INTERCOMPARISON OF GRIDDED PRODUCTS: A MULTI-DIMENSIONAL PERSPECTIVE

Sixto HERRERA GARCÍA[1], Fidel GONZÁLEZ ROUCO[2], Roberto SERRANO-NOTIVOLI[3], Juan Luís GARRIDO[4], Santiago BEGUERÍA[5], José M. GUTIÉRREZ[6], Pere QUINTANA-SEGUÍ[7], Maialen ITURBIDE[6], Esteban RODRÍGUEZ[8] y Ana MORATA[8]

[1]Applied Mathematics and Computer Science Department, Universidad de Cantabria,
[2]Instituto de Geociencias (IGEO), Universidad Complutense de Madrid (UCM-CSIC),
[3]Dpto. de Geografía y Ordenación del Territorio, Instituto Universitario de Ciencias Ambientales (IUCA), Universidad de Zaragoza,
[4]División de Contaminación Atmosférica, Centro de Investigaciones Energéticas, Medioambientales Y Tecnológicas (CIEMAT),
[5]Estación Experimental de Aula Dei (EEAD), CSIC,
[6]Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria.
[7]Observatori de l'Ebre (OE), Universitat Ramon Llull – CSIC.
[8]Spanish Meteorological Agency (AEMET).

sixto.herrera@unican.es, fidelgr@ucm.es, roberto.serrano@unizar.es, juanluis.garrido@ciemat.es, santiago.begueria@csic.es, manuel.gutierrez@unican.es, pquintana@obsebre.es, miturbide@ifca.unican.es, erodriguezg@aemet.es, amoratag@aemet.es

ABSTRACT

Observational uncertainty has become an increasingly significant concern in climate studies. In this context, mainland Spain has several publicly available daily gridded datasets of precipitation, which have been utilized in various regional and national studies. Recent research, like the latest CLIVAR-Spain report, underscores the need for thorough intercomparison among these datasets, highlighting their respective strengths and limitations while discussing findings in light of existing observational uncertainties.

This study seeks to address this issue by performing a comprehensive analysis of 10 gridded datasets of daily precipitation currently available for mainland Spain. It is the first study to conduct an extensive comparison across these datasets, focusing on multiple dimensions: statistical distribution, extreme events and spells.

Our findings indicate that national datasets are more similar to each other than to global or European-wide datasets. While mean values primarily differ in intensity, these differences extend to spatial patterns when extremes or spells are considered, thereby expanding or reducing the affected regions. These variations remain mostly consistent with altitude, with most datasets exhibiting similar behavior, except for E-OBS v27e, EMO-1arcmin, and ERA-Land.

Key words: Interpolation, intercomparison, precipitation, maximum and minimum temperatures, Iberian Peninsula, mainland Spain

1. INTRODUCTION

In recent decades, the Spanish climate research community has made substantial efforts to develop historical datasets for essential meteorological variables, such as precipitation and temperature, across Spain, ensuring high spatial and temporal resolution. To achieve this, researchers have employed various methods, from interpolation-only techniques (SPREAD: Serrano-Notivoli et al., 2017; Iberia01: Herrera et al., 2019a; PTI-Clima: Beguería et al., 2025) to more sophisticated approaches that integrate observational data with outputs from numerical models, such as reanalyses or short-range weather forecasts (ROCIO-IBEB: Peral et al., 2017; SAFRAN and HUMID01: Quintana-Segui et al., 2017). Additionally, global datasets (ERA5-Land: Muñoz Sabater, 2019; CHELSA W5E5: Karger et al., 2023) and European datasets (E-OBS v27e: Cornes et al., 2018; regional reanalysis CERRA-SFC: El-Said et al., 2021; EMO-1arcmin: Thiemig et al., 2020) also cover the region, providing useful complements to the national data.

This variety of datasets introduces a new source of uncertainty in analyzing current climate conditions (Burton et al., 2018; Lledó et al., 2024; Newman et al., 2019; Tanarhte et al., 2012) and in calibrating and evaluating climate change projections (Kotlarski et al., 2019; Herrera et al., 2020). Intercomparison studies are essential to assess these datasets, highlighting their differences, limitations, characteristics, and appropriate applications to ensure accurate interpretation of results (Newman et al., 2019).

