozzi, D., Polcher, J., Ryu, D., Santanello, J., Satoh, Y., Seneviratne, S., Singh, D. and Yokohata, T. (2023) Irrigation in the earth system. *Nature Reviews Earth and Environment*, **4**, 435–453.
- Minnett, P. J., Alvera-Azcárate, A., Chin, T. M., Corlett, G. K., Gentemann, C. L., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., Santoleri, R., Picart, S. S., Steele, M. and Vazquez-Cuervo, J. (2019) Half a century of satellite remote sensing of sea-surface temperature. *Remote Sens. Environ.*, **233**, 111366.
- Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C. and Haghdoust, S. (2025) Gleam4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present. *Scientific Data*, **12**.
- Naz, B. S., Sharples, W., Ma, Y., Goergen, K. and Kollet, S. (2022) Continental-scale evaluation of a fully distributed coupled land surface and groundwater model ParFlow-CLM (v3.6.0) over Europe. *Geosci. Model Dev. Discuss.*, 1–29.
- Oleson, K. W., Niu, G.-Y., Yang, Z.-L. and et al. (2008) Improvements to the community land model and their impact on the hydrological cycle. *J. Geophys. Res. Biogeosci.*, **113**.
- Pham, T. V., Steger, C., Rockel, B., Keuler, K., Kirchner, I., Mertens, M., Rieger, D., Zängl, G. and Früüh, B. (2021) Icon in climate limited-area mode (ICON release version 2.6.1): A new regional climate model. *Geosci. Model Dev.*, **14**, 985–1005.
- Pielke Sr., R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., Goldewijk, K. K., Nair, U., Betts, R., Fall, S., Reichstein, M., Kabat, P. and de Noblet, N. (2011) Land use/land cover changes and climate: modeling analysis and observational evidence. *WIREs Climate Change*, **2**, 828–850. URL: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.144>.
- Poshyvailo-Strube, L., Wagner, N., Goergen, K., Furusho-Percot, C., Hartick, C. and Kollet, S. (2024) Impact of groundwater representation on heat events in regional climate simulations over europe. *Earth System Dynamics*, **15**, 167–189.
- Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., Fox Maule, C., van Meijgaard, E., Déqué, M., Nikulin, G., Vautard, R., Colette, A., Kjellström, E. and Jacob, D. (2016) Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: high resolution, high benefits? *Climate Dynamics*, **46**, 383–412. URL: <https://doi.org/10.1007/s00382-015-2589-y>.
- Prein, A. F., Langhans, W., Fossier, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tille, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M. and Leung, R. (2015) A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, **53**, 323–361. URL: <http://doi.org/10.1002/2014RG000475>.
- Raschendorfer, M. (2001) The new turbulence parameterization of LM. COSMO Newsletter, No. 1.
- Rieger, D., Boucouvala, D., Gofa, F., Kolyvas, C., Khain, P., Muskatel, H., Shtivelman, A., Baharad, A., Uzan, L., Levi, Y., Marcucci, F., Sudati, F., Zaccariello, N., Garbero, V., Vela, N., Oberto, E., Minguzzi, E., Cesari, D., Gastaldo, T., Poli, V., Marsigli, C., Tesini, M. S., Cerenzia, I., Linkowska, J., Interewicz, W., Wójcik, D., Iriza-Burcă, A., Maco, B. A., Dragomir, R., Ş Gabrian, Ş Dinicilă, Bogdan, M., Bălăcescu, T., Dumitrache, R. C., Bonatti, G. R., Santos, R. R. D., Kitagawa, Y. K. L. and Silveira, R. B. D. (2022) Consortium for small-scale modelling the priority project c2i, transition of cosmo to icon-final report the priority project c2i, transition of cosmo to icon-final report. *Tech. rep.* URL: www.cosmo-model.org.
- Rockel, B. and Geyer, B. (2022) SPICE (starter package for ICON-CLM experiments). Zenodo.
- Scanlon, T., Geer, A., Bormann, N. and Browne, P. (2024) Improving ocean surface temperature for NWP using all-sky microwave imager observations. *EUMETSAT/ECMWF Fellowship Programme Research Reports RR64*, European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, UK.

- Schneider, U., Hänsel, S., Finger, P., Rustemeier, E. and Ziese, M. (2022) GPCC full data monthly product version 2022 at 0.25°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historical data. URL: http://dx.doi.org/10.5676/DWD_GPCC/FD_M_V2022_025.
- Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stefan, K., Períáñez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the cosmo model (kenda). *Q.J.R. Meteorol. Soc.*, **142**, 1453–1472.
- Schrodin, R. and Heise, E. (2001) The multi-layer version of the DWD soil model TERRA LM. URL: <http://www.cosmo-model.org>. Tech Rep 2, Consortium for Small-Scale Modelling.
- Schulz, J. P., Vogel, G., Becker, C., Kothe, S., Rummel, U. and Ahrens, B. (2016) Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil. *Meteor. Z.*, **25**, 607–620.
- Schulzweida, U. (2023) Cdo user guide. URL: <https://doi.org/10.5281/zenodo.10020800>.
- Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpillot, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried, K., Panosetti, D., Rüdüsühli, S., Schlemmer, L., Schulthess, T., Sprenger, M., Ubbiali, S. and Wernli, H. (2020) Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society*, **101**, E567–E587. URL: <https://doi.org/10.1175/BAMS-D-18-0167.1>.
- Seifert, A. (2008) A revised cloud microphysical parameterization for COSMO-LME. URL: <http://www.cosmo-model.org>. COSMO News Letter No. 7.
- Seneviratne, S. I., Wartenburger, R., Guillod, B. P., Hirsch, A. L., Vogel, M. M., Brovkin, V., van Vuuren, D. P., Schaller, N., Boysen, L., Calvin, K. V., Doelman, J., Greve, P., Havlik, P., Humpenöder, F., Krisztin, T., Mitchell, D., Popp, A., Riahi, K., Rogelj, J., Schleussner, C.-F., Sillmann, J. and Stehfest, E. (2018) Climate extremes, land–climate feedbacks and land-use forcing at 1.5°C. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **376**, 20160450. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2016.0450>.
- Seo, H., O'Neill, L. W., Bourassa, M. A., Czaja, A., Drushka, K., Edson, J. B., Fox-Kemper, B., Frenger, I., Gille, S. T., Kirtman, B. P., Minobe, S., Pendergrass, A. G., Renault, L., Roberts, M. J., Schneider, N., Small, R. J., Stoffelen, A. and Wang, Q. (2023) Ocean mesoscale and frontal-scale ocean-atmosphere interactions and influence on large-scale climate: A review. *J. Climate*, **36**, 1981–2013.
- Shrestha, P., Sulis, M., Masbou, M., Kollet, S. and Simmer, C. (2014) A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. *Mon. Weather Rev.*, **142**, 3466–3483.
- Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P. and Portmann, F. T. (2010) Groundwater use for irrigation – a global inventory. *Hydrology and Earth System Sciences*, **14**, 1863–1880. URL: <https://hess.copernicus.org/articles/14/1863/2010/>.
- Siebert, S. and Döll, P. (2010) Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *J. Hydrol.*, **384**, 198–217.
- Siegismund, F. and Kusche, J. (2024) Collaborative research centre 1502 detect: 'regional climate change: Disentangling the role of land use and water management'. URL: <https://doi.org/10.5194/egusphere-egu24-20232>. EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-20232.
- da Silva Lopes, F., Schindelegger, M., Gutknecht, B. and Kusche, J. (2025) Dynamically downscaled european water budget quantities in the presence of sea surface temperature uncertainty. ESS Open Archive [Preprint]. URL: <https://doi.org/10.22541/essoar.173888329.93971180/v1>. [cited 2025 Nov 03].
- Silwimba, K., Flores, A. N., Cionni, I., Billings, S. A., Sullivan, P. L., Ajami, H., Hirmas, D. R. and Li, L. (2025) Soil parameterization in land surface models drives large discrepancies in soil moisture predictions across hydrologically complex regions of the contiguous united states. *Geoscientific Model Development*, **18**, 7707–7734. URL: <https://gmd.copernicus.org/articles/18/7707/2025/>.

