

MAJOR COMMENTS

The manuscript presents a comprehensive dataset derived from regional climate downscaling addressed to the Greek territory using advanced machine learning techniques. The primary objectives are to enhance the spatial resolution over a 39-year period and to improve the representation of daily temperature and precipitation fields across the complex geography of continental Greece and its islands. The authors employ a hybrid approach combining geostatistical interpolation and statistical downscaling methods with atmospheric modelling, validated against observational datasets. Four methods (FRK, GAM, KNN, and SVM) were evaluated, with FRK ultimately chosen. Evaluation against CHELSA-W5E5 and withheld station data supports the improved spatial accuracy and bias reduction of CLIMADAT-GRid, especially in mountainous regions. The validation strategy, multiple error metrics, and comparison with an established product (CHELSA-W5E5) strengthen the study's robustness. The outcomes suggest improved accuracy in temperature and precipitation at regional scales, supported by the analysis of suitable climate indicators. However, while the objectives are met mainly, certain aspects require further clarification to substantiate the claims thoroughly.

Answer:

We sincerely appreciate the reviewer's thoughtful comments and valuable suggestions. We have carefully considered all the points raised and have addressed each of them thoroughly in the revised manuscript. We believe that these revisions have significantly improved the clarity, rigor, and overall quality of the work. A detailed point-by-point response is provided below, highlighting how each comment was incorporated or clarified.

1) The introduction chapter might discuss similar datasets produced for the same purpose. This would help clarify the expectations surrounding this exercise, including its benefits, drawbacks, and potential challenges. The datasets E-OBS and IBERIA01 referenced by the authors in Chapter 3 could serve as a starting point.

Answer:

Following the reviewer's suggestion, the specific section in the introduction has been modified in the revised manuscript **from:**

“As for the second category, which is the focus of this study, the remarkable advances in computing power and software have led to the development and creation of gridded observational datasets at both global and regional/national levels. In a recent study, Varotsos et al. (2023a) compiled a list of available observational gridded datasets for Europe, with E-OBS (Cornes et al., 2018) being the state-of-art gridded observational dataset for the entire European domain. However, it is crucial for users of gridded observational datasets to recognize that these products are model-generated rather than direct observations, and as such have a number of limitations (Hofstra et al., 2010). In particular, the quality of the gridded datasets depends on the quality of the station data and its spatial coverage (density of the

meteorological station network), as interpolation methods degrade in performance in areas with sparse station data and/or in areas with complex topography (Hofstra et al., 2010; Begueria et al., 2016; Herrera et al., 2019).”

to :

“As for the second category, which is the focus of this study, the remarkable advances in computing power and software have led to the development and creation of gridded observational datasets at both global, regional/national and sub-national levels. These datasets include E-OBS (Cornes et al., 2018) which is the state-of-art daily gridded observational dataset for the entire European domain with a resolution of 0.1° , while on the regional/national and sub-national scale a number of datasets have recently emerged in Europe. These include Iberia01 (Herrera et al., 2019) for the Iberian Peninsula (daily gridded dataset for temperatures and precipitation at 0.1° grid), SPREAD (Serrano-Notivoli et al., 2017) and STEAD (Serrano-Notivoli et al., 2019) for Spain (daily datasets for precipitation and temperatures at $5\text{km} \times 5\text{km}$, respectively), SiCLIMA (Serrano-Notivoli et al., 2024) for Aragon, Spain (daily dataset for precipitation and temperatures at $500\text{m} \times 500\text{m}$), PTHRES (Fonseca and Santos 2018) for Portugal (daily dataset for temperatures at $1\text{km} \times 1\text{km}$), HYRAS (Krähenmann et al., 2018) for Germany (hourly dataset for a number of variables at $1\text{km} \times 1\text{km}$), HadUK-Grid (Hollis et al. 2019) for the United Kingdom (daily dataset for a number of variables at $1\text{km} \times 1\text{km}$), seNorge2 (Lussana et al., 2018a,b) for Norway (daily dataset for precipitation and temperatures at $1\text{km} \times 1\text{km}$, respectively), SLOCLIM (Škrk et al., 2021) for Slovenia (daily dataset for precipitation and temperatures at $1\text{km} \times 1\text{km}$), MeteoSerbia1km (Sekulić et al. 2021) for Serbia (daily dataset for a number of variables at $1\text{km} \times 1\text{km}$) and GAA.HRES (Varotsos et al., 2023a) for Attica, Greece (daily dataset for precipitation and temperatures at $1\text{km} \times 1\text{km}$). It is important for users to recognize that these gridded observational products are geostatistically generated, rather than direct observations. Consequently, they are subject to several limitations and the accuracy of these datasets largely depends on the quality and spatial density of the underlying meteorological station network. In particular, interpolation methods tend to perform poorly in regions with sparse station coverage or complex topography (Hofstra et al., 2010; Beguería et al., 2016; Herrera et al., 2019). While most of these datasets are built upon dense networks of ground-based observations, in areas with limited station density or insufficient representation of elevation gradients it is often required enhancement through the integration of satellite data, reanalysis products, and atmospheric models to improve spatial coverage and reliability (Doblas-Reyes et al., 2021; Varotsos et al., 2023a). It should be noted that Serrano-Notivoli and Tejedor (2021) analyzing the performance of 48 gridded products proposed a general workflow to transform observations into grid estimates, which includes four steps: i) quality control, ii) data series reconstruction, ii) gridding and iv) assessment of the uncertainty.”

2) In the subsequent chapter, the datasets are delineated without any preceding explanation or introduction regarding their presentation, which may lead to confusion for the readers. For example, WRF parachutes in Section 2.3 without any justification or preangle (abstract is not part of the manuscript), leading readers to assume that the model will be employed for temperature and precipitation analysis. I recommend incorporating an introductory paragraph between items 2 and 2.1 to bridge this gap, as was implemented in Chapter 3.

Answer:

Following the reviewer's suggestion, the following lines have been added between 2 and 2.1 in the revised manuscript.

“In this section, the datasets utilized in the analysis are presented. Subsection 2.1 summarizes the daily observational data, including maximum (TX), minimum (TN), and average (TG) temperatures, as well as daily precipitation (PR). Subsection 2.2 outlines the procedures applied for quality control, gap filling, and homogenization of the datasets. Subsection 2.3 describes the Weather Research and Forecasting (WRF) model simulation, whose output is blended with the available temperature observational data using gridding techniques, as detailed in Section 3. This approach was preferred over relying solely on observational data due to the sparse spatial coverage of in situ measurements, especially at higher altitudes (above 1000 m) as presented in Subsection 2.1.”

3) Continuing with the discussion on WRF, only during Section 3.2 (the second section before the results), the readers are informed that the atmospheric model was utilised exclusively for the temperature field, which may reduce the audience rejection regarding precipitation. The decision to employ solely one year to represent the overall study period is somewhat contentious, as a significant amount of variability is forfeited. Nevertheless, this approach is permissible, given that the model ultimately functions as a spatial interpolator driven by physical laws, subsequently manipulated to incorporate the seasonal and interannual variations delineated by the observational data. It may be beneficial for the authors to include a map illustrating the participation of the observational data at each grid point, as it could mitigate the discussion concerning the employment of the atmospheric model to cover regions lacking a station. Furthermore, this addition would facilitate the analysis of the specific areas to which the results can be attributed through the model. It would be essential to formally present the domain's limits, since they may circumstantially restrict the representation of some large-scale atmospheric phenomena transiting from the boundary condition. Furthermore, the authors reference several studies on WRF applications in Greece but do not elaborate on their findings or the model calibration expressed by the selected set of physical parameterisations. This potentially makes the WRF application quite questionable for these purposes.

Answer:

We appreciate the reviewer's constructive suggestions. Given the 1 km resolution of the WRF model, it is indeed the case that the vast majority of $1 \text{ km} \times 1 \text{ km}$ grid cells lack direct in situ observations. This

spatial sparsity is a primary reason for our reliance on model-derived output, which offers a continuous and physically consistent representation of the temperature field. The absence of observational stations, particularly in areas above 1000 meters, further justifies our methodological choice, as the model helps fill observational gaps in complex terrain. Please see our response in the previous comment.

Regarding the reviewer's comment on the model extent, we agree that this clarification enhances the transparency of our modeling framework. Accordingly, we have included the following figure (Figure 2, new) in the revised manuscript that clearly depicts the model domain and its geographical boundaries. This addition enables the reader to assess the potential influence of boundary conditions on the study area and better understand the spatial context of the simulations.

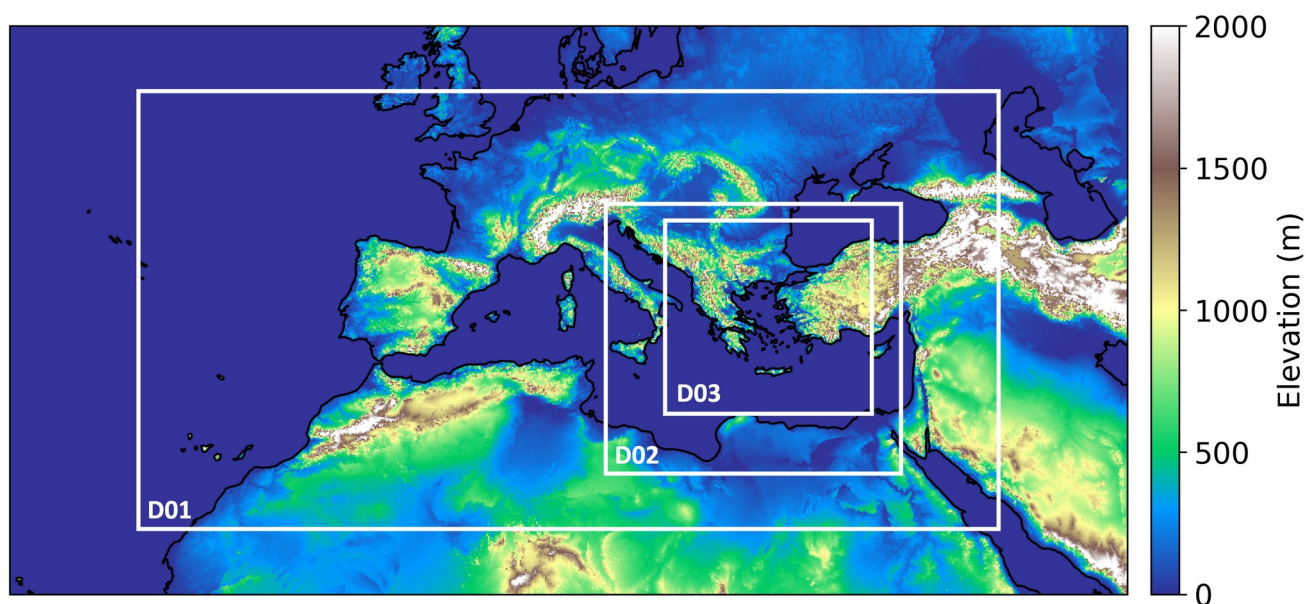


Figure 2: WRF-ARW model domains

Finally, in response to the comment on model performance, we have added references to studies where similar WRF setups have been used and evaluated. These studies report robust performance in comparable climatic and topographic settings. A clarifying sentence has also been included to highlight the prior validation of the model configuration used in our study.

Therefore, section 2.3 has been modified **from:**

“For the atmospheric simulations, the Advanced Weather Research and Forecasting Model (WRF-ARW) version 4.1.3 (Powers et al., 2017; Skamarock et al., 2008; Skamarock et al., 2019) was employed. WRF-ARW serves as a limited-area atmospheric model, utilized for both operational

forecasting (Sofia et al., 2024; Patlakas et al., 2023) and scientific research (Stathopoulos et al., 2023; Otero-Casal et al., 2019). It is based on a fully compressible, non-hydrostatic dynamic core. On the vertical plane it has terrain-following, mass-based, hybrid sigma-pressure vertical coordinates based on dry hydrostatic pressure, with vertical grid stretching permitted while for the horizontal grid, the Arakawa C-grid staggering is employed.

In this analysis, the WRF model was configured to run with three two-way nested grids. The coarser one has a resolution of 9 km, covering a large area that includes parts of North Africa and Central Europe. The inner grids are focused on the Eastern Mediterranean and Greece, with spatial resolutions of 3 km and 1 km, respectively. Vertically, the model consists of 48 layers.”

to :

“For the atmospheric simulations, the Advanced Weather Research and Forecasting Model (WRF-ARW), version 4.1.3 (Skamarock et al., 2019), was employed. WRF-ARW is a limited-area atmospheric model based on a fully compressible, non-hydrostatic dynamic core. Vertically, it utilizes terrain-following, mass-based hybrid sigma-pressure coordinates based on dry hydrostatic pressure, with support for vertical grid stretching. Horizontally, the model applies an Arakawa C-grid staggering.