While most of these products are publicly accessible, there has not yet been a comprehensive intercomparison study assessing their ability to represent various precipitation regimes across the Iberian Peninsula. Existing studies (Quintana-Seguí et al., 2017; Herrera et al., 2019a; Herrera et al., 2019b) have partially addressed intercomparison, usually focusing on describing new datasets rather than thoroughly comparing available products and evaluating the relevant characteristics of each dataset.

This study aims to address this gap and build upon the work by Herrera et al. (2024) on minimum and maximum temperatures, by focusing on precipitation and extreme precipitation events as well as maximum and minimum temperatures. Section 2 outlines the methods and datasets used, Section 3 presents the primary results, and Section 4 discusses the conclusions and implications.

2. METHODS

2.1. Datasets

Table 1 summarizes the main properties of the ten datasets analyzed in this study. These datasets represent three primary methodological approaches:

1. Observation-based interpolation (O), which uses statistical or geostatistical methods to interpolate observed data.
2. Analysis-type (A), which combines observations with numerical model outputs from reanalyses or short-range weather forecasts through optimal interpolation.

3. Reanalysis-based (R), which consists of direct outputs from global or regional reanalyses or calibrated reanalysis products.

Name	Spatial Res.	Temporal Res.	Referencia
SPREAD (O)	5 km	Daily (1950-2012)	Serrano-Notivoli et al. 2017
E-OBS v27e (O)	0.1º	Daily (1951-Present)	Cornes et al. 2018
Iberia01 (O)	0.1º	Daily (1971-2015)	Herrera et al. 2019[a]
PTI-Clima (O)	0.025º	Daily (1961-2022)	Beguería et al. 2025
HUMID01 (A)	0.05º	Hourly (1979-2017)	Quintana-Seguí et al. 2017
ROCIO-IBEB (A)	5 km	Daily (1951-2022)	Peral et al. 2017
ERA5-Land (R)	0.1º	Hourly (1950-Present)	Muñoz Sabater, 2019
CHELSA-W5E5 (R)	0.01º	Daily (1979-2016)	Karger et al. 2023
EMO-1arcmin (R)	1.5 km	Daily (1990-2022)	Thiemig et al. 2020
CERRA-SFC (R)	5.5 km	Hourly (1984-2021)	El-Said et al. 2021

Table 1: Gridded datasets considered over the mainland Spain and their main properties: name (approach), spatial resolution, temporal resolution and coverage, and Reference. Global and/or European wide datasets have been highlighted in bold. The datasets have been grouped according to the methods used to build them: observational interpolation (O), analysis-type (A) and reanalysis-based (R) products. Source: Herrera et al., 2024.

2.1.a. Data Availability

The SPREAD datasets are publicly accessible through the Digital.CSIC repository at https://doi.org/10.20350/digitalCSIC/7393 (Serrano-Notivoli et al., 2016). E-OBS v27e is available through the Copernicus Climate Data Store (Cornes et al., 2020), and Iberia01 is distributed under the Open Database License here: http://hdl.handle.net/10261/183071 (Herrera et al., 2019b). The ROCIO-IBEB dataset can be accessed via AEMET's data server at https://www.aemet.es/es/serviciosclimaticos/cambio_climat/datos_diarios?w=2.

ROCIO-IBEB serves as a reference dataset in this study, as it is developed, maintained, and utilized by AEMET, Spain's official authority for meteorological information and services.

The EMO-1arcmin dataset, currently supported, is provided under a Creative Commons Attribution 4.0 International license and is accessible via the public repository at https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/CEMS-EFAS/meteorological_forcings/EMO-1arcmin/. ERA5-Land is also available through

the Copernicus Climate Data Store at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land (Hersbach et al., 2023), following Copernicus product terms. The CHELSA-W5E5 dataset is accessible through the ISIMIP server at https://chelsa-climate.org/ under the CC0 1.0 Universal Public Domain Dedication license. CERRA-SFC is available via the Copernicus Climate Data Store at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-cerra-single-levels, and this study uses the single-level subset, including both atmospheric and surface quantities. The precipitation forecast for CERRA-SFC was obtained at the analysis time of 00:00 with a 6-hour spin-up, calculating the difference between the 30-hour and 6-hour lead times.
While SAFRAN-Spain (Quintana-Segui et al., 2017) is publicly available for research purposes at https://www.obsebre.es/en/en-safran, the HUMID01 dataset has not yet been published, and the PTI-Clima dataset remains undisclosed to the public.