- Tapiador, F. J., Roca, R., Genio, A. D., Dewitte, B., Petersen, W. and Zhang, F. (2019) Is precipitation a good metric for model performance? *Bulletin of the American Meteorological Society*, **100**, 223–233. URL: <https://journals.ametsoc.org/view/journals/bams/100/2/bams-d-17-0218.1.xml>.
- Taylor, K. E. (2024) Truly conserving with conservative remapping methods. *Geoscientific Model Development*, **17**, 415–430.
- Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Weather Rev.*, **117**, 1779–1800.
- Titchner, H. A. and Rayner, N. A. (2014) The met office hadley centre sea ice and sea surface temperature data set, version 2: HadISST.2.0.0.0. *Climate Dyn.*, **42**, 2097–2118.
- Tuinenburg, O. A. and de Vries, J. P. (2017) Irrigation patterns resemble era-interim reanalysis soil moisture additions. *Geophysical Research Letters*, **44**, 10,341–10,348.
- Valcke, S. (2013) The OASIS3 coupler: A european climate modelling community software. *Geosci. Model Dev.*, **6**, 373–388.
- Valmassoi, A., Keller, J. D., Potthast, R., Anlauf, H. and A. Cress, A. (2024) Towards a new generation of regional reanalyses for europe. In *EMS Annual Meeting 202*. Barcelona, Spain. URL: <https://doi.org/10.5194/ems2024-27>. EMS2024-27.
- Valmassoi, A. and Keller, J. D. (2022) A review on irrigation parameterizations in earth system models. *Frontiers in Water*, **4**.
- Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S. and Bierkens, M. F. P. (2010) Global depletion of groundwater resources. *Geophysical Research Letters*, **37**. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL044571>.
- Wada, Y. and Bierkens, M. F. P. (2014) Sustainability of global water use: past reconstruction and future projections. *Environmental Research Letters*, **9**, 104003. URL: <https://dx.doi.org/10.1088/1748-9326/9/10/104003>.
- Wada, Y., de Graaf, I. E. M. and van Beek, L. P. H. (2016) High-resolution modeling of human and climate impacts on global water resources. *J. Adv. Model. Earth Syst.*, **8**, 735–763.
- Winkler, K., Yang, H., Ganzenmüller, R., Fuchs, R., Ceccherini, G., Duveiller, G., Grassi, G., Pongratz, J., Bastos, A., Shvidenko, A., Araza, A., Herold, M., Wigneron, J.-P. and Ciais, P. (2023) Changes in land use and management led to a decline in Eastern Europe's terrestrial carbon sink. *Communications Earth & Environment*, **4**, 1–14. URL: <https://www.nature.com/articles/s43247-023-00893-4>.
- Yang, W., Xue, Y., Li, S., Zhao, M., Li, J., Liu, Y., Liu, Y., Douville, H., Jr., W. J. G., Schiemann, R., Ruiz, J. A. and Giorgi, F. (2021) Impacts of different SST datasets on the simulated regional climate in CORDEX over East Asia. *Climate Dyn.*, **57**, 1029–1051.
- Zaitchik, B. F., Rodell, M., Biasutti, M. and Seneviratne, S. I. (2023) Wetting and drying trends under climate change. *Nature Water*, **1**, 502–513. URL: <https://doi.org/10.1038/s44221-023-00073-w>.
- Zeng, X., Wang, Y. and Dai, A. (2005) A Satellite-based Daily SST Climate Dataset for Climate Models. *J. Climate*, **18**, 5189–5204.
- Zhang, Y., Wagner, N., Goergen, K. and Kollet, S. (2024) Summer evapotranspiration-cloud feedbacks in land-atmosphere interactions over Europe. *Climate Dynamics*, **62**, 10767–10783. URL: <https://doi.org/10.1007/s00382-024-07475-w>.
- Zhou, T., Wu, L., Zhou, B., Zhang, C., Wang, W., Dong, B., Xue, W., Chen, X., Weber, S. L., Wang, Z., Chen, J. and Zou, L. (2019) The effect of Pacific SST anomalies on the water cycle in East Asia: A review. *Adv. Atmos. Sci.*, **36**, 1209–1223.
- Zhuang, J., raphael dussin, Huard, D., Bourgault, P., Banihirwe, A., Raynaud, S., Malevich, B., Schupfner, M., Filipe, Gauthier, C., Levang, S., Jüling, A., Almansi, M., RichardScottOZ, RondeauG, Rasp, S., Smith, T. J., Mares, B., Stachelek, J., Plough, M., Pierre, Bell, R., Caneill, R. and Li, X. (2025) pangeo-data/xesmf: v0.8.10. URL: <https://doi.org/10.5281/zenodo.15304267>.

Zipper, S. C., Keune, J. and Kollet, S. J. (2019) Land use change impacts on European heat and drought: Remote land-atmosphere feedbacks mitigated locally by shallow groundwater. *Environ. Res. Lett.*, **14**, 044012.

Zängl, G., Reinert, D., Rípodas, P. and Baldauf, M. (2015) The ICON (icosahedral nonhydrostatic) modelling framework of DWD and MPI-M: Description of the nonhydrostatic dynamical core. *Q. J. R. Meteorol. Soc.*, **141**, 563–579.

A | APPENDIX

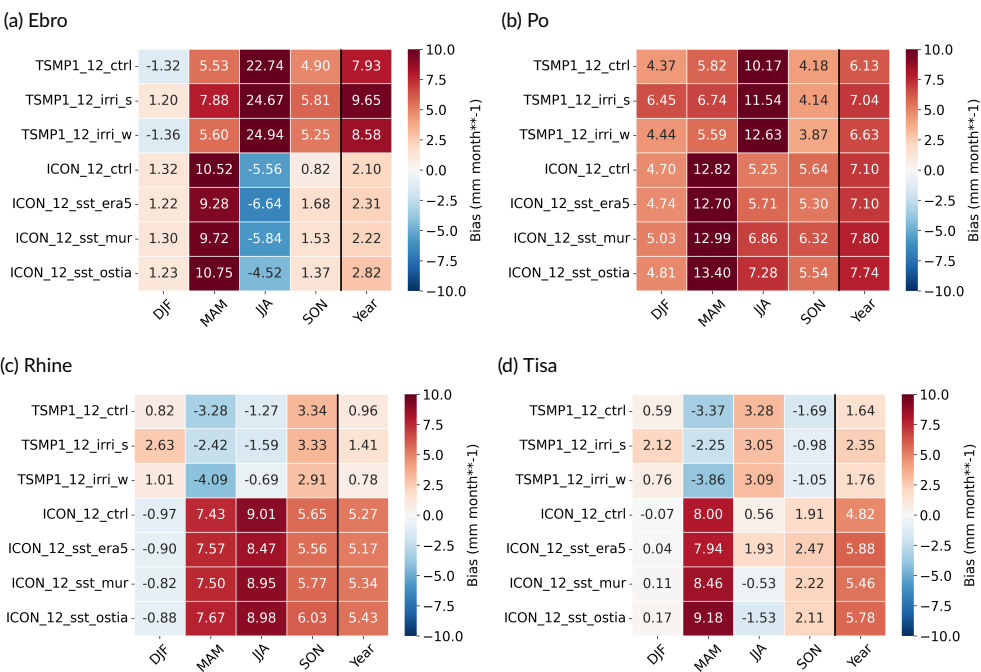


FIGURE S1 Mean seasonal ET biases for different watersheds for the period 2003-2010. The column "Year", shows the yearly MAE for the whole period.