WRF is widely used in both operational forecasting (Sofia et al., 2024; Patlakas et al., 2023) and scientific research (Pantillon et al., 2024; Patlakas et al., 2024; Politi et al., 2021; Stathopoulos et al., 2023; Otero-Casal et al., 2019). These studies include comprehensive evaluations of the model’s performance not only over the present study area but also in regions with similar topographic and climatic characteristics, demonstrating its reliability in representing climatological fields. In this analysis, the WRF model was configured with three two-way nested grids to adequately capture both regional and local-scale processes. The inner grids are focused on the Eastern Mediterranean and Greece, with spatial resolutions of 3 km and 1 km, respectively (Figure 2). Vertically, the model consists of 48 layers.”

4) The description of geostatistical interpolation and ML methods, lacking detail and overly restricted to their respective R packages citation, is both awkward and limiting (Sect. 3.1 and 3.2). This narrow focus obscures the wide variety of options and parameters defined in each method and, as a result, hinders and prevents the proper reproducibility of the study. Thus, presenting these details in the manuscript is crucial.

Answer:

Following the reviewer’s suggestion, the following lines have been modified in section 3.1 of the revised manuscript **from:**

“The following approaches were examined to calculate the monthly precipitation fields: (i) a “fixed rank Kriging approach” (hereafter FRK, Nychka et al., 2015), (ii) generalized additive models (hereafter GAM, Wood, 2017) and (iii) two ML algorithms namely k-nearest neighbors (hereafter KNN) and

support vector machines (hereafter, SVM). The analysis is performed using the R Language and Environment for Statistical Computing (R core team 2024) and the packages LatticeKrig (Nychka et al., 2019) for FRK, mgcv (Wood and Wood, 2015) for GAM, while the two ML algorithms are tuned using the caret R package (Kuhn, 2008).”

to :

“The following approaches were examined to calculate the monthly precipitation fields:

(i) a “Fixed Rank Kriging approach” (FRK). FRK is a geostatistical interpolation technique that approximates a spatial field using a low-rank representation of the underlying spatial process. It models the spatial covariance structure through a set of basis functions, allowing efficient estimation even with large datasets (Nychka et al., 2015). In this study, FRK is implemented using the LatticeKrig package (Nychka et al., 2019) in R (R Core Team, 2024), where the model parameters, including variance components and spatial range parameters, are estimated using maximum likelihood estimation.

(ii) Generalized Additive Models (hereafter GAM) are a semi-parametric extension of Generalized Linear Models that assume the underlying relationships are additive and smooth. Their primary strength lies in their ability to capture highly non-linear and non-monotonic relationships between the response variable and explanatory variables (Wood, 2017). In this study, monthly precipitation sums are modeled as smooth functions of longitude, latitude and elevation using thin plate regression splines, with smoothing parameters estimated using restricted maximum likelihood (mgcv R package, Wood and Wood, 2015).

(iii) two ML algorithms namely K-Nearest Neighbors (KNN) and Support Vector Machines (based on an exponential radial basis function, SVM). KNN estimates the value of an unknown data point by identifying its k closest neighbors in the spatial dataset, where k is a user-defined hyperparameter (Nwaila et al., 2024). The predicted value is computed as a weighted average of these neighbors' values, with the weights typically based on the distance to the target point, i.e. closer neighbors have greater influence. In this study, k ranges from 2 to 30 in increments of 1. SVM is a ML algorithm, effective in capturing non-linear spatial trends, which seeks a function that predicts the value of an unknown data point, while balancing accuracy and model complexity (Bonsoms and Ninyerola, 2024). The complexity is regulated by the cost parameter C (tested values: 0.1, 1, 5, 10), while the smoothness of the kernel is governed by σ (tested values: 0.01, 0.025, 0.05, 0.075, 0.1). For both algorithms, optimal hyperparameters (k for KNN, and C and σ for SVM) are selected using tenfold cross-validation combined with grid search, using the caret package in R (Kuhn, 2008; R Core Team, 2024).”

5) Concerning the metrics employed in the evaluation, specifically RMSE, MAE, and KGE, it is noteworthy that the most fundamental among them, Bias, was not considered. Relying solely on RMSE and MAE does not allow assessing whether the method underestimates or overestimates the observed values. This significantly impacts the analysis of temperature fields, particularly precipitation fields. Therefore, including this metric is also mandatory.

Answer:

Following the reviewer's suggestion BIAS values are also reported in Tables 2 and 3 of the revised manuscript. Moreover, the discussion in sections 4.1.1 and 4.1.2 have been modified to include the BIAS values. In particular, for precipitation in section 4.1.1 the following lines have been modified **from:**

“The values of the root mean square error (RMSE), the mean absolute error (MAE) and the KGEs are presented in Table 2 for the precipitation grids produced using the different approaches described in Sect. 3.1. These results are compared against the withheld observations on an annual and seasonal basis for the test period. From Table 2 it is clear that the methods yield similar values across all statistical measures with annual RMSE (MAE) being lower than 1.45 (0.65) for all methods. On a seasonal basis, the highest RMSE and MAE are observed during the DJF (December-January-February) and SON (September-October-November) periods. Overall, SVM and FRK methods exhibit the lowest values for both metrics on an annual and seasonal basis. Regarding the KGEs, with the exception of GAMs that exhibit a KGE value of 0.7 in JJA (June-July-August), for the rest of the seasons and on an annual basis the reported KGEs are higher than 0.8. Similar to RMSEs and MAEs, the highest KGE values are found for the SVM and FRK methods indicating that these two methods capture quite well the temporal distributions of the daily precipitation.”

to:

“ The values of BIAS, RMSE, MAE and KGE are presented in Table 2 for the precipitation grids produced using the different approaches described in Sect. 3.1. These results are evaluated against withheld observations on both an annual and seasonal scale over the test period. From Table 2, it is evident that on the annual scale, all four approaches exhibit strong and relatively consistent performance across all evaluation metrics. BIAS values are minimal, indicating that none of the models significantly overestimate or underestimate precipitation. FRK shows the smallest annual BIAS (-0.01 mm), closely followed by GAM (-0.02 mm) and SVM (-0.05 mm), while KNN shows a negative BIAS of -0.08 mm. Regarding RMSE and MAE, it is shown that the different approaches yield similar values across both statistical measures with annual RMSE (MAE mm) being lower than 1.45 (0.65 mm), while KGE values are higher than 0.85 for all approaches, with SVM and FRK reaching the highest values 0.93 and 0.92, respectively. At the seasonal scale, greater variability is shown. During winter (DJF, December-January-February), SVM and GAM tend to significantly underestimate precipitation, with SVM showing the most negative BIAS (-0.16 mm) and a KGE of 0.91, while GAM also underperforms with a lower KGE of 0.81 and relatively high MAE (1.01 mm). In contrast, KNN tends to overestimate during this period (BIAS = 0.12 mm), but still maintains a moderate KGE of 0.84. FRK, once again, shows stability with low BIAS (-0.02 mm), competitive RMSE (1.51 mm), and a strong KGE of 0.88. During spring (MAM, March-April-May) and summer (JJA, June-July-August), all approaches show

better agreement with observations. FRK continues to perform robustly with the lowest RMSE (MAE) values (0.94 mm (0.50 mm) for MAM and 0.65 mm (0.29 mm) for JJA), and high KGE values (0.90 and 0.82, respectively). The rest of the approaches perform reasonably, though GAM's MAE remains higher in MAM, and SVM and KNN show slight deviations in JJA. In autumn (SON, September-October-November), all approaches tend to overestimate precipitation, with KNN showing the highest BIAS (0.16 mm). SVM shows the lowest BIAS (0.02 mm) and maintains a KGE of 0.94, the highest among the methods for this season. Nevertheless, FRK demonstrates consistent performance with a balanced BIAS (0.05 mm), relatively low RMSE and MAE of 1.27 mm and 0.66 mm, respectively, and a strong KGE of 0.93. Overall, the results indicate FRK as the most stable and reliable method, delivering consistently low BIAS and error across both annual and seasonal scales while maintaining high KGE values throughout the year.”

Table 2: Precipitation annual and seasonal BIAS, RMSE, MAE and KGE statistics based on daily values between the 30 reference station and the interpolated ones as derived from the different interpolation approaches.

	FRK				GAM				KNN				SVM			
PR	BIAS	RMSE	MAE	KGE	BIAS	RMSE	MAE	KGE	BIAS	RMSE	MAE	KGE	BIAS	RMSE	MAE	KGE
Annual	-0.01	1.13	0.60	0.92	-0.02	1.45	0.63	0.87	-0.08	1.18	0.62	0.87	-0.05	1.09	0.59	0.93
DJF	-0.02	1.51	0.95	0.88	-0.06	1.46	1.01	0.81	0.12	1.56	0.97	0.84	-0.16	1.43	0.91	0.91
MAM	0.03	0.94	0.50	0.90	-0.03	1.47	0.93	0.88	-0.11	0.99	0.54	0.86	-0.04	0.92	0.51	0.92
JJA	-0.09	0.65	0.29	0.82	-0.16	1.47	0.3	0.70	-0.06	0.63	0.29	0.86	-0.08	0.67	0.31	0.82
SON	0.05	1.27	0.66	0.93	0.07	1.43	0.71	0.87	0.16	1.34	0.69	0.85	0.02	1.22	0.66	0.94

For temperatures (subsection 4.1.2) the following lines have been modified **from:**

“From Table 3 and Fig. 5-7, it is evident that all methods perform well in capturing the temporal temperature characteristics for *TX*, *TG* and *TN*. For *TX* the methods yielding the best performance in terms of RMSE, MAE and KGE are KNN and SVM. KNN shows better performance during the colder seasons (DJF and SON), while SVM in the warmer ones (March-April-May (MAM) and JJA). FRK is the third best performing method, while GAM exhibits the highest RMSEs at all periods examined. Considering the annual biases (Fig. 5) a different perspective emerges. GAM shows the lowest absolute biases (below 0.2 °C), while FRK displays the highest biases (up to 0.6 °C), highlighting variations in the strengths of the methods depending on the evaluation metric.”

to :

“From Table 3 and Fig. 5-7, it is evident that all methods perform well in capturing the temporal temperature characteristics for *TX*, *TG* and *TN*. Table 3 presents the values of the metrics as calculated from daily values, which offer insight into the different methods systematic errors at the finer temporal scale. For *TX*, KNN and SVM exhibit the best overall performance across RMSE, MAE, and KGE,

with seasonal RMSE values consistently below 0.67 °C and high KGE values (≥ 0.93). KNN performs particularly well in colder seasons (DJF and SON), while SVM shows better results in warmer periods (MAM and JJA). FRK ranks third, showing competitive RMSE and MAE values, though it consistently underestimates TX with a negative bias across all seasons, most notably in SON (-0.47 °C). GAM, while displaying the highest RMSE and MAE values in every season, has the lowest annual BIAS (0.00 °C) and relatively low seasonal BIAS values (e.g., -0.04 °C in DJF, -0.27 °C in MAM), indicating alignment with the observed annual mean but poor performance in capturing daily variability. When examining the annual differences between the observations and the various methods (Fig. 5), GAM demonstrates the lowest absolute annual deviations, remaining below 0.2 °C, indicating strong agreement with long-term averages. In contrast, FRK, exhibits the highest annual deviations, reaching up to 0.6 °C.

For TG and TN, FRK emerges as the most robust method, outperforming others across all metrics on both annual and seasonal scales. It consistently delivers the lowest RMSE, MAE, and KGE values. Importantly, FRK also exhibits the lowest annual BIAS values for TG (0.08 °C) and TN (0.10 °C), underscoring its minimal systematic deviation from observations. In comparison, other methods present substantially higher BIAS values, particularly KNN (0.57 °C for TG; 0.74 °C for TN) and SVM (0.71 °C for TG; 0.56 °C for TN). Furthermore, FRK's mean absolute annual deviations remain lower than 0.3 °C for both TG and TN (Fig. 6-7), whereas other methods show deviations reaching up to 0.8 °C, depending on the variable and method.”

Table 3: Daily maximum (TX), mean (TG) and minimum (TN) temperatures annual and seasonal RMSE, MAE and KGE statistics between the 13 reference station values and the interpolated ones as derived from the different interpolation approaches.