2.2. Data Homogenization
As can be seen in Table 1, each dataset has its own temporal and spatial resolution, including variations in map projections and temporal references. Therefore, all the datasets have been homogenized defining a common grid and time reference to enable effective comparative analysis.
Due to the different nature and temporal resolution of the nine datasets, the differences between them are not only on the spatial dimension but also on the temporal reference, so the common period of the 10 datasets has been considered: 1990-2012.
To ensure a fair comparison between the datasets, we established a common grid that none of the datasets had originally used. We opted for a coarser regular grid covering mainland Spain with a resolution of 0.25º. All datasets were then redefined to fit this grid using a conservative interpolation scheme (Zhuang et al., 2023).

Code	Description	Units	Dimension
pr_{clim}	Mean of daily values	mm	SD
$pr_{4\text{-}6Nov}$	Total precipitation amount in event 4-6 Nov 1997	mm	EE
lds	Average Maximum Length of Dry Spells	days	SS

Table 2: Indices considered for the validation of the different datasets. The reference considered has always been ROCIO-IBEB. The var code refers to the variable considered to obtain the index. SD: Statistical distribution; EE: Extreme events; SS: Spells. Source: Created by the authors for this contribution.

2.3. Intercomparison Parameters
Table 2 outlines the parameters considered, along with the dimensions evaluated for each. To assess the statistical significance of differences in mean values and distribution shapes, the t-test and the Kolmogorov-Smirnov test were used for comparing two samples. Additionally, the percentage of grid points with statistically

significant differences relative to the reference dataset, ROCIO-IBEB, was calculated at a 95% confidence level. For the Kolmogorov-Smirnov test, time series were centered to isolate distribution shape effects from mean differences, which are explicitly addressed by the t-test.

In order to analyze the extreme regime, an extreme precipitation event that occurred in the Iberian Peninsula in November 1997 has been used to evaluate how each dataset reproduces the intensity and spatial pattern, and better understand the uncertainties related to how different datasets can reproduce extreme events. In order to complete the evaluation, a index of spells (lds: average maximum length of dry spells) has been considered, defined as the average of the annual maximum length of consecutive dry days (precipitation ≥ 1 mm).

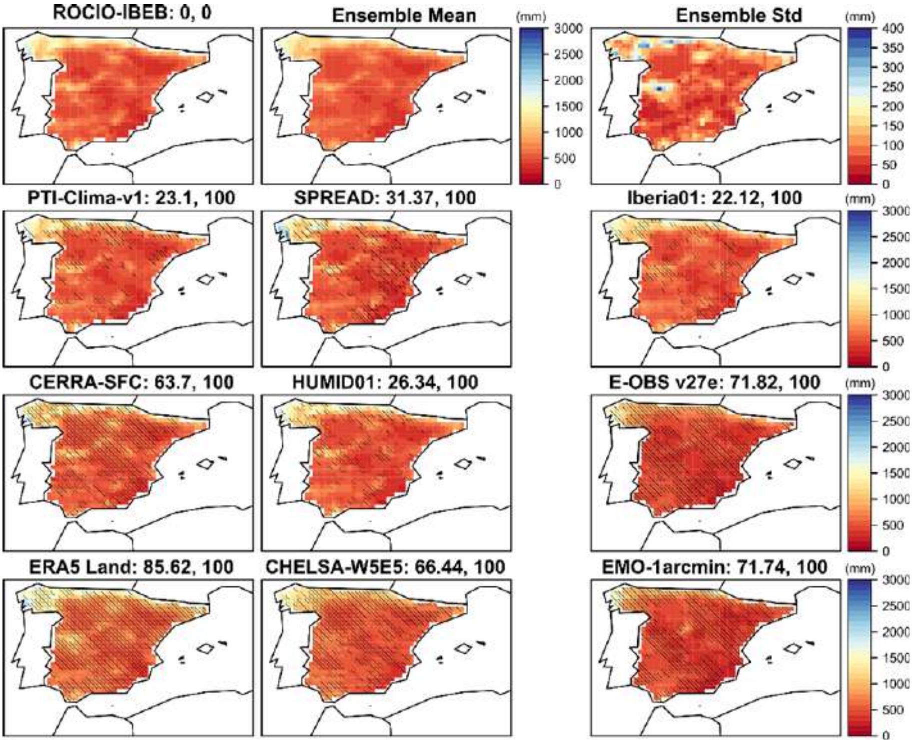

Figure 1. Climatology of anual precipitation for the period 1990-2012 (mm). Grid-points with statistically significant differences (95%) in the mean w.r.t. ROCIO-IBEB according to the t-test results have been hatched. The percentage of the grid-points is also shown for each dataset in the title. Source: Created by the authors for this contribution.