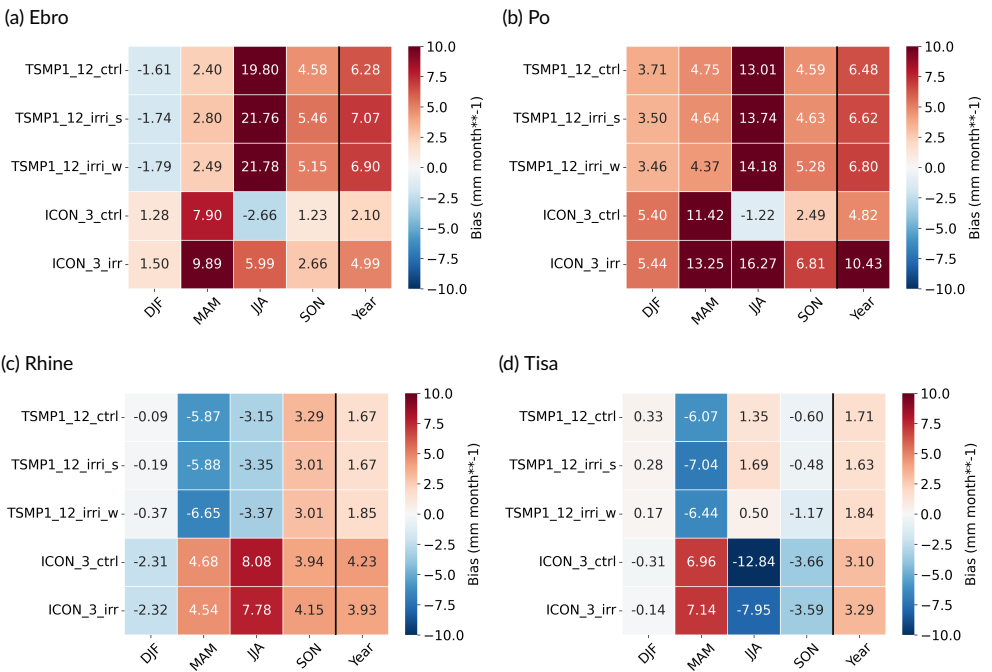


FIGURE S2 Mean seasonal ET biases for different watersheds for the period 2011-2020. The column "Year", shows the yearly MAE for the whole period.

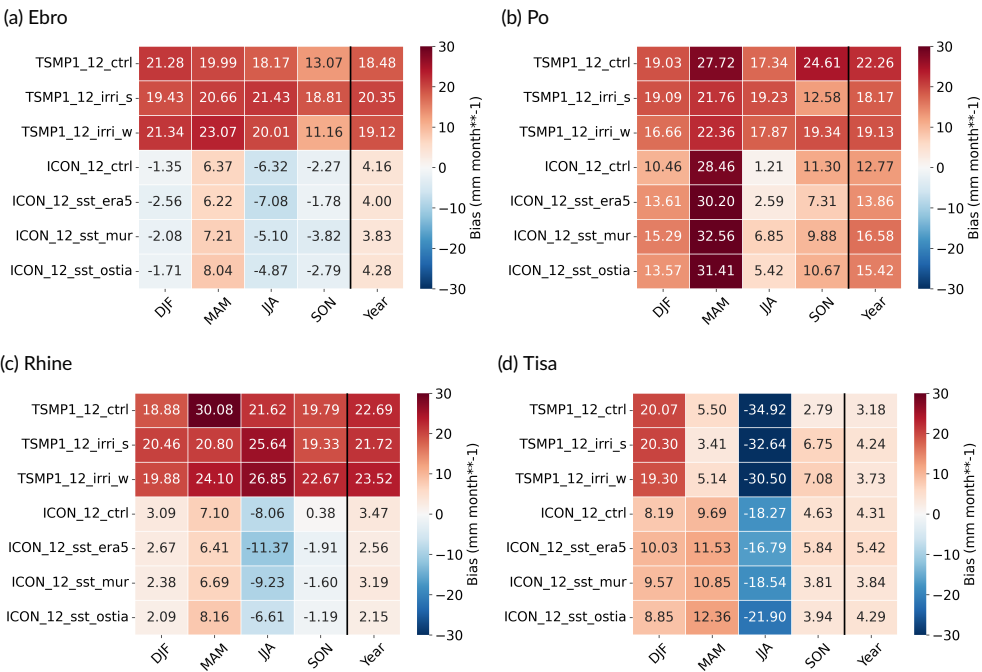


FIGURE S3 Mean seasonal precipitation biases for different watersheds for the period 2003-2010. The column "Year", shows the yearly MAE for the whole period.

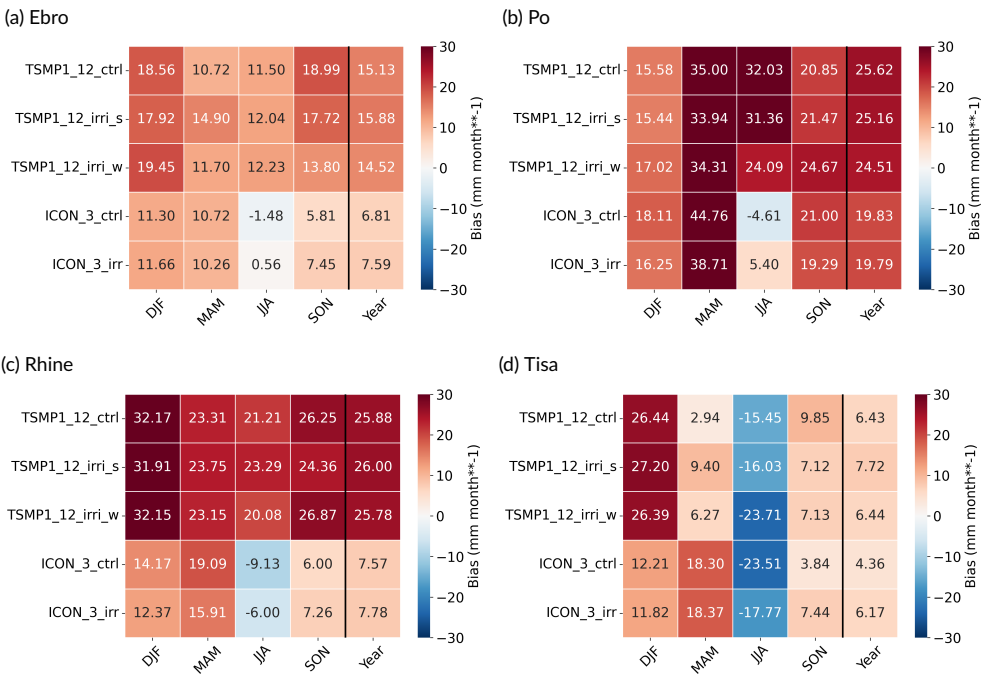


FIGURE S4 Mean seasonal precipitation biases for different watersheds for the period 2011-2020. The column "Year", shows the yearly MAE for the whole period.