	FRK				GAM				KNN				SVM			
	BIAS	RMSE	MAE	KGE	BIAS	RMSE	MAE	KGE	BIAS	RMSE	MAE	KGE	BIAS	RMSE	MAE	KGE
TX																
Annual	-0.33	0.69	0.55	0.98	0.00	1.63	0.48	0.98	0.11	0.60	0.50	0.98	-0.12	0.60	0.49	0.98
DJF	-0.41	0.68	0.55	0.91	-0.04	1.6	0.41	0.93	-0.30	0.60	0.49	0.93	-0.38	0.64	0.53	0.94
MAM	-0.06	0.60	0.47	0.97	-0.27	1.53	0.53	0.97	-0.33	0.67	0.56	0.98	-0.21	0.62	0.51	0.98
JJA	-0.40	0.72	0.58	0.97	-0.14	1.35	0.49	0.98	-0.18	0.60	0.50	0.98	-0.12	0.56	0.46	0.98
SON	-0.47	0.76	0.61	0.96	-0.15	1.48	0.47	0.98	-0.02	0.55	0.45	0.99	0.03	0.57	0.46	0.98
TG																
Annual	0.08	0.45	0.36	0.98	0.41	2.81	0.48	0.97	0.57	0.69	0.59	0.97	0.71	0.82	0.73	0.96
DJF	-0.18	0.45	0.36	0.98	0.24	2.56	0.38	0.97	0.53	0.67	0.56	0.93	0.64	0.77	0.66	0.92
MAM	0.23	0.45	0.36	0.97	0.53	2.43	0.57	0.96	0.66	0.75	0.67	0.95	0.70	0.80	0.71	0.95
JJA	0.23	0.48	0.37	0.97	0.51	2.32	0.55	0.97	0.53	0.66	0.57	0.97	0.78	0.88	0.79	0.96
SON	0.23	0.42	0.35	0.96	0.51	2.53	0.43	0.97	0.53	0.67	0.58	0.97	0.78	0.83	0.74	0.96
TN																
Annual	0.10	0.52	0.41	0.99	0.51	3.85	0.59	0.96	0.74	0.88	0.77	0.94	0.56	0.77	0.64	0.95
DJF	-0.04	0.63	0.52	0.92	0.39	3.85	0.58	0.88	0.75	0.97	0.80	0.80	0.66	0.93	0.75	0.82
MAM	0.22	0.52	0.43	0.98	0.66	3.85	0.70	0.93	0.75	0.88	0.78	0.92	0.49	0.72	0.60	0.95
JJA	0.17	0.42	0.34	0.98	0.55	3.79	0.57	0.97	0.73	0.83	0.73	0.96	0.49	0.65	0.54	0.95
SON	0.06	0.48	0.38	0.99	0.45	3.85	0.54	0.96	0.74	0.88	0.77	0.94	0.62	0.79	0.68	0.95

6) The results chapter is well-structured, and its figures and tables are clear, readable, and sufficient to justify the main findings. Nevertheless, it is also true that omitting some information and making certain choices in preparing the outputs reduces their respective impacts. For example, why was Figure 4 presented only for 2016 and not for the 10 years of validation (2010-2019)? Besides lacking a direct relation to Table 3, this leads readers to believe that the differences pointed out by the authors are limited and specific to that year and may not be as pronounced in an overall analysis. Furthermore, including a fifth panel (e) with the map that highlights only the station locations using colours representing the observed values on the same scale as precipitation would significantly enhance this analysis and assist in determining which method was definitively superior. Keep in mind that spatial variability does not imply a better result.

Answer:

Following the reviewer’s suggestion, a new Figure 4 has been added in the revised manuscript replacing Figure 4 from the first version of the manuscript. In addition, the text in section 4.1.1 has been modified **from:**

“In terms of spatial distribution, the west to east gradient of precipitation in Greece, which exhibits the largest and lowest amounts of yearly precipitation (Gofa et al., 2019), is captured by all approaches. However, as Fig. 4 for the year 2016 illustrates, GAM, KNN and SVM demonstrate reduced spatial variability in regions with significant altitude differences in both the northeastern and southern regions of Greece (e.g. Crete). This spatial variability is also evident in other years examined within the 2010–2019 period (not shown).”

to:

“ When all available stations are used for the interpolation for the period 2010-2019 (Fig. 4), the results indicate that the west to east gradient of precipitation in Greece, which exhibits the largest and lowest amounts of yearly precipitation (Gofa et al., 2019) is captured by all approaches. However, as it is evident from Fig. 4, FRK maintains strong predictive skill when applied to the full set of available stations, with only a modest increase in error, reflecting its ability to handle spatial heterogeneity and non-stationarity common in precipitation fields. GAM error metrics increase somewhat on the full dataset, however, retaining a reasonable predictive skill, indicating its capacity to capture important spatial patterns. For KNN, an opposite sign BIAS is evident when compared to the results of Table 2 indicating that the method is strongly dependent on the proximity to known data. As a result, the method exhibits poorer extrapolation ability than FRK and GAM. In contrast, SVM shows a significant decline in performance when applied to the full station network, as it lacks explicit modeling of spatial structures despite its capability to capture complex nonlinear relationships (Heinke et al. 2023), resulting in higher errors and biases across diverse climatic and geographic regions. This behavior is highlighted in the mountainous area in Crete where total annual precipitation is much lower than the other methods.

Overall, this comparison highlights that FRK is better suited for robust precipitation interpolation across regions with complex topography.”

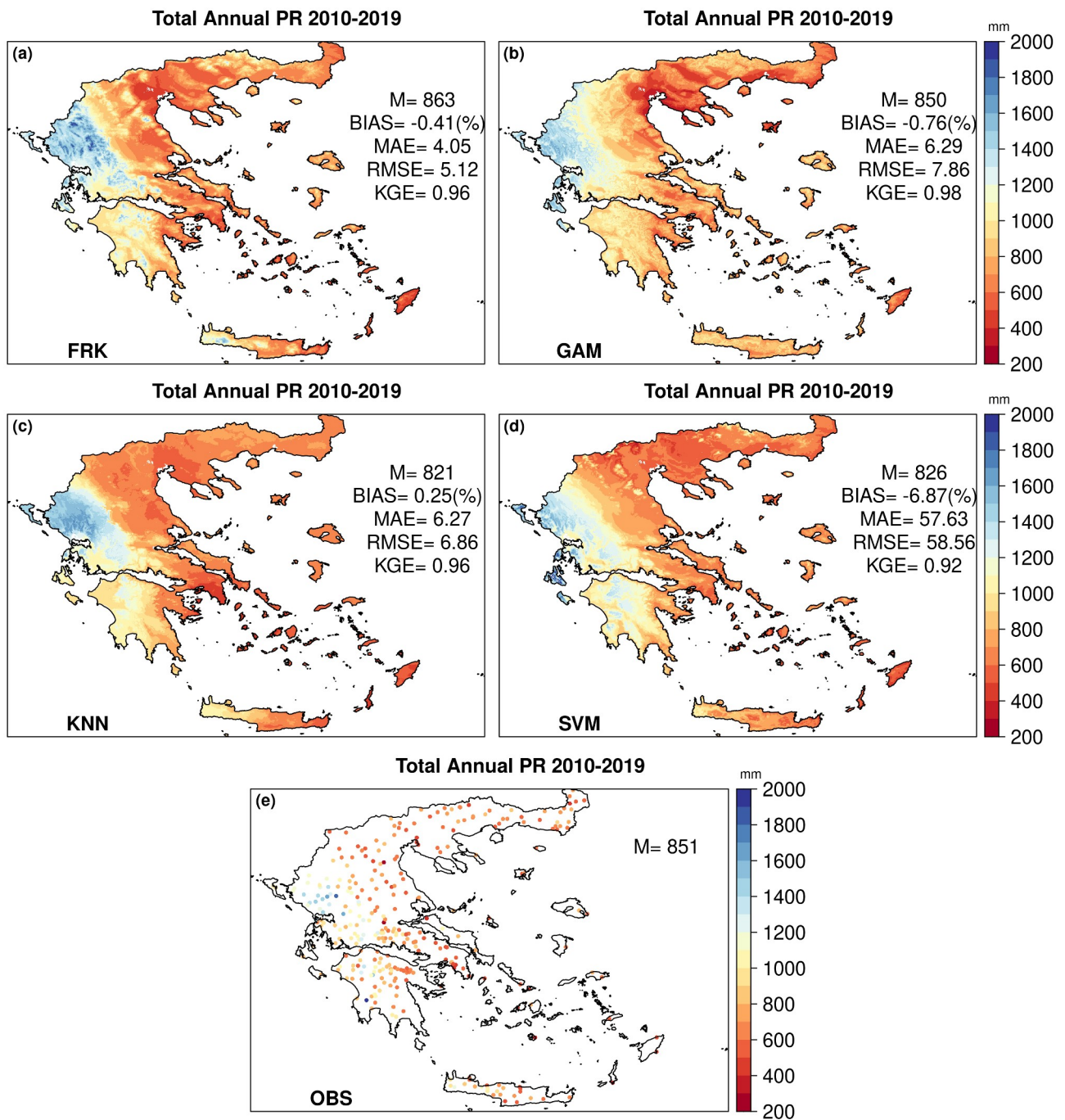


Figure 4. Spatial distribution of total annual precipitation for the period 2010–2019, as estimated by the different interpolation methods (panels a–d) and observed data (panel e). Each panel includes the spatial average (M) calculated over all grid points (or over all stations in panel e). For panels a–d, the relative BIAS, MAE (in mm), RMSE (in mm), and KGE are provided, based on comparisons between the interpolated values at the nearest grid points and the corresponding station observations.

In addition, the following lines 285-289 in subsection 4.1.2 have been modified **from**:

“In terms of spatial distribution, similar patterns to those observed for precipitation are identified. In particular GAM, KNN and SVM demonstrate reduced spatial variability between the mountainous areas of Greece and the lower altitude surrounding areas indicating limited ability to accurately represent temperature gradients influenced by topography (not shown).”

to:

“ In terms of spatial distribution, similar patterns to those observed for precipitation are identified. In particular GAM, KNN and SVM demonstrate limited ability to accurately represent temperature gradients influenced by topography (not shown).”

7) Another relevant point regards the sentence “Despite this, CHELSA underestimates the observed number of SU by about 10 days/yr, while CLIMADAT-GRid closely aligns with the observed values.” (L319-320). Which results substantiate this conclusion? It may be pertinent to present and discuss the evaluation metrics (Bias, RMSE, MAE, KGE) for both CLIMADAT-GRid and CHELSA throughout the entire study period. Only a brief and vague text in L350-354 may not be sufficient.

Answer:

In the first version of the manuscript in each panel of the Figures 8-12 the following values were reported within the panels: M which denoted the spatial average over the grid points covering the area, M.o which denoted the station mean values and M.c which denoted the mean values for the closest grid points to the stations locations. Therefore, the difference between the interpolated values at the nearest grid points (M.c) and the corresponding station observations (M.o) could be calculated. Nevertheless, in the revised version of the manuscript all the evaluation metrics are presented within each panel in addition to TG (the new figures can be found at the end of this response letter). Moreover, the differences between CLIMADAT-GRid and CHELSA can be found in the supplementary material (the new figures can be found at the end of this response letter). Consequently, the discussion in section 4.2 has been modified to incorporate all the changes. In particular, the following lines have been modified **from:**

“4.2.1 Daily maximum and minimum temperature results

In Fig. 8 and 9 the average annual and seasonal temperature results for *TX* and *TN* are shown, respectively. For *TX*, both datasets show comparable temperature values averaged over the whole domain of interest (denoted as M in the figures), with the maximum difference between the two datasets not exceeding 0.5 °C (DJF). When compared to the station data (M.o in the figures) CLIMADAT-GRid exhibits values of similar magnitude for the closest grid point to the station locations (M.c in the

figures), while CHELSA systematically underestimates the observations on both the annual and the seasonal timescales, with average discrepancies between 0.5 °C to 0.7 °C. In terms of spatial variability, CLIMADAT-GRid with WRF blended in the gridded dataset better captures the orographical gradient of temperatures, indicating lower temperatures than CHELSA in mountainous areas ranging from the northwest of Greece to the central Peloponnese and west of Crete. Furthermore, CHELSA appears to be cooler than CLIMADAT-GRid in the Ionian Sea islands of Zakynthos and Kefallonia, as well as the Aegean sea islands of Cyclades, but warmer in Rhodes and Samos.

For *TN* (Fig. 9), both datasets show nearly identical temperature values averaged across the entire domain of interest, with the highest difference between them not exceeding 0.2 °C (MAM and SON). CLIMADAT-GRid produces values of the same magnitude as the station data for the grid points nearest to the station locations, whereas CHELSA slightly overestimates the observations annually, DJF and MAM. Nonetheless, the maximum overestimation in daily minimum temperatures is less than 0.3 °C. In contrast, JJA and SON produce temperatures that are similar to those observed. The highest discrepancy between the two datasets is obvious while examining the maps in Fig. 8. In particular, CHELSA indicates notably higher temperatures compared to CLIMADAT-GRid, from west to east, in the Ionian Sea's Zakynthos and Kefallonia, as well as Crete, and the majority of the Aegean Sea islands. The higher temperatures in CHELSA can be attributed to the implementation of a basic statistical downscaling approach that employs atmospheric temperature lapse rates, B-spline interpolations, and high-resolution orography rather than a full physical scheme (Karger et al., 2023). Furthermore, according to the authors, constant lapse rates were utilized for all air temperature variables impacting minimum temperatures to a greater extent, as minimum temperatures in high altitudes are frequently the result of nighttime inversions.