3. RESULTS

The intercomparison results among the datasets are presented below. In all figures, the first row of maps shows the parameter for the reference dataset (ROCIO-IBEB), along with the ensemble mean and standard deviation of the remaining nine datasets. The subsequent rows display the parameter for each of these nine datasets individually.

Figure 1 shows the mean annual precipitation and the percentage of grid cells with statistically significant (95%) differences between both the means and the shapes of the distributions of each dataset compared to the reference, ROCIO-IBEB. Notably, while mean differences are minimal in some cases, higher-moment differences affect the entire domain, as indicated by the Kolmogorov-Smirnov test. Global and European datasets exhibit more substantial differences than the national datasets. Although all datasets show a common spatial pattern, with maximum precipitation in the northern regions and mountain ranges of the Iberian Peninsula, the main discrepancies lie in precipitation amounts, leading to substantial regional variability among datasets.

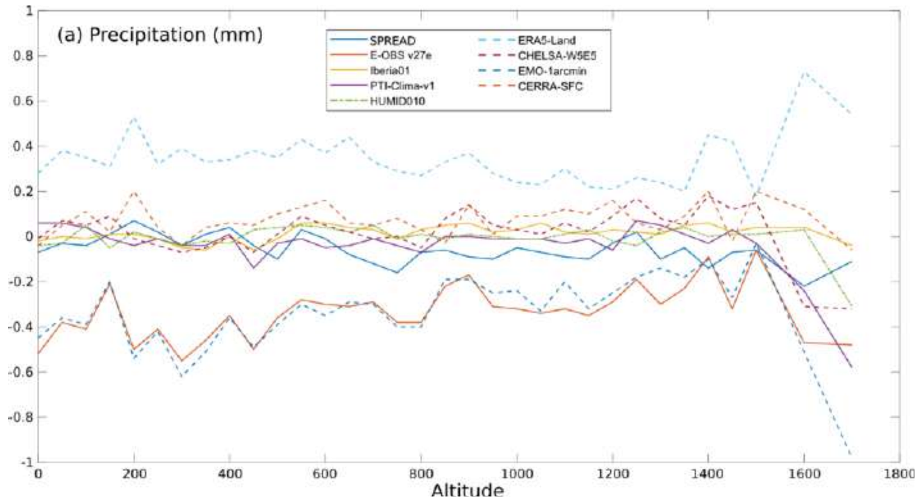

Figure 2: Mean spatial bias depending on the grid cell height for precipitation. All the bias have been obtained using ROCIO-IBEB as reference. Source: Created for this contribution.

When analyzing these differences by altitude (Figure 2), we observe that, although there are some fluctuations, the differences remain relatively stable up to the highest elevations, where they become more pronounced in one direction or another. Most datasets exhibit similar behavior, with the exceptions of E-OBS v27e and EMO-1arcmin, which tend to underestimate precipitation, and ERA-Land, which distinctly overestimates precipitation.

Figure 4 depicts an extreme precipitation event over the Iberian Peninsula in November 1997, characterized by a diagonal pattern of intense rainfall from the southwest to central Iberia, as well as high precipitation in the north, from Asturias to the Pyrenees. With the exception of E-OBS v27e, all interpolation and analysis-type datasets closely reproduce this spatial pattern and intensity. While E-OBS v27e captures part of the spatial pattern, it underestimates intensity due to its less dense observational network compared to other interpolated datasets. Among the reanalysis-based datasets, some pattern displacement results in a dipole effect, with areas of overestimation and underestimation across the peninsula. EMO-1arcmin shows high values in southwestern mainland Spain but mainly concentrated there, while ERA5-Land and CHELSA W5E5 both underestimate the event, capturing parts of the spatial pattern but not the full intensity.