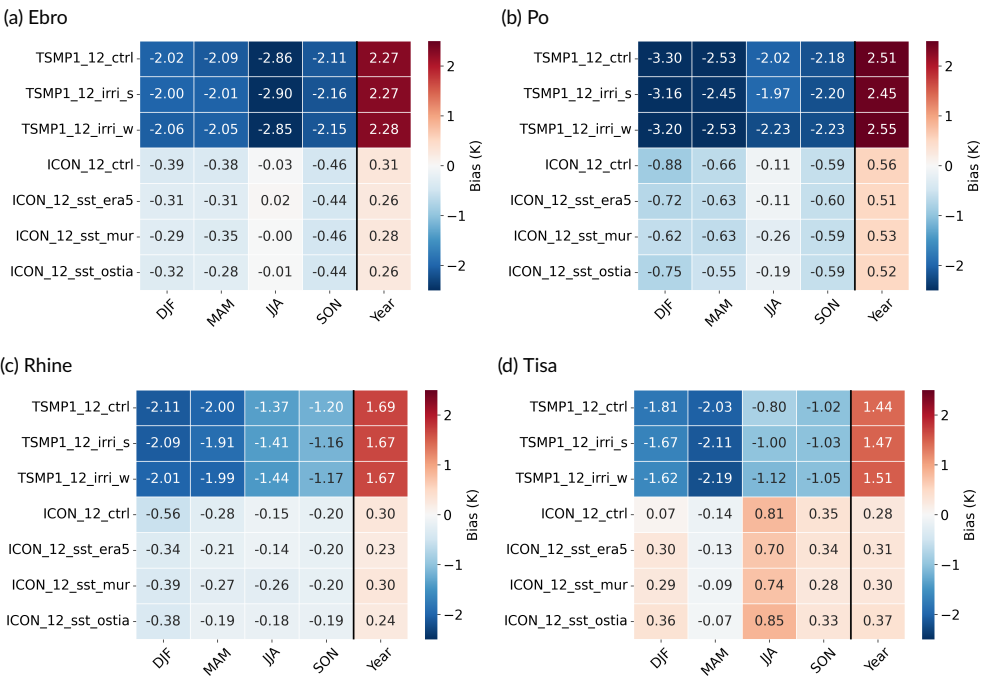


FIGURE S5 Mean seasonal T2m for different watersheds for the period 2003-2010. The column "Year", shows the yearly MAE for the whole period.

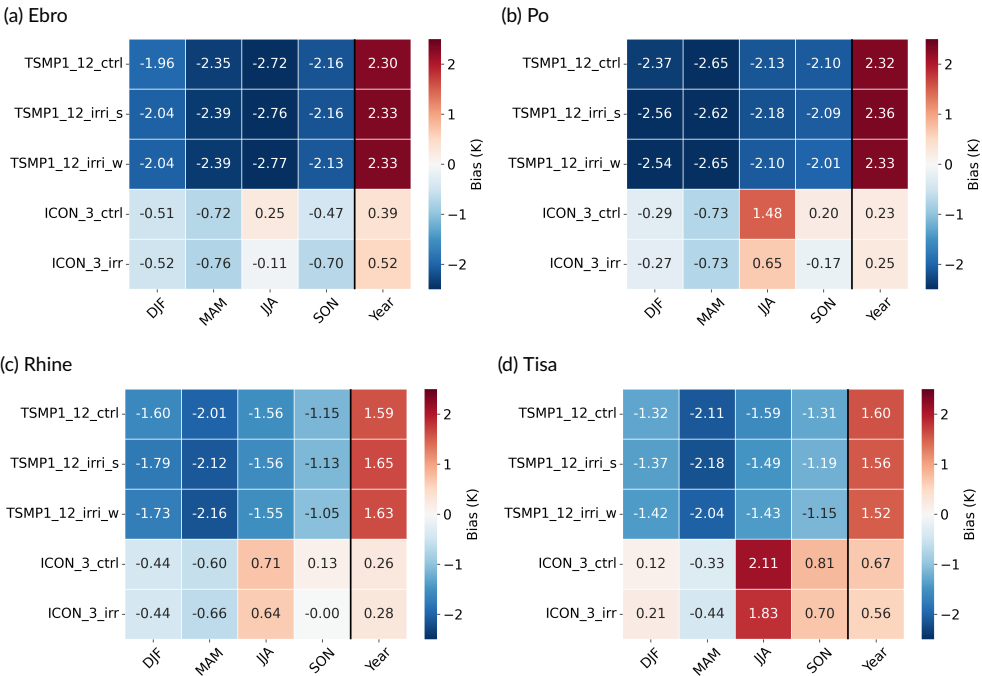
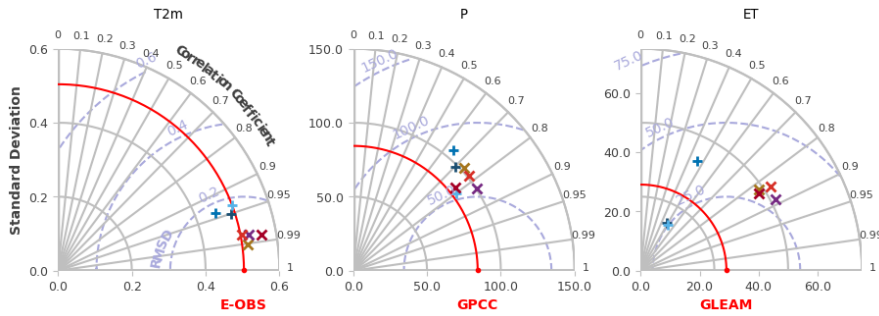
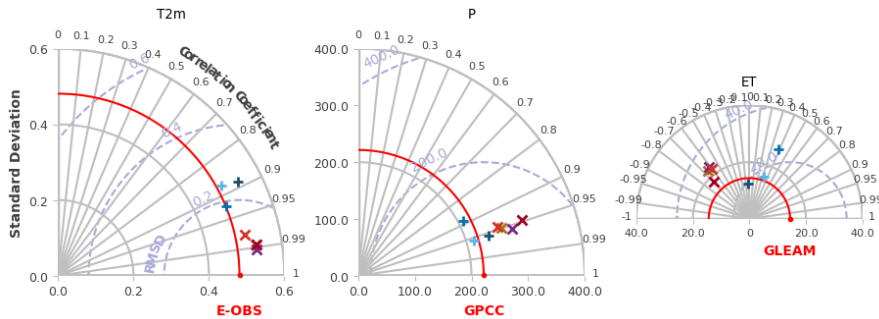


FIGURE S6 Mean seasonal T2m for different watersheds for the period 2011-2020. The column "Year", shows the yearly MAE for the whole period.