Figure 10 illustrates the results for the number of days with daily $TX > 25$ °C (SU) and $TX > 35$ °C (SU35) and the number of days with daily $TN > 20$ °C (TR). From the figure it is evident that both datasets yield similar domain averaged values for SU and SU35. However, for TR, CHELSA estimates a higher number of days/yr compared to CLIMADAT-GRid (23 and 18 days/yr, respectively). Despite this, CHELSA underestimates the observed number of SU by about 10 days/yr, while CLIMADAT-GRid closely aligns with the observed values. Differences are most noticeable in the spatial variability of the results, which are most evident in the Ionian Sea Zakynthos and Kefallonia islands, as well as the Aegean Sea Cyclades for SU and in the Ionian Sea Zakynthos and Kefallonia islands, as well as Crete and the majority of the Aegean Sea's islands for TR. Furthermore, CLIMADAT-GRid catches pretty well the spatial variability of SU35, pinpointing well-known warm hot spots in the Greek territory, as well as the urban heat island effect in the Athens urban area, which is also evident in CHELSA, but to a lesser extent.”

to :

“4.2.1 Daily maximum, mean and minimum temperature results

Figures 8, 9 and 10 present the annual and seasonal average temperature results for TX, TG, and TN for the CLIMADAT-GRid and CHELSA datasets, respectively, while their differences are shown in Fig. S1 of the Supplementary Material. For TX, both datasets show broadly comparable spatial average temperatures over the entire domain (denoted as M in the figures) (Fig. 8). CLIMADAT-GRid consistently matches station observations well, as indicated by zero BIAS and minimal error metrics across all seasons. MAE and RMSE values remain at or below 0.02 °C, and KGE values are close to 0.99 throughout. In contrast, CHELSA systematically underestimates TX, with biases ranging from -0.49 °C in MAM to -0.69 °C in DJF, and RMSE values up to 0.72 °C. CHELSA's lower KGE values (e.g., 0.86 in DJF) further suggest reduced agreement with observations. CLIMADAT-GRid, blended with WRF in the gridded dataset, captures the orographic temperature gradients more effectively exhibiting lower temperatures in elevated regions such as the northwest of Greece, the central Peloponnese, and western Crete (Fig. S1). Conversely, CHELSA tends to show cooler conditions in the Ionian and Cycladic islands but warmer conditions in Rhodes and Samos.

For TG, the pattern is similar but slightly more pronounced in favor of CLIMADAT-Grid (Fig. 9). The mean annual TG is 14.3 °C in CLIMADAT-GRid and 14 °C in CHELSA. CLIMADAT-GRid demonstrates extremely low errors and near-zero bias across all seasons, with RMSE values consistently at 0.01–0.02 °C and KGE values at or near 0.99. In contrast, CHELSA underestimates TG with average annual and seasonal biases between -0.29 °C (DJF) and -0.79 °C (JJA), and RMSE values reaching up to 0.81 °C. The KGE for CHELSA is lower, particularly in DJF (0.90) and SON (0.92), indicating less accurate temperature modeling compared to CLIMADAT-GRid. The spatial differences also reflect the better performance of CLIMADAT-GRid, especially in mountainous regions where it more accurately captures lower mean temperatures (Fig. S1).

For TN, both datasets report nearly identical domain-averaged values, with the largest difference being only 0.2 °C in MAM and SON (Fig. 10). CLIMADAT-GRid slightly underestimates TN (bias from -0.01 °C to -0.03 °C), whereas CHELSA slightly overestimates it during DJF and MAM (bias up to 0.27 °C), though both datasets perform well in JJA and SON. Despite the small average differences, error metrics again favor CLIMADAT-GRid, which shows low MAE and RMSE (0.01–0.03 °C) and perfect or near-perfect KGE values (≥ 0.99). CHELSA, by contrast, displays larger errors (RMSE up to 0.34 °C in MAM) and lower KGE, particularly annually (0.85) and in JJA (0.85), indicating a modest degradation in its representation of minimum temperatures. Regionally, the most significant TN discrepancies appear in coastal and island areas (Fig. S1). CHELSA shows notably higher TN values than CLIMADAT-GRid across Zakynthos, Kefallonia, Crete, and many of the Aegean islands. The higher temperatures in CHELSA can be attributed to the implementation of a basic statistical downscaling approach that employs atmospheric temperature lapse rates, B-spline interpolations, and high-resolution orography rather than a full physical scheme (Karger et al., 2023). Furthermore, according to the authors, constant lapse rates were utilized for all air temperature variables impacting minimum temperatures to a greater extent, as minimum temperatures in high altitudes are frequently the result of nighttime inversions.

Figure 11 presents a comparative analysis of the number of days exceeding key temperature thresholds, $TX > 25^{\circ}\text{C}$ (SU), $TX > 35^{\circ}\text{C}$ (SU35), and $TN > 20^{\circ}\text{C}$ (TR) based on the CLIMADAT-GRid and CHELSA datasets. While both datasets show close agreement in domain-averaged values for SU and SU35, notable discrepancies emerge for TR, where CHELSA reports a higher frequency (23 days/year) compared to CLIMADAT-Grid (18 days/year).

When benchmarked against observations, CLIMADAT-GRid demonstrates stronger agreement, particularly for SU. CHELSA underestimates SU by approximately 10 days/year, whereas CLIMADAT-Grid closely tracks observed values, reflecting its higher reliability for this metric. Statistical evaluation supports this, since for SU, CLIMADAT-Grid achieves a low BIAS (1°C), MAE (1.11°C), RMSE (1.4°C), and a high KGE (0.97), outperforming CHELSA, which shows a significant negative BIAS (-9.34°C), higher MAE (9.34°C), RMSE (9.69°C), and a slightly lower KGE (0.91).

For TR, the results are more nuanced. CLIMADAT-GRid shows a greater negative BIAS (-4.02 days/year), but CHELSA, despite the smaller BIAS (-1.28 days/year), presents higher MAE and RMSE values (3.09 days/year and 3.72 days/year, respectively), suggesting differences in how each dataset captures nighttime temperature extremes. Both datasets perform comparably in terms of KGE, with values of 0.84 (CLIMADAT-GRid) and 0.82 (CHELSA).

Spatial differences further reveal key distinctions between the two datasets (Fig. S2). Discrepancies in SU are concentrated over the Ionian Sea islands (Zakynthos and Kefallonia) and the Cyclades, while differences in TR are more widespread, notably affecting Crete, the Aegean islands, and again the Ionian Sea region. In addition CLIMADAT-GRid captures the spatial distribution of TR capturing the urban heat island effect in Athens. In contrast, this urban signature is less pronounced in the CHELSA data, suggesting limitations in its resolution or calibration over complex urban terrains. Similarly, CLIMADAT-Grid effectively captures the spatial distribution of SU35, accurately identifying thermal hotspots across Greece.

Together, these findings highlight the greater consistency and spatial sensitivity of the CLIMADAT-GRid dataset, particularly in reflecting observed heat extremes and local variability across the Greek region.”

Moreover, the following lines of subsection **4.2.2 Precipitation results** have been modified **from:**

“Figure 11 presents the total annual and seasonal precipitation results averaged over the period 1981–2016 for both datasets. In general, CLIMADAT-GRid indicates higher precipitation values compared to CHELSA on both the annual and the seasonal timescales. Both datasets capture the west-to-east precipitation gradient in Greece, with wetter conditions prevailing in the west and drier conditions in the east. However, the differences are more pronounced in CLIMADAT-GRid, particularly during Greece's rainy seasons (SON and DJF). When compared to the observations CLIMADAT-GRid indicates negligible biases, while for CHELSA the relative biases are about 18 % for the annual total precipitation and for the seasonal precipitation the biases range from about 15 % in SON to about 24 % in JJA with the rest of the seasons indicating intermediate biases.

Regarding the number of wet days (RR1, Fig. 12) the results are different since both datasets indicate a clear overestimation compared to the observed spatial means, with CLIMADAT-GRid indicating a more pronounced orographic pattern than CHELSA. This highly positive bias in the number of wet days has also been found in other gridded products (e.g., IBERIA01) and it is a byproduct of the selected interpolation methods. One way to reduce the inflated number of wet days is to introduce a third term in the interpolation scheme of precipitation by interpolating the daily occurrence of rainfall (0 or 1 depending on whether $PR > 0.1$ mm) considering a threshold between 0.1 and 0.9 for assigning a wet day to a grid point (Cornes et. al., 2018; Varotsos et al., 2023a). For instance, if we assign a value of 0.2 for the wet days and multiply the interpolated fields with the daily precipitation product the average number of wet days is reduced to 90 days/year with however increased underestimation in the annual and seasonal precipitation sums (not shown). For future studies utilizing the CLIMADAT-GRid precipitation dataset, a threshold of 2 mm/day could be considered when analyzing the number of wet days.

Finally, for the number of days with precipitation amounts higher or equal than 10 mm (R10mm, Fig. 12) both datasets exhibit similar results as it is shown from the spatial means with CLIMADAT-GRid indicating higher values for the specific index than CHELSA and closer to the observed spatial means. The highest values are shown for both datasets in western Greece, while the lowest in the eastern areas of the Greek domain.”

to :

“Figure 12 presents the total annual and seasonal precipitation results averaged over the period 1981–2016 for both datasets. In general, CLIMADAT-GRid indicates higher precipitation values compared to CHELSA on both the annual and seasonal timescales (their relative differences are shown in Fig. S3 of the Supplementary Material). Both datasets capture the west-to-east precipitation gradient in Greece, with wetter conditions prevailing in the west and drier conditions in the east.

When evaluated against observations, CLIMADAT-GRid shows minimal biases. Specifically, the annual BIAS is -1.56%, with seasonal biases ranging from -1% in DJF to -9.27% in JJA. CLIMADAT-GRid also maintains low MAE and RMSE values across all seasons, for example, annual MAE is 11.5 mm and RMSE is 15.17 mm, with high KGE values near 0.98, indicating strong agreement with observations.

In contrast, CHELSA demonstrates significant underestimations. The annual precipitation BIAS is -19%, and seasonal BIAS ranges from -11.97% in SON to -38.27% in JJA. The errors are also substantially larger, with annual MAE and RMSE at 142.51 mm and 147.02 mm, respectively. Seasonal MAE and RMSE values are consistently higher than those of CLIMADAT-GRid, particularly during DJF, MAM and SON. Additionally, CHELSA's KGE values are lower across all seasons, with a peak of 0.82 in SON and a low of 0.65 in JJA, indicating comparatively reduced reliability in capturing observed precipitation patterns.

Regarding the number of wet days (RR1, Fig. 13 and Figure S4 of the Supplementary Material), both datasets demonstrate a systematic overestimation relative to the observed spatial means. For CLIMADAT-GRid a positive BIAS of about 49 days/year is shown, while for CHELSA the bias reaches about 31 days/year. MAE and RMSE for CLIMADAT-GRid are about 49 and 50 days/year, respectively, with a negative KGE of -0.51, indicating poor agreement with observations despite its more pronounced orographic pattern. CHELSA performs somewhat better in this respect, with a lower MAE (31 days/year), RMSE (31 days/year), and a positive KGE of 0.31. This highly positive bias in the number of wet days has also been found in other gridded products (e.g., IBERIA01), and it is a byproduct of the selected interpolation methods. One way to reduce the inflated number of wet days is to introduce a third term in the interpolation scheme of precipitation by interpolating the daily occurrence of rainfall (0 or 1 depending on whether $PR > 0.1$ mm) considering a threshold between 0.1 and 0.9 for assigning a wet day to a grid point (Cornes et. al., 2018; Varotsos et al., 2023a). For instance, if we assign a value of 0.2 for the wet days and multiply the interpolated fields with the daily precipitation product the average number of wet days is reduced to 90 days/year with however increased underestimation in the annual and seasonal precipitation sums (not shown). For future studies utilizing the CLIMADAT-GRid precipitation dataset, a threshold of 2 mm/day could be considered when analyzing the number of wet days.

In terms of the number of days with daily precipitation equal to or greater than 10 mm (RR10, Fig. 13, Fig. S4 of the Supplementary Material), the two datasets display similar spatial distributions, with both indicating the highest frequencies in western Greece and the lowest in the east. However, CLIMADAT-GRid performs better quantitatively, with a mean annual R10mm of 23 days and a bias of about -3 days/year, compared to CHELSA's 19 days/year and a larger negative bias of about -7 days/year. Additionally, CLIMADAT-GRid exhibits lower MAE and RMSE (3 days/year for both metrics, respectively), along with a higher KGE of 0.86, indicating close agreement with observations. CHELSA, in contrast, yields a higher MAE, RMSE (7 days/year for both metrics, respectively), and a lower KGE of 0.73, reinforcing the overall tendency of CLIMADAT-GRid to more accurately represent moderate-to-heavy precipitation events.”