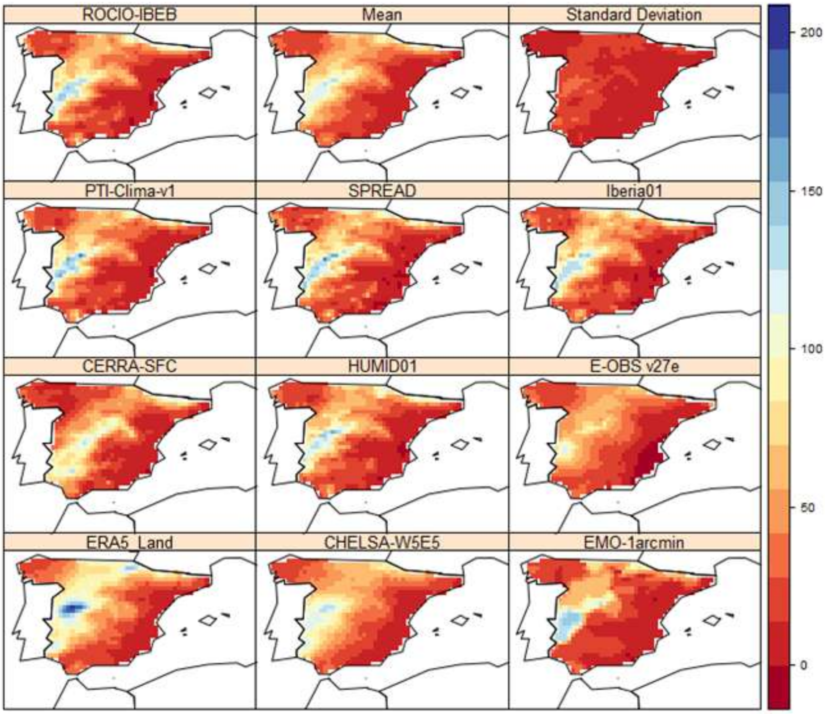

Figure 4. Extreme event of precipitation occurred between November 4th and 6th 1997 (mm). Source: Created by the authors for this contribution.

Figure 5 shows the index *lds* with the average maximum length of dry spells reflecting two groups of datasets based on their spatial pattern. On one hand, most of the datasets have an incremental gradient from northwest to south-southeast reaching the maximum value in Almeria and the Guadalquivir river basin, in the south-southeastern coast of the Iberian Peninsula. Other datasets (e.g. CERRA-SFC, ERA5

Land or E-OBS v27e) present a more homogeneous pattern with lower values than the rest and/or with a more continous pattern on the south.

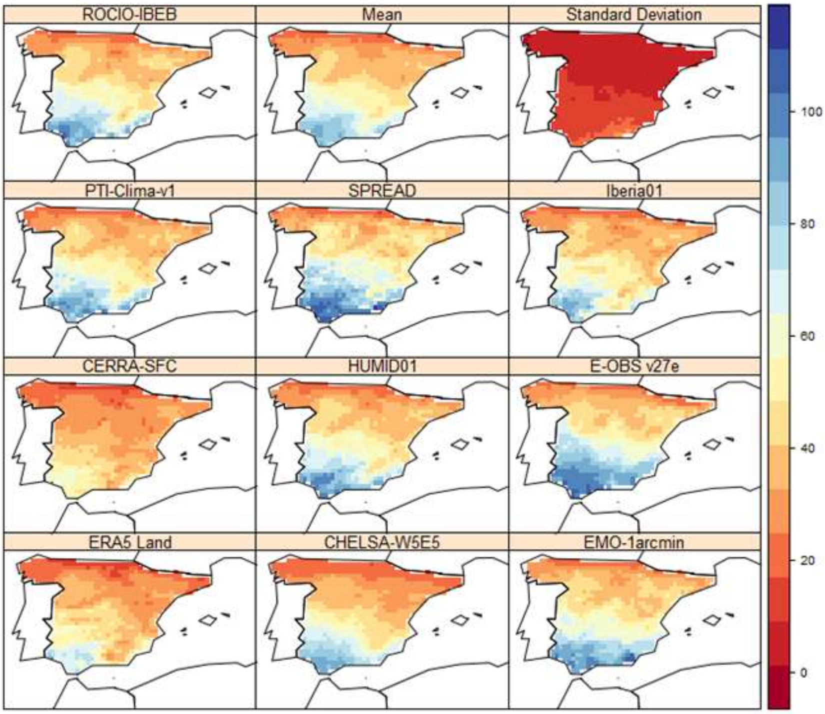

Figure 5. Average Maximum Length of Dry Spells for the period 1990-2012. Source: Created for this contribution.