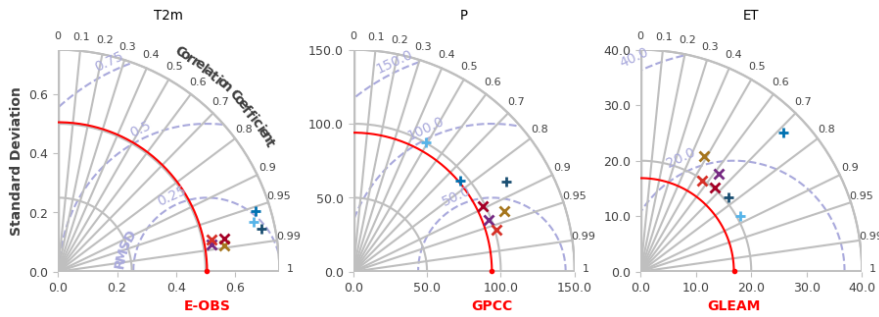
(a) Ebro



(b) Po



(c) Rhine



(d) Tisa

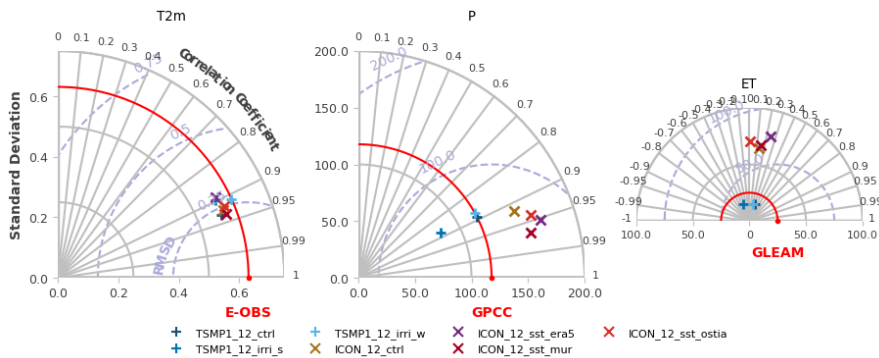
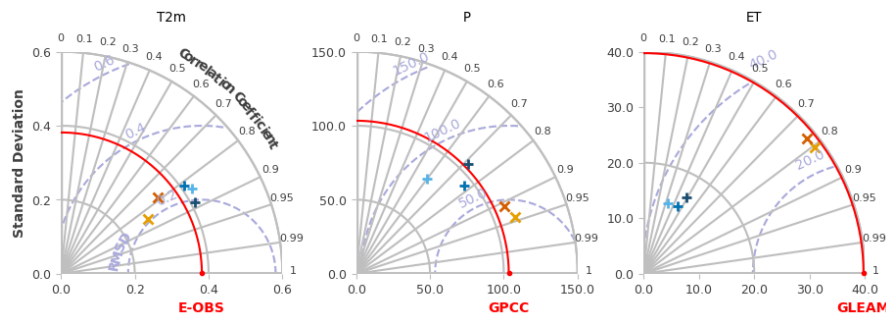
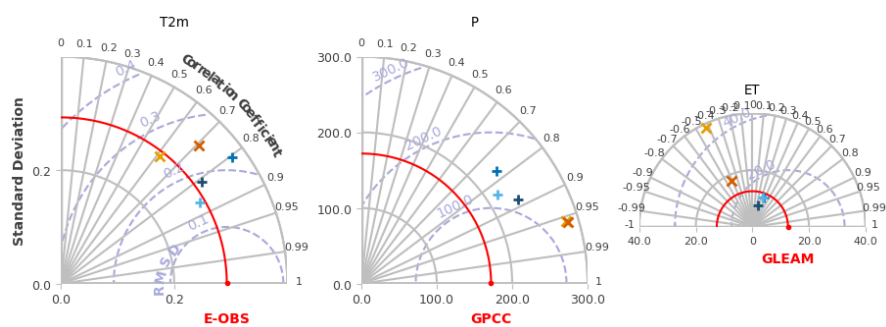


FIGURE S7 Taylor diagrams comparing simulation results to reference datasets for 2 m temperature (T2m), precipitation (P) and evapotranspiration (ET), over the period 2003–2010, for each basin. The simulations are compared against E-OBS (for T2m), GPCC (for P) and GLEAM (for ET)

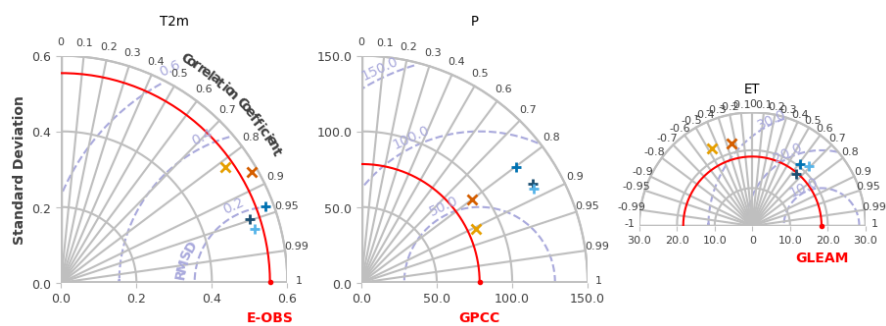
(a) Ebro



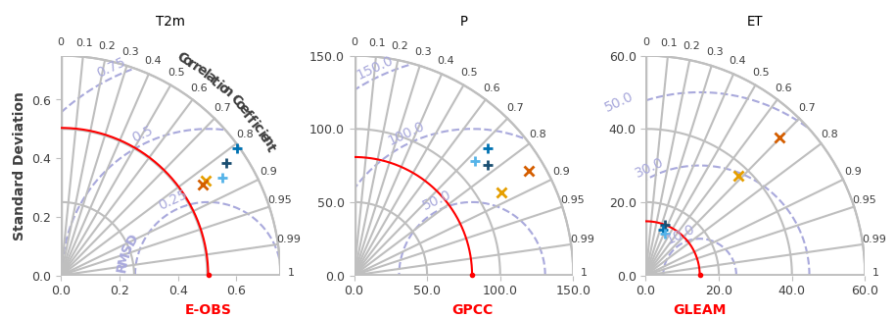
(b) Po



(c) Rhine



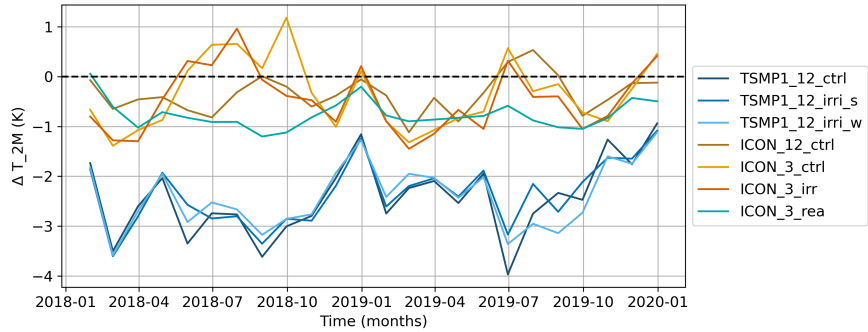
(d) Tisa



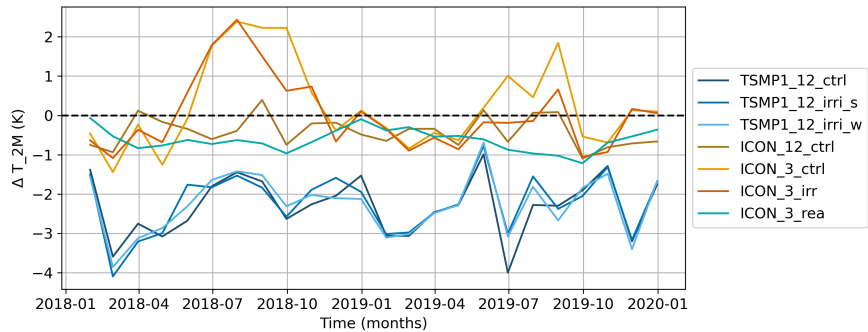
+ TSMP1_12_ctrl + TSMP1_12_irri_s + TSMP1_12_irri_w x ICON_3_ctrl x ICON_3_irri

FIGURE S8 Taylor diagrams comparing simulation results to reference datasets for 2 m temperature (T2m), precipitation (P) and evapotranspiration (ET), over the period 2011–2020, for each basin. The simulations are compared against E-OBS (for T2m), GPCC (for P) and GLEAM (for ET)