8) Finally, the conclusions chapter effectively fulfills its intended role, although it does not provide commentary on an essential aspect of the work regarding the various geostatistical and machine learning techniques employed in developing the temperature and precipitation datasets. This oversight may be attributed to the insufficient detail in the preceding chapters. Incorporating these elements, whether in the methodology or the conclusions, would substantially enhance the manuscript's value. Nonetheless, it is essential to underscore that the analyses provided are devoid of any fallacies or significant flaws and, in any case, compromise the integrity of the study. It is simply a matter of refinement.

Answer:

Following the reviewer's suggestion the following lines have been added in the Conclusions section:

“To produce the gridded fields, we evaluated four interpolation methods, Fixed Rank Kriging (FRK), Generalized Additive Models (GAM), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), using independent station data for validation. FRK emerged as the most reliable method, demonstrating consistent performance across variables and time scales, particularly for precipitation. It also best captured spatial patterns, especially over the complex terrain of Greece. For temperatures, SVM and KNN performed well for maximum temperatures, while FRK was more consistent for mean and minimum temperatures. FRK was ultimately chosen as the method for constructing the CLIMADAT-Grid.”

Given the innovative approach and the potential contributions to regional climate, I recommend acceptance upon a comprehensive review of major comments. The paper presents a high-quality, methodologically sound dataset likely to be of great use in regional climate research, impact modelling, and policy work in Greece. The authors must tackle the previously mentioned concerns to enhance their transparency, reproducibility, and broader significance. Addressing these issues will fortify the manuscript and increase its contribution to the scientific community.

Answer:

We thank the reviewer for their positive evaluation and constructive feedback. We have carefully addressed all issues raised, improving the manuscript's transparency, reproducibility, and overall clarity.

MINOR COMMENTS

L23, L26 CHELSA still CHELSA-W5E5 up to this point.

Corrected

L31 The phrase “... are becoming...” requires modification. This citation originates from 2012. Currently, it represents a prevailing reality.

Modified

L32-33 Add a comma in “(Herrera et al. 2012)”.

Corrected

L55 Remove the E-OBS citation that was previously introduced in L41.

Corrected

L63 Citation missing for “MeteoSerbia1km”.

Corrected

L73 The acronym "CLIMADAT-GRid" is used without prior definition. Please define it upon first use.

CLIMADAT-GRid is a designated name and not an acronym

Fig 1 Figures 1 and 2 are redundant; only Figure 2 is sufficient if it replaces Figure 1. Furthermore, the blue colouration on terrain elevation maps is typically attributed to regions situated below the mean sea level ($h < 0$). Consequently, it is advisable to redefine the scale to initiate with green tones.

Following the reviewer's suggestion, the following figure has been added in the revised manuscript which replaces Figures 1 and 2 of the first version of the manuscript.

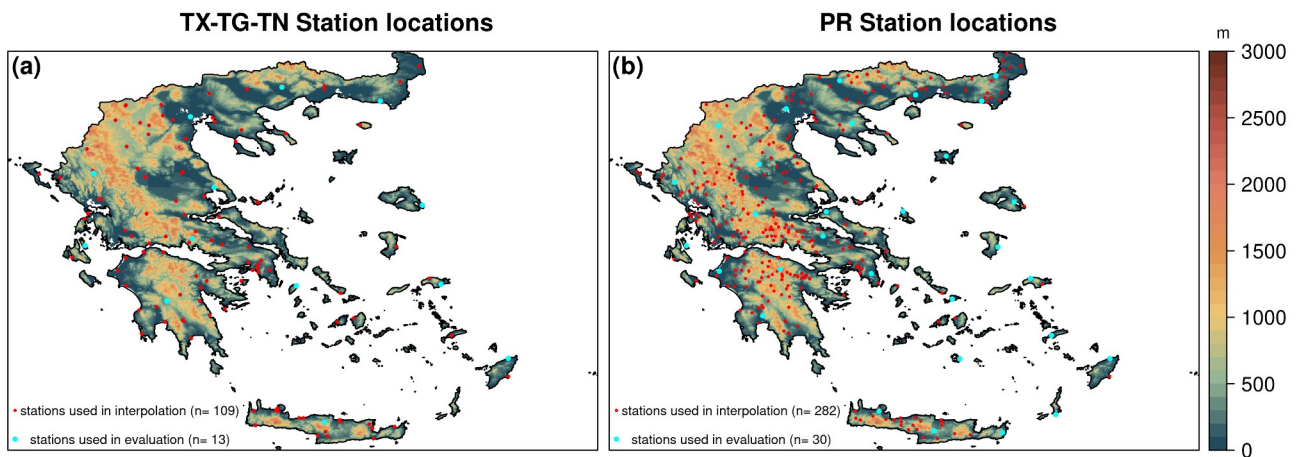


Figure 1. Locations of meteorological stations used for (a) temperature and (b) precipitation measurements, including both the stations used in the interpolation and the withheld stations used for evaluation. The background shows elevation data from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010).

L120 Citing Skamarock et al. (2019) may be sufficient.

The reviewer's suggestion has been implemented.

L136 Remove the term “approximately” once ERA5 has a precise horizontal resolution of 0.25°. In this case, the approximation regards the resolution in km, which varies from 25 to 31 km, roughly estimated at 28 km.

The reviewer's suggestion has been implemented.

L137 Following the standard presented in the manuscript, replace “*USGS (United States Geological Survey) (Slater et al., 2011)*” with “*United States Geological Survey (USGS, Slater et al., 2011)*”.

The reviewer’s suggestion has been implemented.

L137 Following the standard presented in the manuscript, replace “*CORINE (Coordination of Information on the Environment) database (2010)*” with “*Coordination of Information on the Environment (CORINE, CLMS 2018)*” if the authors have used the latest version (<https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac>). Additionally, this citation may be included in the references.

The reviewer’s suggestion has been implemented.

L153 The acronym “GMTED2010” is used without prior definition. Please define it upon first use (Figure’s caption doesn’t count).

The reviewers’ suggestion has been implemented.

L202 Remove the comma in “*Papa and Koutroulis (2025,)*”.

The reviewers’ suggestion has been implemented.

L205 Replace “*Climate Change Detection and Indices (ETCCDI) (Zhang et al. 2011).*” with “*Climate Change Detection and Indices (ETCCDI, Zhang et al., 2011).*”.

The reviewers’ suggestion has been implemented.

Fig 2 In addition to the aforementioned comments regarding Figure 1, it is advisable to change the colour used for the markers on the evaluation stations, as they tend to blend with the background.

The reviewers’ suggestion has been implemented.

L217 Replace “*The values of the root mean square error (RMSE), the mean absolute error (MAE) and the KGEs...*” with “*The values of RMSE, MAE and the KGEs...*”, once they were defined previously.

The reviewers’ suggestion has been implemented.

Fig 8,9,11 Since the objective of these figures is to compare the temperature and precipitation fields of two different datasets, wouldn't it be better to present the difference between them instead of the entire fields? This way, the differences pointed out by the authors would be clearer. Furthermore, it would enable the presentation of TN, TG, and TX in the same figure without losing quality. That is, Figures 8 and 9 would be merged, with the addition of TG, which was omitted without explanation.

We have retained the original structure for presenting the comparison between the two datasets in the revised manuscript. This approach not only facilitates the comparison itself but also serves to illustrate how temperature and precipitation fields are distributed across the complex terrain of Greece, which is a key objective of the study. However, to address the reviewer's concern, we have included the differences between the two datasets for the annual and seasonal means, as well as for the climate indices, in the Supplementary Material. Additionally, a new figure related to TG (new Figure 9) has been incorporated and discussed in the revised manuscript. Please also refer to our response to Comment 7 in the list of major comments for further clarification.

General The recursive use of “*hereafter*” is inappropriate in most instances. Typically, this expression is employed to redefine a name or acronym. Only in L199 does it appear to be correctly utilised to redefine the acronym CHELSA-W5E5 as CHELSA.

The reviewers' suggestion has been implemented.

General In the scientific literature on climate and meteorology, the prevailing terminology for the temporal aggregation of precipitation over a day is “*daily accumulated precipitation*” or simply “*daily precipitation*”. Although the procedure is referred to as the precipitation sum, its use can lead to different interpretations.

The reviewers' suggestion has been implemented. Daily precipitation is used in the revised manuscript.

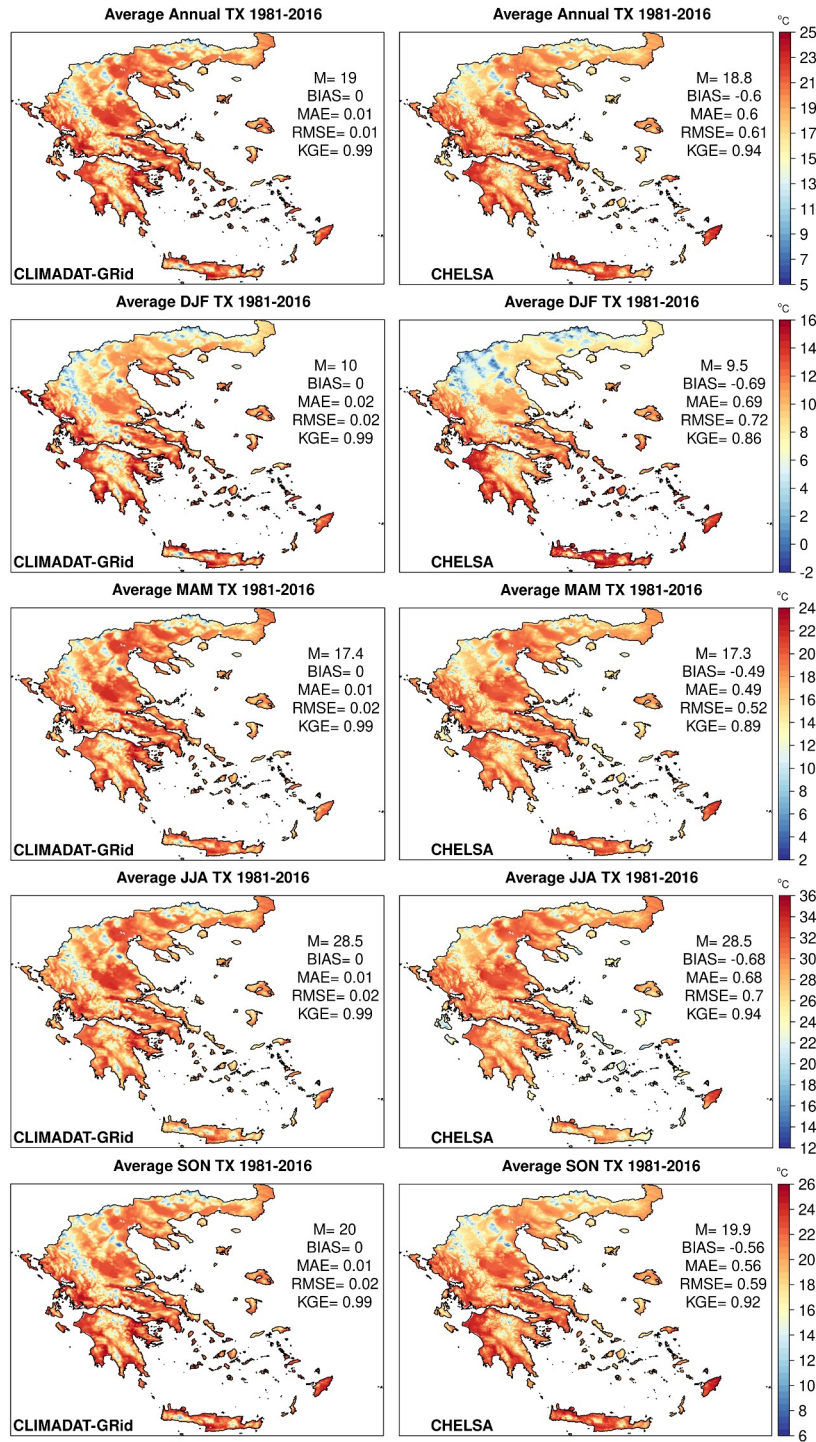


Figure 8. Average annual and seasonal TX (winter (DJF), spring (MAM), summer (JJA) and autumn (SON)) for the period 1981–2016 for CLIMADAT-GRid (left column) and CHELSA (right column). In each panel, M denotes the spatial average

over the grid points covering the area. In addition, the evaluation metrics between the stations and the data for the closest grid points to the stations locations are shown within each panel.