4. DISCUSIÓN

This study addresses the need for an intercomparison analysis of various existing daily precipitation datasets to assess their strengths, limitations, potential applications, and possible misuses. The findings reveal that, unlike the temperature results shown in Herrera et al. (2024), national datasets tend to be more consistent with each other than with global or European datasets. While the main differences in mean values are related to intensity, these differences also extend to spatial patterns when examining extremes or precipitation spells, thereby expanding or contracting the affected regions.

Notably, the differences remain largely consistent with altitude, with most datasets displaying similar behavior. However, certain exceptions, such as E-OBS v27e, EMO-1arcmin, and ERA-Land, highlight issues potentially stemming from factors beyond the chosen methodology, such as the observational network used in constructing the dataset.

AGRADECIMIENTOS
This research work was funded by the European Commission – NextGenerationEU (Regulation EU 2020/2094), through CSIC's Interdisciplinary Thematic Platform Clima (PTI Clima)/ Development of Operational Climate Services. RSN is funded by grant RYC2021-034330-I funded by MCIN/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR". J.M.G. and S.H. acknowledge the support of the Spanish Government through the Agencia Estatal de Investigación (project PID2019-111481RB-I00 - Contribución española al Atlas del IPCC-AR6; Desarrollo y problemas científicos). We acknowledge the data providers in the ECA\&D project. Data and metadata available at \url{https://www.ecad.eu}.

REFERENCIAS

Beguería, S., and colabs. (2025). Undefined. Undefined (In preparation)

Burton C, Rifai S, Malhi Y., (2018), Inter-comparison and assessment of gridded climate products over tropical forests during the 2015/2016 El Niño. Philos Trans R Soc Lond B Biol Sci., 373 (1760), 20170406. doi: 10.1098/rstb.2017.0406

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D., (2018), An ensemble version of the E-OBS temperature and precipitation data sets. Journal of Geophysical Research: Atmospheres, 123, 9391-9409. doi: https://doi.org/10.1029/2017JD028200

Cornes, R., Schrier, G., Besselaar, E.J.M., Jones, P., (2020), E-OBS daily gridded meteorological data for europe from 1950 to present derived from in-situ observations. Technical report, Copernicus Climate Change Service, Climate Data Store, doi: https://doi.org/10.24381/cds.151d3ec6

El-Said, A., Brousseau, P., Ridal, M., Randriamampianina, R., (2021), A new temporally flow-dependent eda estimating background errors in the new copernicus european regional re-analysis (CERRA). Earth and Space Science Open Archive, 1-28, doi: https://doi.org/10.1002/essoar.10507207.1

Herrera, S., Cardoso, R.M., Soares, P.M.M., Espírito-Santo, F., Viterbo, P., Gutiérrez, J.M. (2019). Iberia01: a new gridded dataset of daily precipitation and temperaturas over Iberia. Earth System Science Data, 1947–1956, doi: https://doi.org/10.5194/essd-11-1947-2019

Herrera, S., Cardoso, R.M., Soares, P.M.M., Espírito-Santo, F., Viterbo, P., Gutiérrez, J.M., (2019), Iberia01: Daily gridded (0.1 resolution) dataset of precipitation and temperatures over the Iberian peninsula [dataset], DIGITAL.CSIC, doi: https://doi.org/10.20350/digitalCSIC/8641

Herrera, S., Soares, P.M.M., Cardoso, R.M., Gutirrez, J.M. (2020). Evaluation of the Euro-CORDEX regional climate models over the iberian peninsula: Observational uncertainty analysis. Journal of Geophysical Research: Atmospheres, 125 (12), 2020-032880, doi: https://doi.org/10.1029/2020JD032880