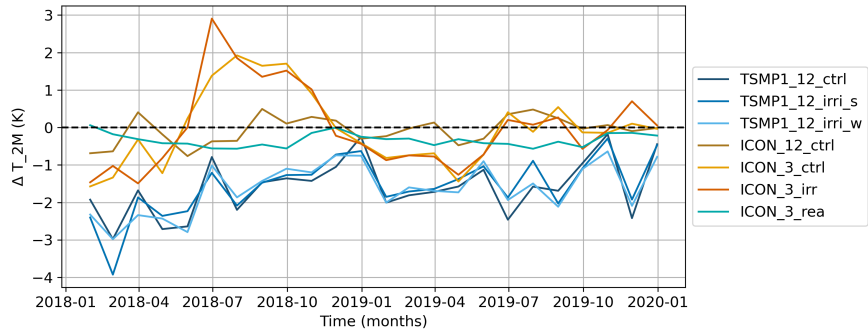
(a) Ebro



(b) Po



(c) Rhine



(d) Tisa

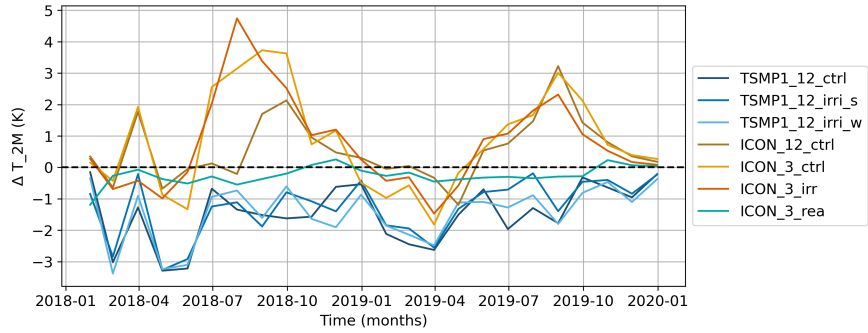
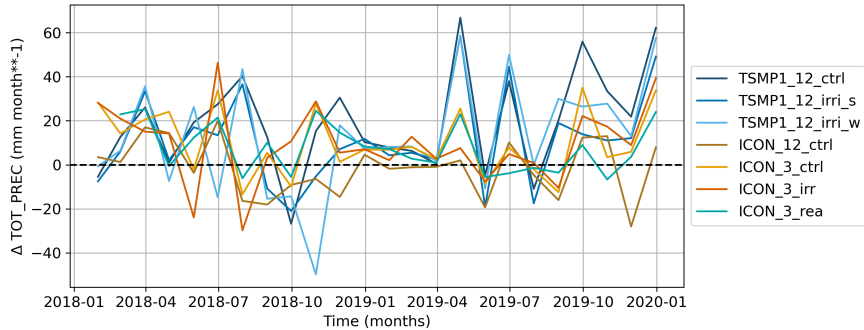
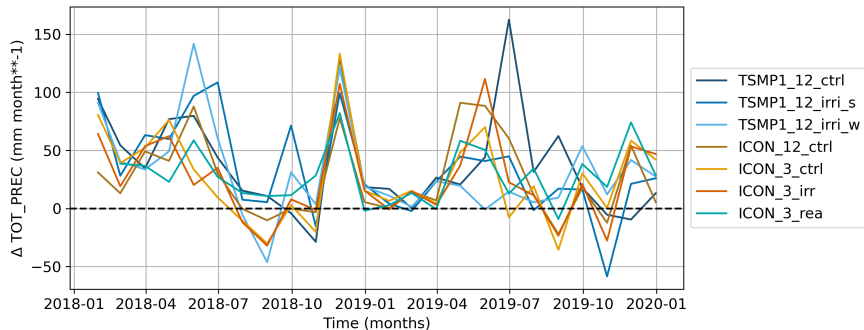


FIGURE S9 Timeseries of monthly T2m biases (Sim. - E-OBS) for representative watersheds.

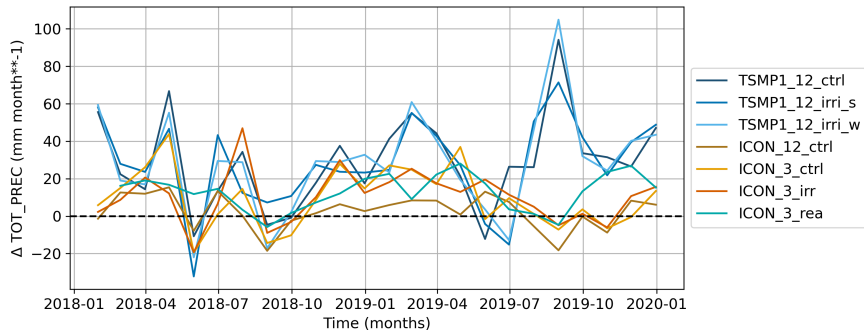
(a) Ebro



(b) Po



(c) Rhine



(d) Tisa

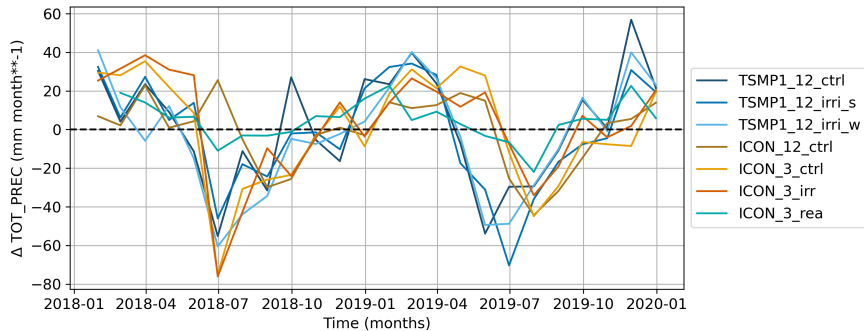
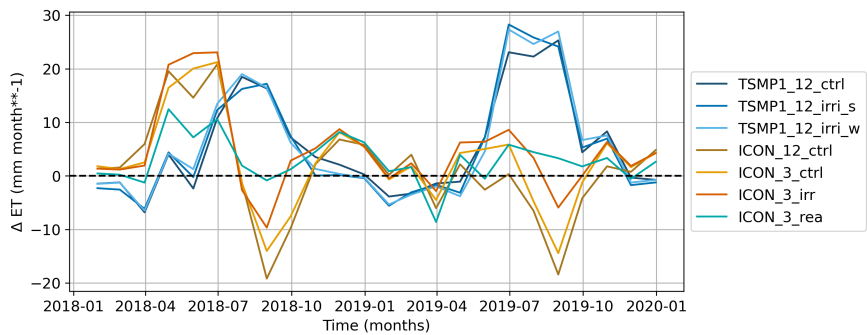
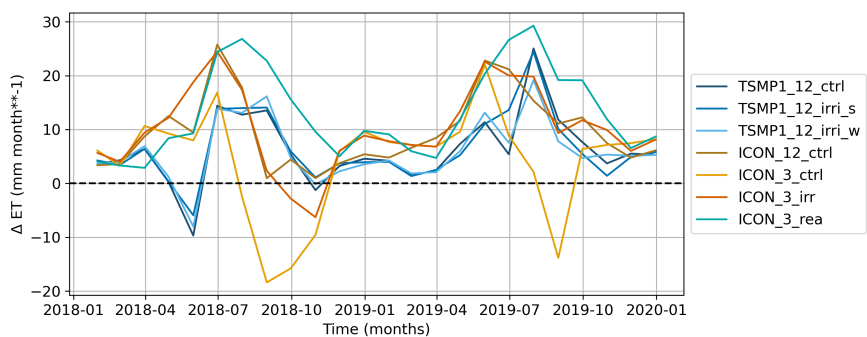


FIGURE S10 Timeseries of monthly precipitation biases (Sim. - GPCC) for the representative watersheds.

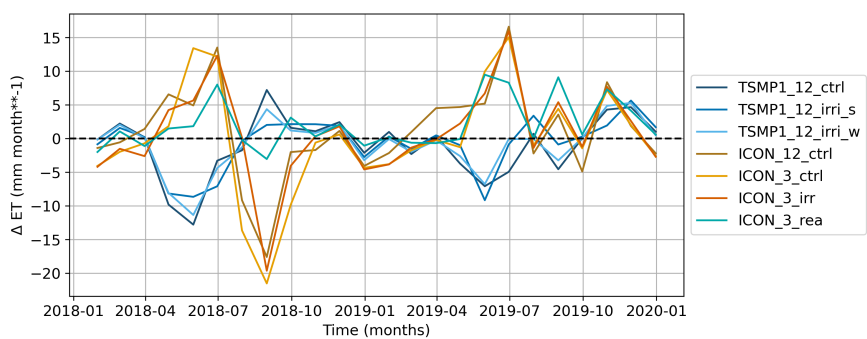
(a) Ebro



(b) Po



(c) Rhine



(d) Tisa

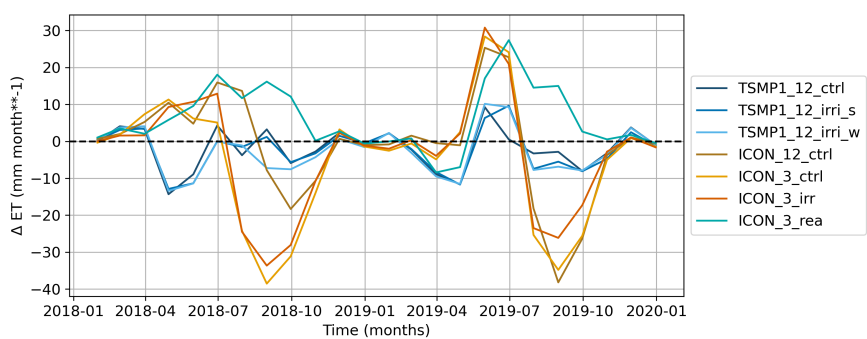


FIGURE S11 Timeseries of monthly evapotranspiration biases (Sim. - GLEAM) for the representative watersheds.