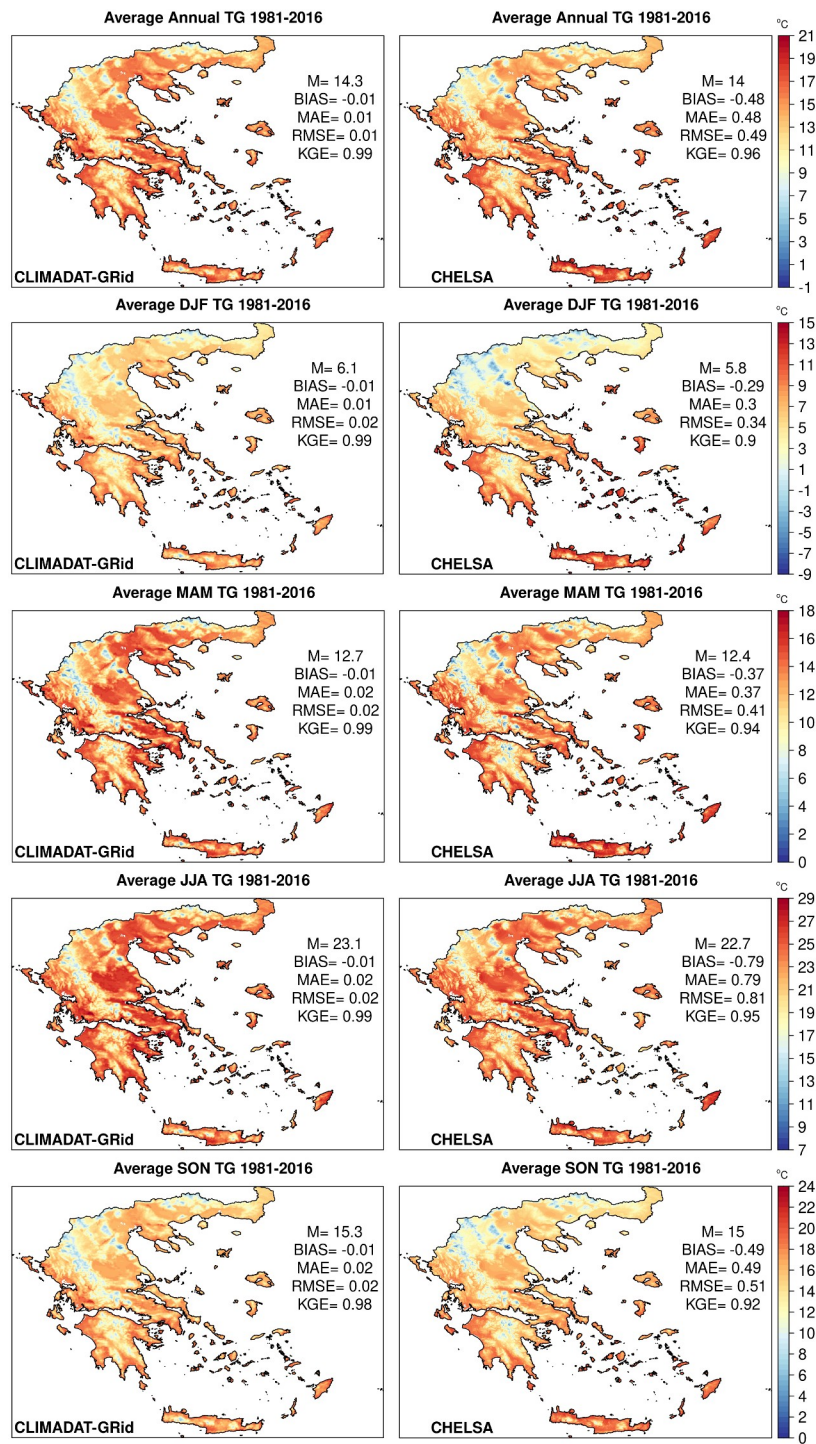


Figure 9. Average annual and seasonal TG (winter (DJF), spring (MAM), summer (JJA) and autumn (SON)) for the period 1981–2016 for CLIMADAT-GRid (left column) and CHELSA (right column). In each panel, M denotes the spatial average

over the grid points covering the area. In addition, the evaluation metrics between the stations and the data for the closest grid points to the stations locations are shown within each panel.

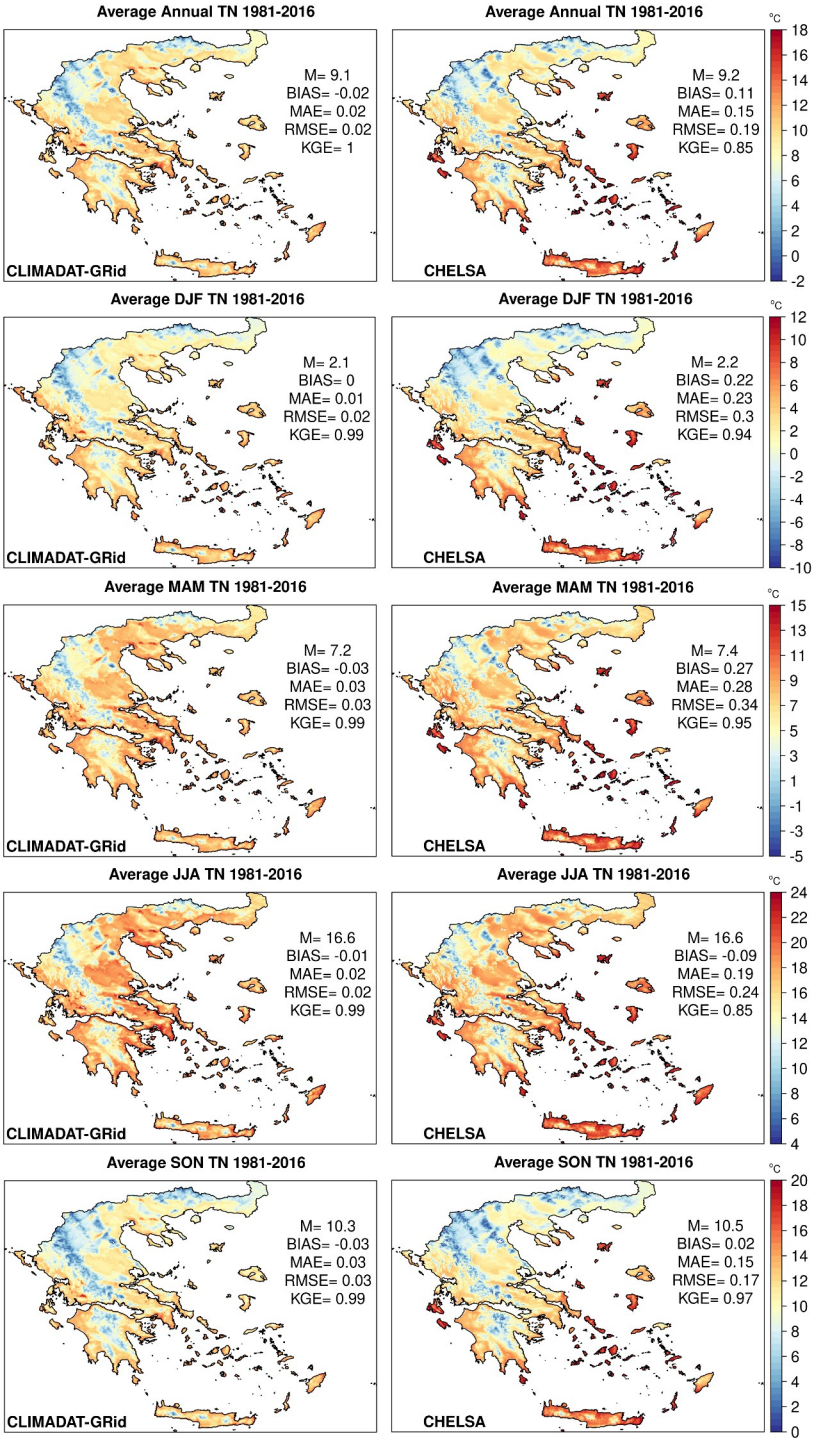


Figure 10. Average annual and seasonal TN (winter (DJF), spring (MAM), summer (JJA) and autumn (SON)) for the period 1981–2016 for CLIMADAT-GRid (left column) and CHELSA (right column). In each panel, M denotes the spatial average

over the grid points covering the area. In addition, the evaluation metrics between the stations and the data for the closest grid points to the stations locations are shown within each panel.

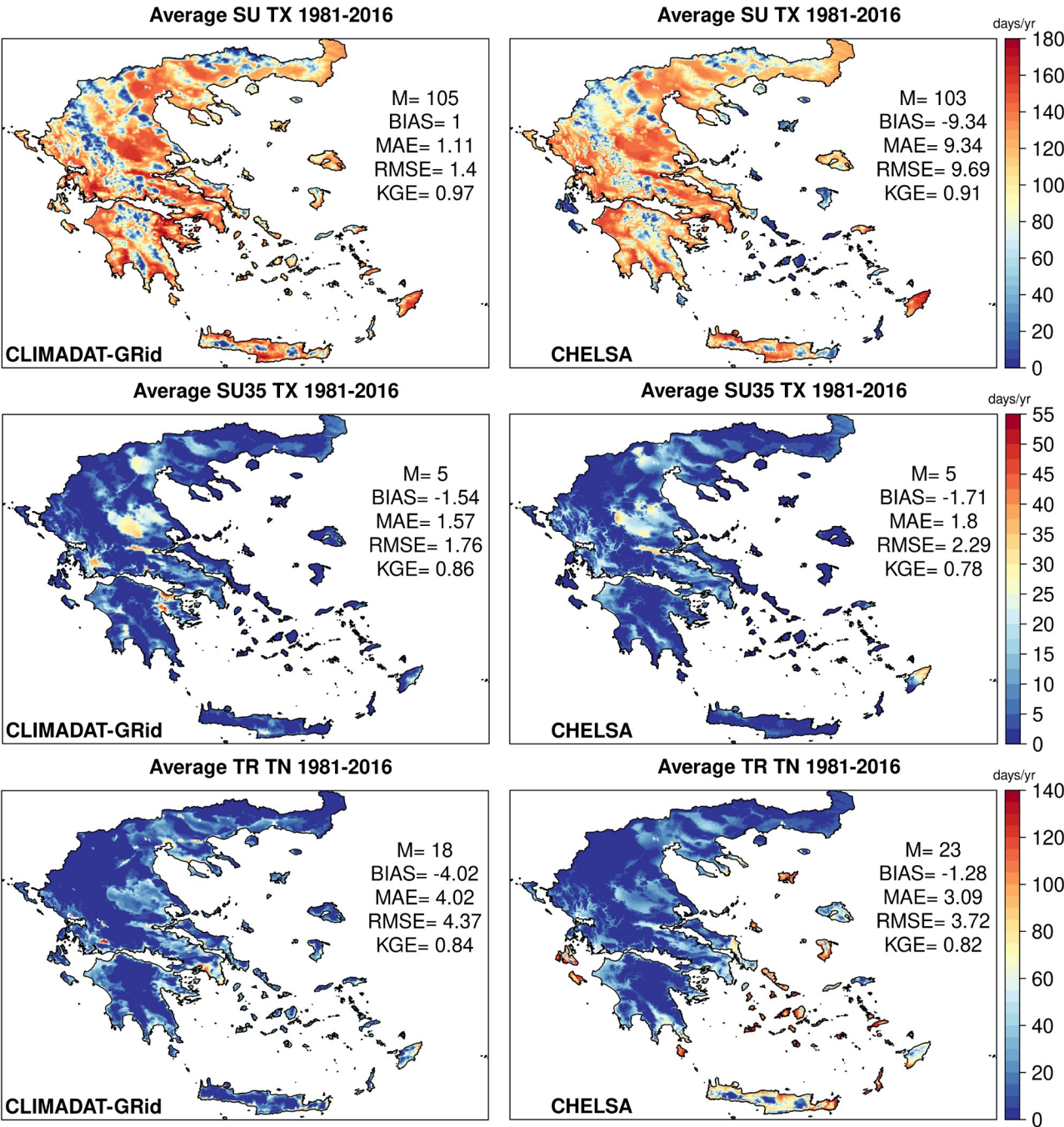


Figure 11. Average annual number of days $TX > 25^{\circ}C$ (SU), number of days $TX > 35^{\circ}C$ (SU35) and number of days $TN > 20^{\circ}C$ (TR) for the period 1981–2016 for CLIMADAT-GRid (left column) and CHELSA (right column). In each panel, M denotes the spatial average over the grid points covering the area. In addition, the evaluation metrics between the stations and the data for the closest grid points to the stations locations are shown within each panel.