Herrera, S., González Rouco, F., Serrano-Notivoli, R., Garrido, J.L., Beguería, S., Gutiérrez, J.M., Quintana-Seguí, P., Iturbide, M., Rodríguez, E., Morata, A. (2024), Intercomparison of daily maximum and mínimum temperature gridded products over mainland, Submitted to Climate Change.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Hornyi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thpaut, J.-N., (2023), ERA5 hourly data on single levels from 1940 to present. Technical report, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), doi: https://doi.org/10.24381/cds.adbb2d47

Karger, D.N., Lange, S., Hari, C., Reyer, C.P.O., Conrad, O., Zimmermann, N.E., Frieler, K. (2023). CHELSA-W5E5: daily 1 km meteorological forcing data for climate impact studies. Earth System Science Data, 15, 2445-2464, doi: https://doi.org/10.5194/essd-15-2445-2023

Kotlarski, S., Szab, P., Herrera, S., Rty, O., Keuler, K., Soares, P.M., Cardoso, R.M., Bosshard, T., Pag, C., Boberg, F., Gutiérrez, J.M., Isotta, F.A., Jaczewski, A., Kreienkamp, F., Liniger, M.A., Lussana, C., Pianko-Kluczyska, K. (2019). Observational uncertainty and regional climate model evaluation: A pan-european perspective. International Journal of Climatology, 39 (9), 3730–3749, doi: https://doi.org/10.1002/joc.5249

Lledó, L., Haiden, T., Chevallier, M. (2024). An intercomparison of four gridded precipitation products over Europe using the three-cornered-hat method. EGUsphere ([preprint]), doi: https://doi.org/10.5194/egusphere-2024-807

Muñoz Sabater, J. (2019) ERA5-Land hourly data from 1950 to present. Technical report, Copernicus Climate Change Service, Climate Data Store, doi: https://doi.org/10.24381/cds.e2161bac

Newman, A.J., Clark, M.P., Longman, R.J., Giambelluca, T.W. (2019). Methodological intercomparisons of station-based gridded meteorological products: Utility, limitations, and paths forward. Journal Of Hydrometeorology, 20, 531-547, doi: https://doi.org/10.1175/JHM-D-18-0114.1

Peral, C., Navascués, B., Ramos, P. (2017). Serie de precipitación diaria en rejilla con fines climáticos. Technical Report 24, AEMET. https://www.aemet.es/documentos/es/conocermas/recursos en linea/publicaciones y estudios/publicaciones/NT 24 AEMET/NT 24 AEMET.pdf (In Spanish)

Quintana-Seguí, P., Turco, M., Herrera, S., Míguez-Macho, G. (2017). Validation of a new SAFRAN-based gridded precipitation product for spain and comparisons to Spain02 and ERA-Interim. Hydrology and Earth System Sciences, 21 (4), 2187–2201, doi: https://doi.org/10.5194/hess-21-2187-2017

Serrano Notivoli, R., De Luís, M., Beguería, S., & Saz, M. Á., (2016), SPREAD (Spanish PREcipitation At Daily scale) [Dataset]. Digital.CSIC, doi: http://doi.org/10.20350/DIGITALCSIC/7393

Tanarhte, M., Hadjinicolaou, P., Lelieveld, J., (2012), Intercomparison of temperature and precipitation data sets based on observations in the mediterranean and the middle east. Journal of Geophysical Research: Atmospheres, 117 (D12), doi: https://doi.org/10.1029/2011JD017293

Thiemig, V., Ramos Gomes, G.N., Skien, J.O., Ziese, M., Rauthe-Schch, A., Rustemeier, E., Rehfeldt, K., Walawender, J., Kolbe, C., Pichon, D., Schweim, C., Salamon, P., (2020), EMO-1arcmin: A high-resolution multi-variable gridded meteorological data set for Europe (1990-2021). European Commission, Joint Research Centre (JRC) [Dataset], doi: https://doi.org/10.2905/0BD84BE4-CEC8-4180-97A6-8B3ADAAC4D26

Zhuang, J., Dussin, R., Huard, D., Bourgault, P., Banihirwe, A., Raynaud, S., Malevich, B., Schupfner, M., Fernandes, F., Levang, S., J¨uling, A., Almansi, M., Scott, O.Z.R., Rondeau, G., Rasp, S., Smith, T.J., Stachelek, J., Plough, M., Manchon, P., Bell, R., Li, X., (2023), xESMF: Universal Regridder for Geospatial Data. Zenodo, doi: https://doi.org/10.5281/zenodo.7800141