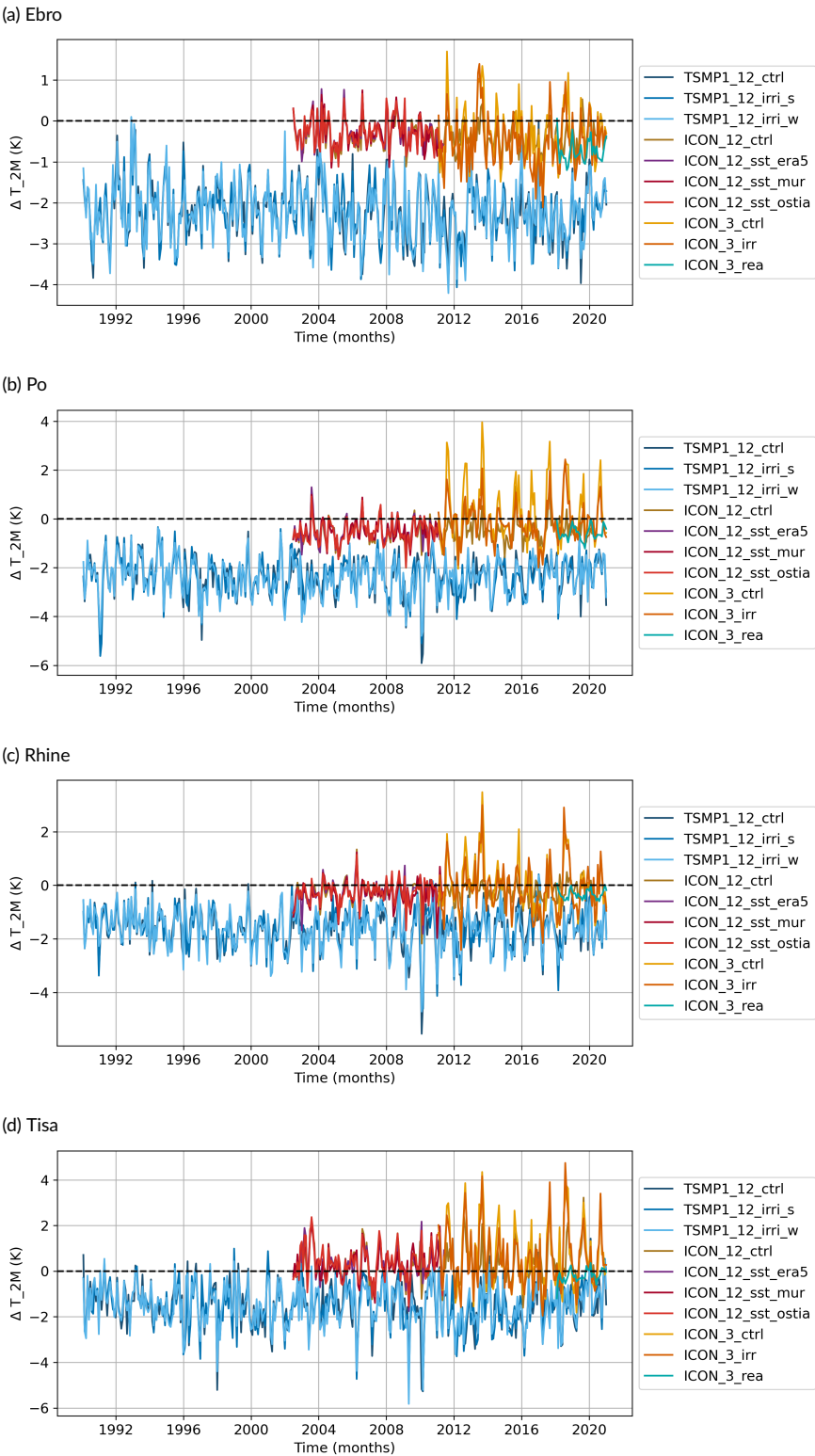


FIGURE S12 Timeseries of monthly T2m biases (Sim. - E-OBS) for the representative watersheds.

TABLE S1 Evaluation metrics for T2m (2018–2020).

Simulations	Ebro			Po			Rhein			Tisa		
	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE
TSMPl_12_ctrl	-2.43	0.75	2.55	-2.31	0.76	2.43	-1.60	0.74	1.77	-1.48	0.92	1.74
TSMPl_12_irri_s	-2.36	0.63	2.44	-2.23	0.76	2.35	-1.57	0.76	1.74	-1.28	0.88	1.55
TSMPl_12_irri_w	-2.40	0.64	2.49	-2.27	0.74	2.39	-1.63	0.66	1.76	-1.45	0.87	1.69
ICON_12_ctrl	-0.35	0.37	0.51	-0.42	0.38	0.56	-0.07	0.36	0.37	0.56	0.98	1.13
ICON_3_ctrl	-0.36	0.69	0.78	0.24	1.09	1.12	-0.04	0.98	0.98	0.85	1.56	1.78
ICON_3_irr	-0.47	0.62	0.78	0.003	0.91	0.91	-0.02	1.09	1.09	0.76	1.44	1.63
ICON_3_rea	-0.76	0.28	0.81	-0.63	0.27	0.68	-0.34	0.18	0.38	-0.25	0.29	0.38

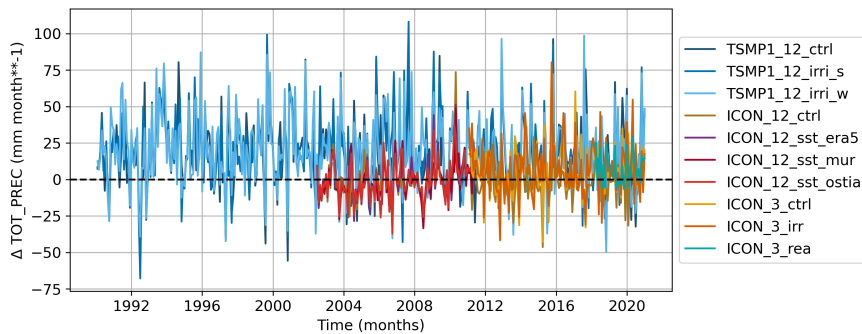
TABLE S2 Evaluation metrics for P (2018–2020).

Simulations	Ebro			Po			Rhein			Tisa		
	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE
TSMP1_12_ctrl	19.19	22.43	29.52	37.82	44.12	58.11	29.09	24.17	37.82	1.67	28.10	28.15
TSMP1_12_irri_s	11.12	20.63	23.43	34.24	39.99	52.65	28.18	23.45	36.67	-1.44	27.24	27.28
TSMP1_12_irri_w	12.72	25.49	28.48	31.61	39.94	50.93	28.68	26.93	39.34	-3.28	28.72	28.91
ICON_12_ctrl	-1.42	12.52	12.60	26.61	33.61	42.87	2.49	9.08	9.42	-1.01	18.19	18.22
ICON_3_ctrl	10.14	14.68	17.84	26.46	37.85	46.19	9.68	15.69	18.43	0.44	28.03	28.03
ICON_3_irr	9.45	17.06	19.50	25.44	36.43	44.43	10.19	13.56	16.96	2.69	26.65	26.79
ICON_3_rea	7.69	11.11	13.51	28.31	23.61	36.86	12.84	9.35	15.88	4.62	10.24	11.23

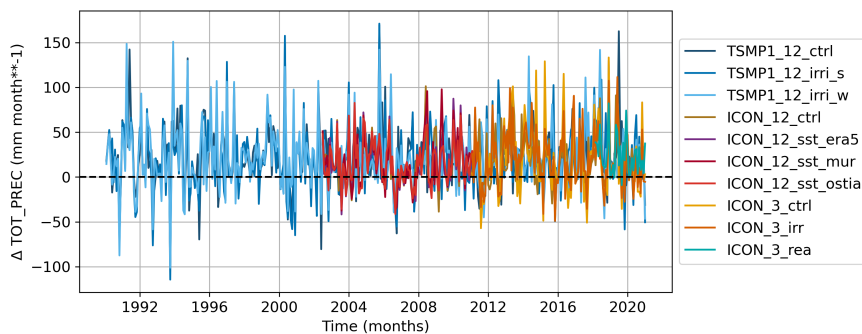
TABLE S3 Evaluation metrics for ET (2018–2020).