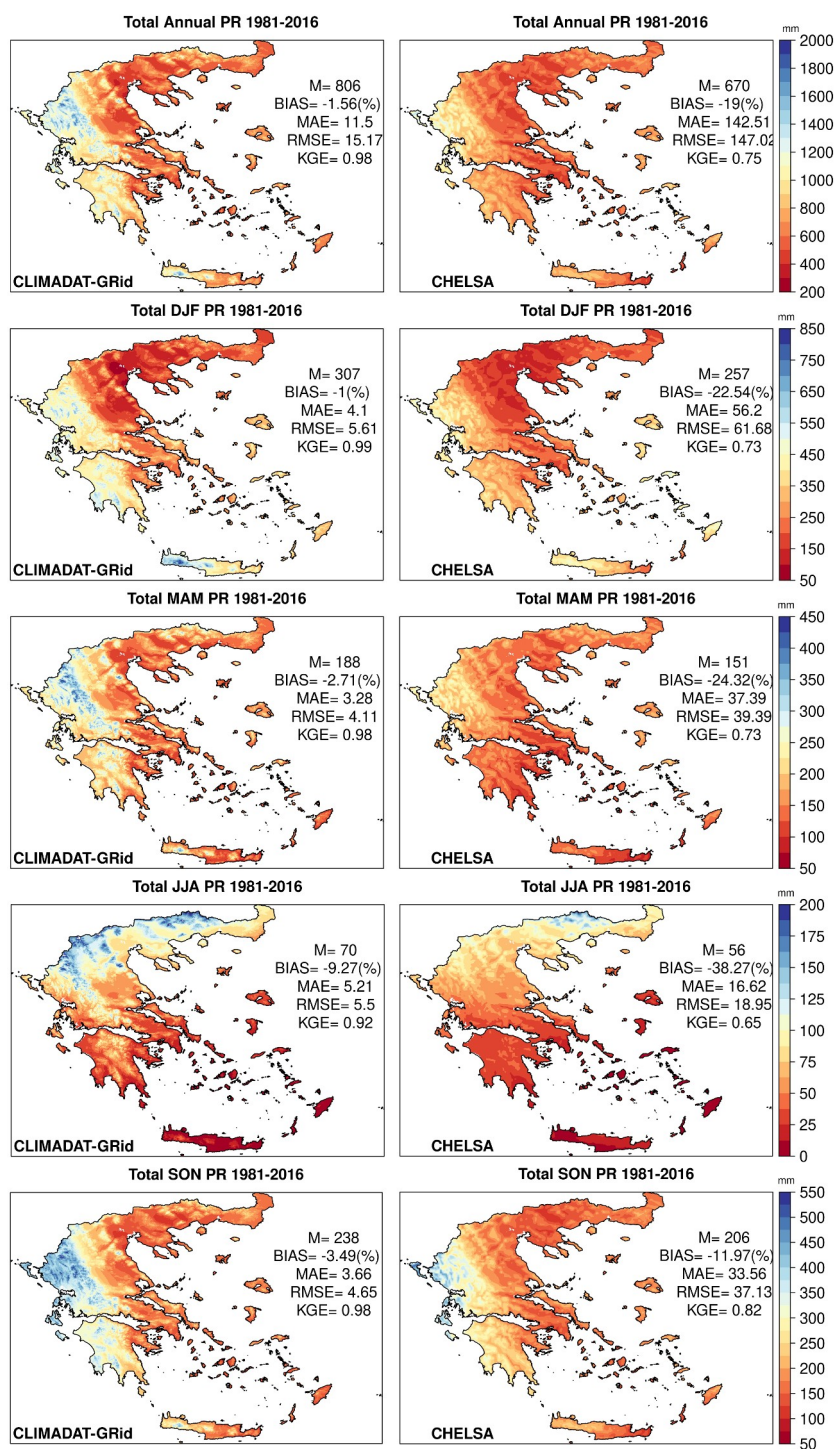


Figure 12. Total annual and seasonal PR (winter (DJF), spring (MAM), summer (JJA) and autumn (SON)) for the period 1981–2016 for CLIMADAT-GRid (left column) and CHELSA (right column). In each panel, M denotes the spatial average

over the grid points covering the area. In addition, the evaluation metrics between the stations and the data for the closest grid points to the stations locations are shown within each panel.

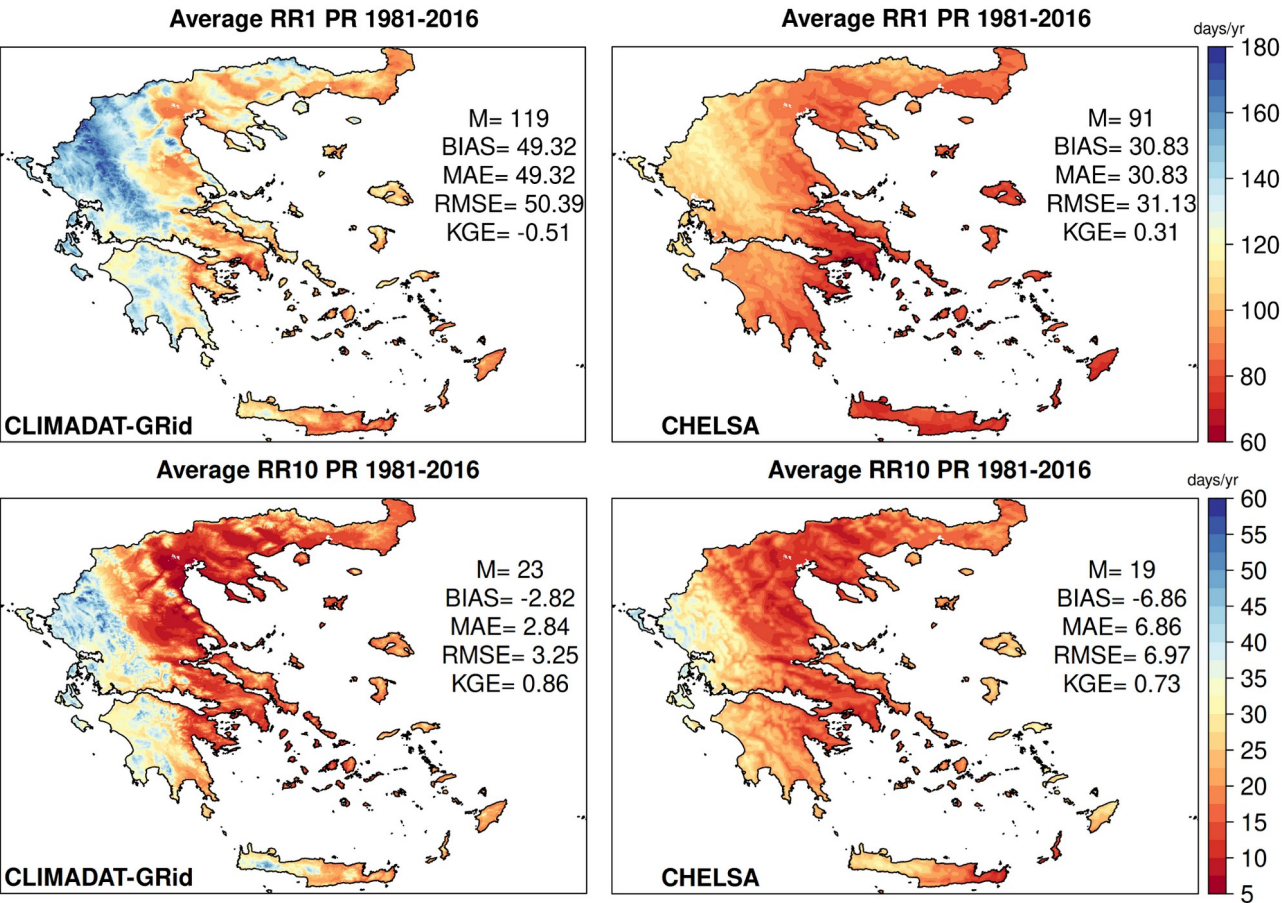


Figure 13. Average annual number of days $PR > 1$ mm (RR1) and number of days $PR \geq 10$ mm (R10mm) for the period 1981–2016 for CLIMADAT-GRid (left column) and CHELSA (right column). In each panel, M denotes the spatial average over the grid points covering the area. In addition, the evaluation metrics between the stations and the data for the closest grid points to the stations locations are shown within each panel.

Supplementary material figures

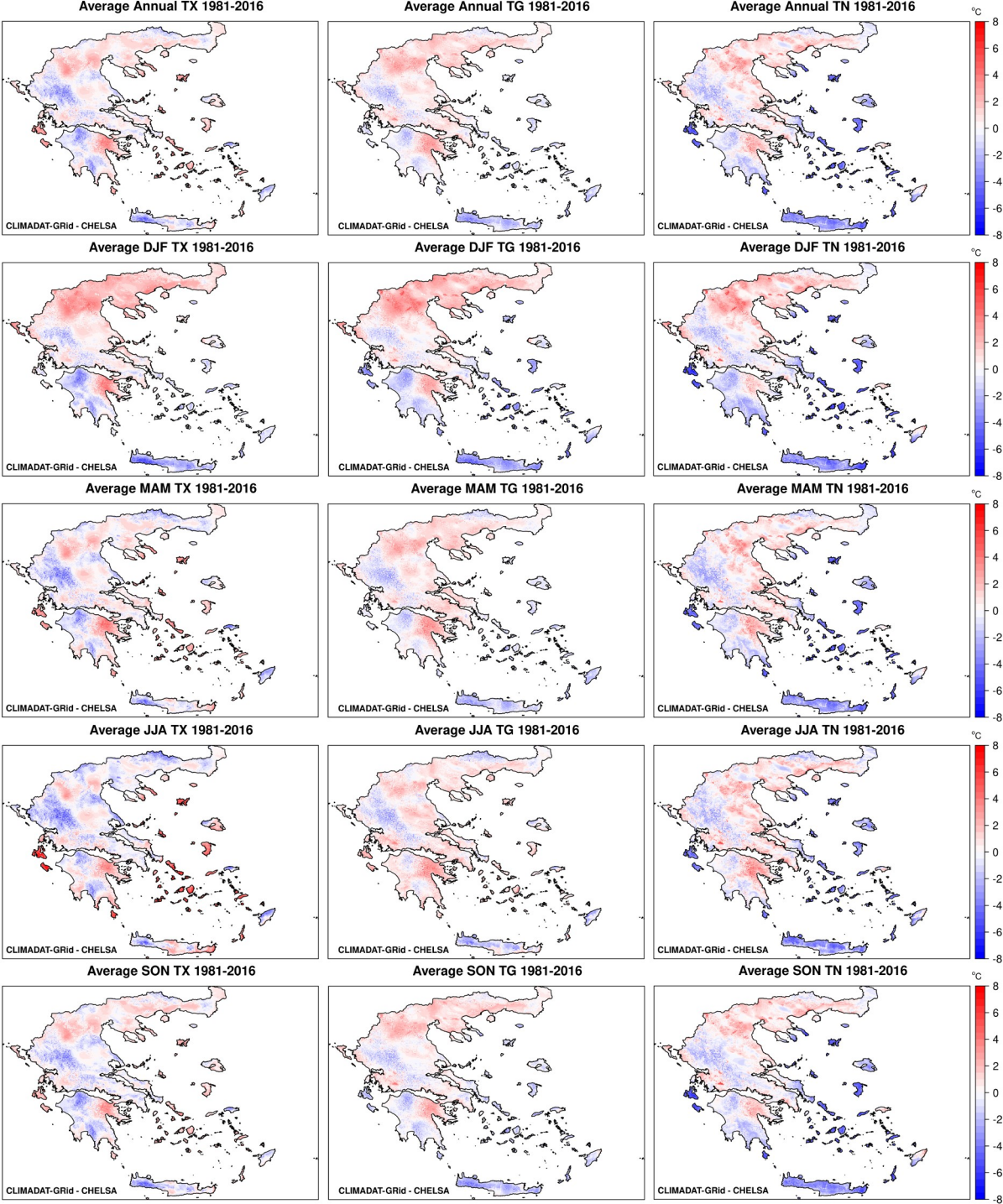


Fig S1. Average annual and Seasonal differences for TX, TG and TN between CLIMADATGRid and CHELSA for the period 1981-2016.