Simulations	Ebro			Po			Rhein			Tisa		
	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE	Bias	STD	RMSE
TSMPl_12_ctrl	5.45	9.11	10.62	6.20	6.44	8.94	-1.07	4.45	4.57	-1.86	5.55	5.86
TSMPl_12_irri_s	5.31	9.99	11.31	6.46	6.14	8.91	-0.73	3.80	3.86	-2.25	5.67	6.10
TSMPl_12_irri_w	5.58	10.08	11.52	5.98	5.68	8.24	-0.93	3.83	3.94	-2.64	6.22	6.76
ICON_12_ctrl	1.04	9.31	9.37	9.55	6.67	11.65	1.06	6.82	6.90	-0.96	14.08	14.11
ICON_3_ctrl	2.49	8.50	8.86	4.29	9.54	10.47	-0.09	8.03	8.03	-4.95	16.92	17.63
ICON_3_irr	4.64	7.89	9.16	10.11	7.41	12.54	0.74	6.67	6.71	-3.40	15.40	15.77
ICON_3_rea	2.88	4.25	5.14	13.13	8.15	15.46	2.04	3.62	4.16	6.09	8.51	10.47

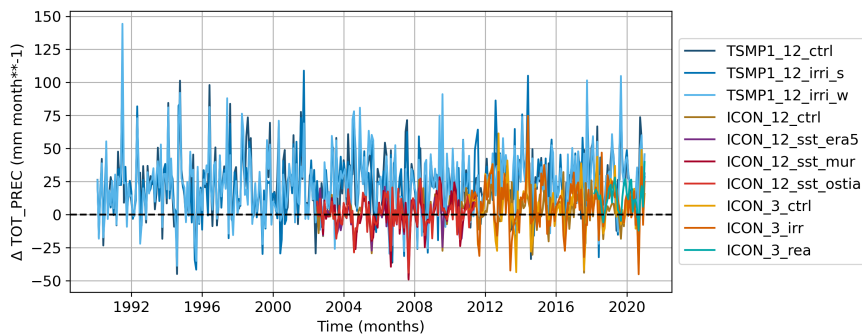
(a) Ebro



(b) Po



(c) Rhine



(d) Tisa

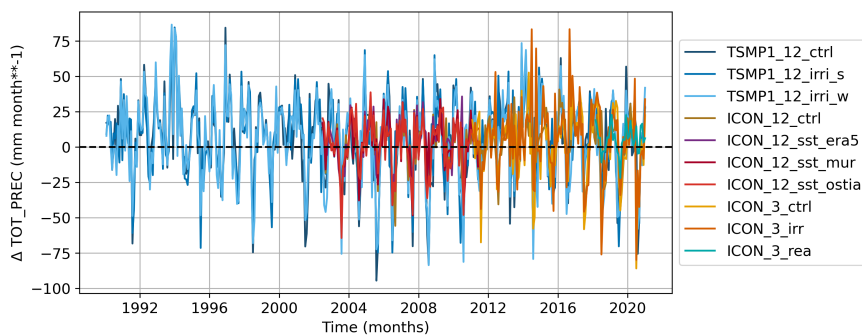
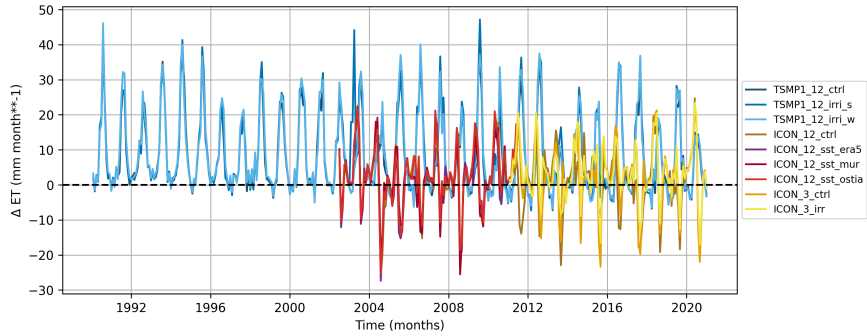
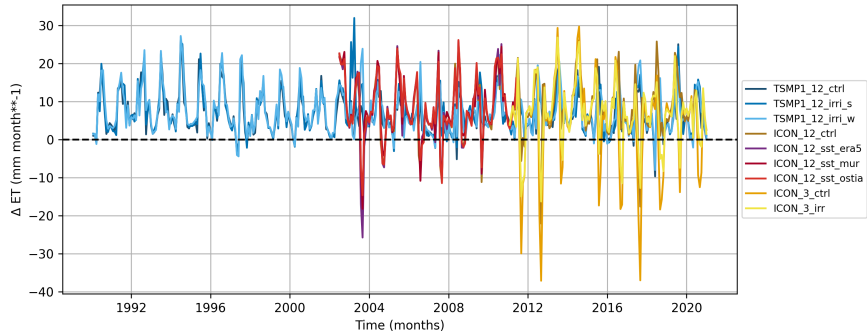


FIGURE S13 Timeseries of monthly precipitation biases (Sim. - GPCC) for the representative watersheds.

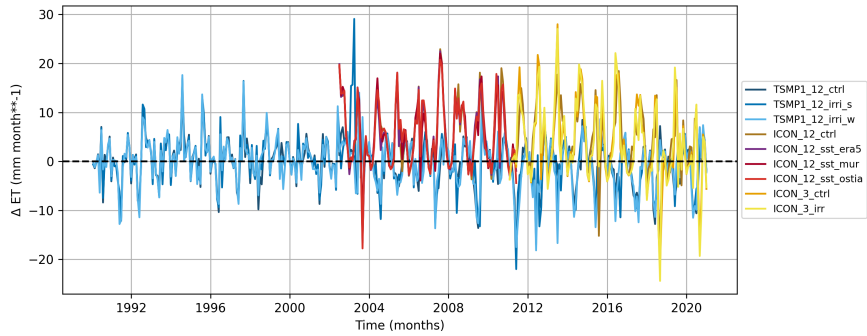
(a) Ebro



(b) Po



(c) Rhine



(d) Tisa

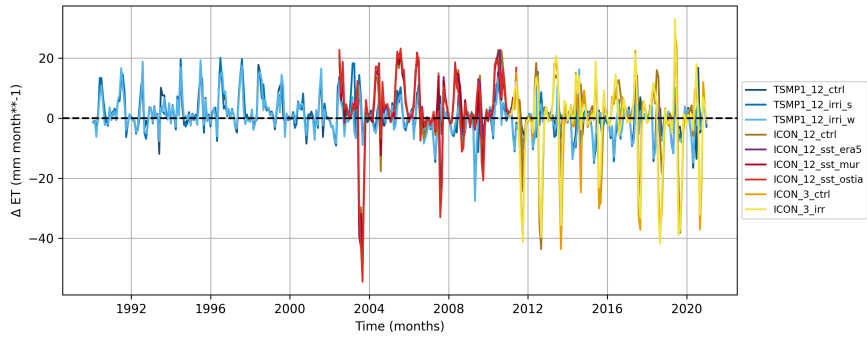


FIGURE S14 Timeseries of monthly ET biases (Sim. - GLEAM) for the representative watersheds.