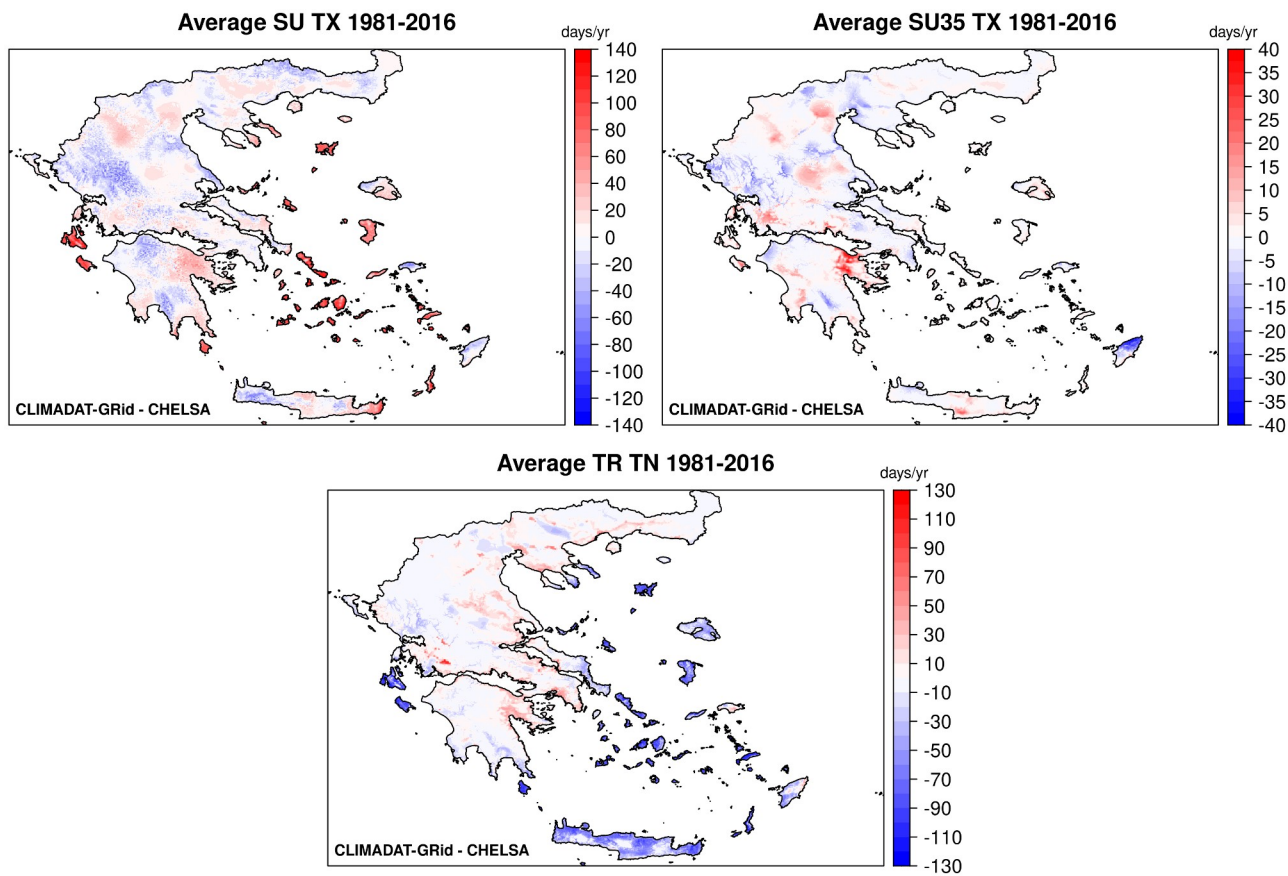


Fig S2. Average annual differences for SU, SU35 and TR between CLIMADATGRid and CHELSA for the period 1981-2016.

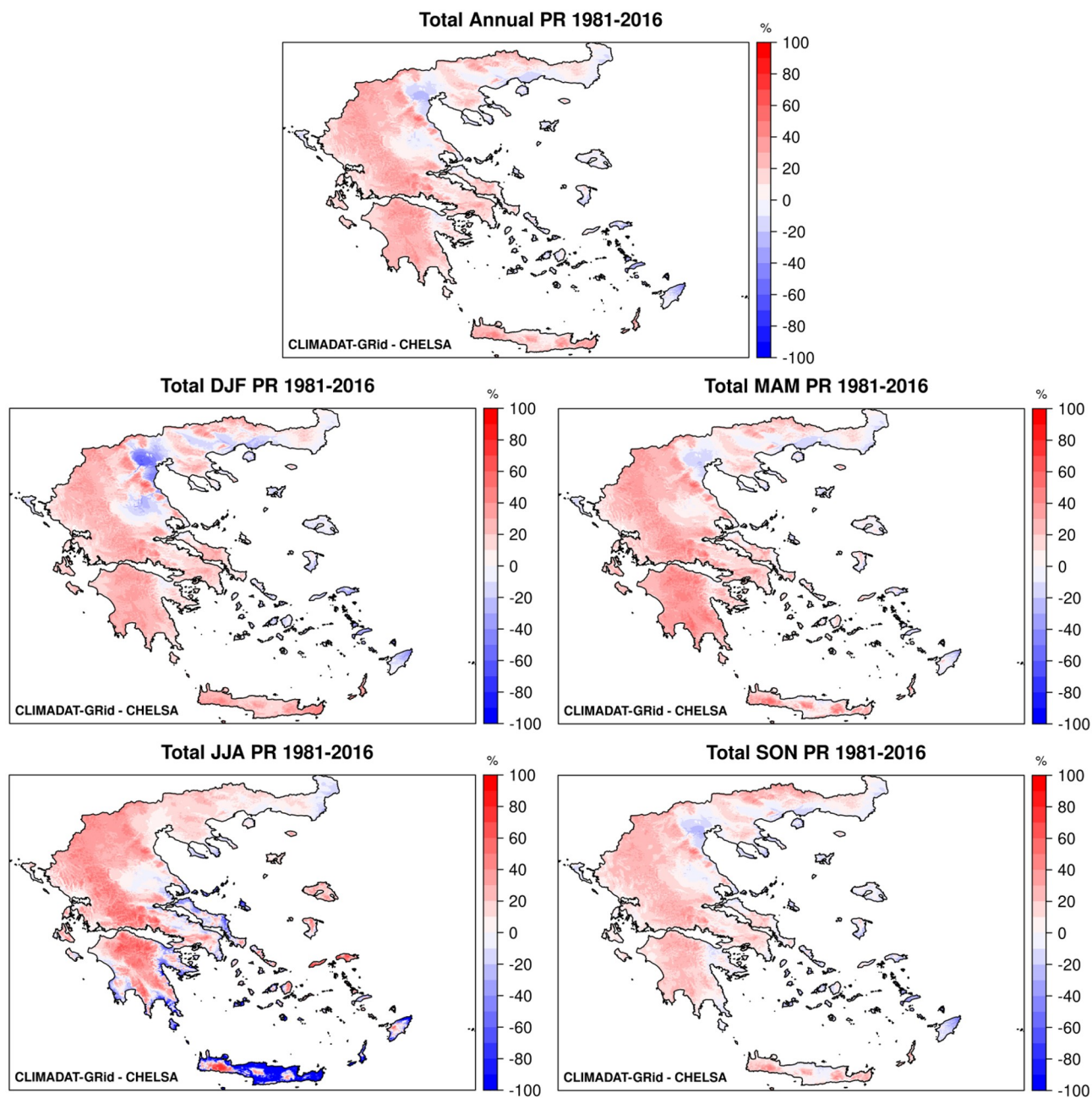


Fig S3. Average annual and seasonal relative differences in precipitation between CLIMADAT-Grid and CHELSA for the period 1981–2016.

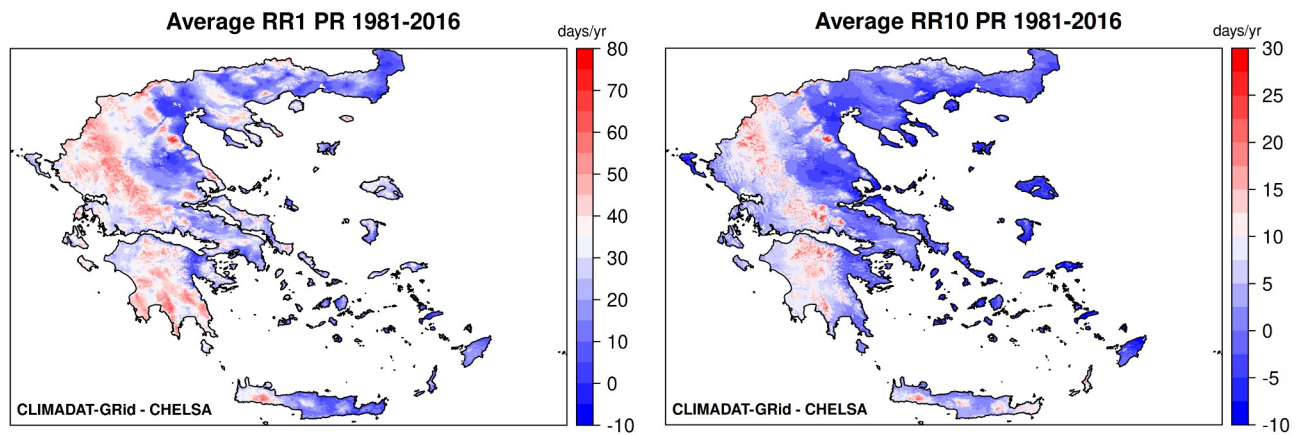


Fig S4. Average annual differences for RR1 and RR10 between CLIMADATGRid and CHELSA for the period 1981-2016.

New references included in the revised version of the manuscript and in the response letter.

Serrano-Notivoli, R., Beguería, S., Saz, M. Á., Longares, L. A., and de Luis, M.: SPREAD: a high-resolution daily gridded precipitation dataset for Spain—an extreme events frequency and intensity overview, *Earth System Science Data*, 9, 721–738, <https://doi.org/10.5194/essd-9-721-2017>, 2017.

Serrano-Notivoli, R., Beguería, S., and de Luis, M.: STEAD: a high-resolution daily gridded temperature dataset for Spain, *Earth System Science Data*, 11, 1171–1188. <https://doi.org/10.5194/essd-11-1171-2019>, 2019.

Serrano-Notivoli, R., Saz, M. Á., Longares, L. A., and de Luis, M.: SiCLIMA: High-resolution hydroclimate and temperature dataset for Aragón (northeast Spain). *Data in Brief*, 56, 110876, <https://doi.org/10.1016/j.dib.2024.110876>, 2024.

Fonseca, A. R., Santos, J. A.: High-Resolution Temperature Datasets in Portugal from a Geostatistical Approach: Variability and Extremes. *Journal of Applied Meteorology and Climatology*, 57, 627–644, <https://doi.org/10.1175/JAMC-D-17-0215.1>, 2018.

Krähenmann, S., Walter, A., Brienens, S., Imbery, F., and Matzarakis, A.: High-resolution grids of hourly meteorological variables for Germany, *Theoretical and Applied Climatology*, 131, 899–926, <https://doi.org/10.1007/s00704-016-2003-7>, 2018.

Hollis, D., McCarthy, M., Kendon, M., Legg, T., and Simpson, I.: HadUK-Grid—A new UK dataset of gridded climate observations, *Geoscience Data Journal*, 6, 151–159. <https://doi.org/10.1002/gdj3.78>, 2019.

Škrk, N., Serrano-Notivoli, R., Čufar, K., Merela, M., Črepinšek, Z., Kajfež Bogataj, L., and de Luis, M.: SLOCLIM: a high-resolution daily gridded precipitation and temperature dataset for Slovenia, *Earth System Science Data*, 13, 3577–3592. <https://doi.org/10.5194/essd-13-3577-2021>, 2021.

Doblas-Reyes, F. J., Sorensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey, B. L., Maraun, D., Stephenson, T. S., Takayabu, I., Terray, L., Turner, A., and Zuo, Z.: Linking global to regional climate change. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Pean, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekci, O., Yu, R., and Zhou, B. (eds.): *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021.

Serrano-Notivoli, R., and Tejedor, E.: From rain to data: A review of the creation of monthly and daily station-based gridded precipitation datasets, *Wiley Interdisciplinary Reviews: Water*, 8, e1555, <https://doi.org/10.1002/wat2.1555>, 2021.

Pantillon, F., Davolio, S., Avolio, E., Calvo-Sancho, C., Carrió, D. S., Dafis, S., Gentile, E. S., Gonzalez-Aleman, J. J., Gray, S., Miglietta, M. M., Patlakas, P., Pytharoulis, I., Ricard, D., Ricchi, A., Sanchez, C., and Flaounas, E.: The crucial representation of deep convection for the cyclogenesis of Medicane Ianos, *Weather and Climate Dynamics*, 5, 1187–1205, <https://doi.org/10.5194/wcd-5-1187-2024>, 2024.

Patlakas, P., Chaniotis, I., Hatzaki, M., Kouroutzoglou, J., and Flocas, H. A.: The eastern Mediterranean extreme snowfall of January 2022: synoptic analysis and impact of sea-surface temperature, *Weather*, 79, 25–33, <https://doi.org/10.1002/wea.4397>, 2024.

Politi, N., Vlachogiannis, D., Sfetsos, A., and Nastos, P. T.: High-resolution dynamical downscaling of ERA-Interim temperature and precipitation using WRF model for Greece, *Climate Dynamics*, 57, 799–825, <https://doi.org/10.1007/s00382-021-05741-9>, 2021.

Heinke, J., Müller, C., and Gerten, D.: Limitations of machine learning in a spatial context, *EGU General Assembly 2023*, Vienna, Austria, 24–28 Apr 2023, EGU23-16096, <https://doi.org/10.5194/egusphere-egu23-16096>, 2023