

We warmly thank the Editor and the Referees for their careful reading of the manuscript and their valuable comments. Please find below our answers (in italic) to all the items raised, and the tracked-changes version, to whom the line numbers refer.

Please note that in the revised version of the paper, numbering and positioning of the figures were reorganized, joining some graphs that were previously presented separately, with the aim to include more results (especially about drought characteristics, as required by Anonymous Referee #2) and increase remarkably the information content of the paper, keeping its overall structure as clear as possible.

Anonymous Referee #1

RC: This paper quantifies the performances of several combinations of regional climate models (RCMs) driven by different general circulation models (GCMs) in two regions of southern Italy. The GCM-RCM combinations are part of the Coordinated Regional Climate Downscaling Experiment (CORDEX) initiative in the European domain (EURO- CORDEX). Performances are evaluated in the ability to capture the spatial variability of mean annual and seasonal precipitation (P) and temperature (T), as well as three drought metrics derived by applying the run theory. I enjoyed reading this paper, which is well written; presents rigorous analyses; critically discusses the results; and has practical utility for impact studies in the study regions, since it provides a list of best GCM-RCM combinations. I recommend its publication and I only have a few minor requests and suggestions.

AC: *We thank the referee for the positive comments.*

RC: The authors should provide more details on how they applied the principal component analysis (PCA): Was it applied on monthly or annual P? The PCA returns spatial patterns that explain most of the P variability. How did the authors derive the subregions within these spatial patterns? This should be clarified.

AC: *We thank the referee for this comment, which allowed us to improve the description of the PCA. At this regard, further information to the manuscript was added as follows:*

Author's changes to the manuscript (LL 168-201): "The precipitation patterns obtained by the interpolation procedure were analyzed adopting a methodology based on the Principal Component Analysis (PCA) to distinguish zones with rather independent climatic variability within the area under investigation. PCA is a well-known statistical tool used to transform an original set of intercorrelated variables into a reduced number of new linearly uncorrelated ones explaining most of the total variance (Rencher, 1998). The latter, derived as linear combinations of the original variables, are the principal components (PCs), while the coefficients of the linear combinations are the loadings, which in turn represent the weight of the original variables in the PCs. From a procedural standpoint, PCA consists of solving an eigenvalue-eigenvector problem applied to the covariance matrix. The eigenvectors, properly normalized, are the loadings of the principal components, while the eigenvalues provide a measure of the total variance explained by each loading (Bordi and Sutera, 2001 and references therein). Under this decomposition, the loadings represent the correlation between the associated PCs and observed time series. Mapping the loading patterns of each PC among the ones selected, based on the percentage of the total explained variance of the process, largely allow to identify independent climatic areas within the study region. Moreover, it may be useful to apply a rotation operation to the eigenvectors, so that the corresponding loadings are more spatially localized. In other words, the

rotation leads to loadings with a high correlation with a smaller set of spatial variables and a low correlation with the remaining variables. Here, only orthogonal rotations were considered, computed by the varimax algorithm in Matlab® R2016. Clearly, each rotated pattern will not explain the same variance of the unrotated one, although the total variance explained remains unchanged.

In the present study, the first nine rotated PCs both at the annual and seasonal (DJF, MAM, JJA, SON) scales were investigated. Regardless of the considered period, the selected PCs always explain more than 78% of the total variance, with a maximum of 85% in the winter season (DJF). The loading patterns of these rotated PCs were mapped for each considered period to identify climatically homogeneous regions. Homogeneous sub-regions were detected at the annual scale and in autumn and winter seasons, whereas a confusing picture arose in spring and summer seasons. Furthermore, since about 75% of the total annual rainfall of the case-study area occurs between autumn and winter (as a result of cyclonic storms), the climatically homogeneous sub-regions identified at the annual scale approximately overlap with the ones identified at seasons SON and DJF. Isolated grid cells showing a different PC correspondence with respect to the surrounding cells, were manually corrected to simplify the delimitation of the homogeneous sub-regions. This approach led in dividing the whole area into six climatically homogenous zones, three for Sicily and three for Calabria (Fig. 1), for which separate performance assessments were carried out. Concerning Sicily region, the three identified sub-regions roughly coincide with the ones detected by Bonaccorso et al. (2003), who investigated the spatial variability of droughts in Sicily region based on SPI series computed on monthly precipitation observed in traditional rain gauges and on NCEP/NCAR reanalysis data from 1926 to 1996. In particular, three distinct areas, namely North-Eastern (identified in the PCA as zone 5, Fig. 1b), South-Central Eastern (zone 4), and Central-Western (zone 1), were identified. In Calabria, three main zones were also determined, namely North-Western (zone 2), North-Eastern (zone 3) and South-Eastern (zone 6), broadly corresponding to climatic homogenous areas found in previous studies (e.g., Versace et al., 1989). Interestingly, the South-Western tip of Calabria is identified as a part of a broader area (zone 5) extending over the North-Eastern Sicily.”

RC: When ranking the models based on performance in reproducing annual P, the authors find that nine models have similar error metrics. Have they tried to compute the mean rank of each model across the zones and even across the five time scales considered (annual and the four seasons)? There may be some models that are consistently in the top (lowest ranks) and these should be mentioned.

AC: This aspect was already partially investigated in a previous version of the manuscript, where we pointed out that the models Had-RACM, ECE-CCLM and Had-CCLM have the overall better performances at both annual and seasonal time scales. Thus, we have enlarged and made clearer the discussion in the revised version of the manuscript, adding also a figure (new Fig. 15) highlighting deviations in the performances of some models (e.g., CM5-ALAD), considering both the annual scale and the average behaviour at the seasonal scale (the higher the deviation, the higher the distance from the bisector).

Author's changes to the manuscript (LL 559-565): “Figure 15 shows a comparison between the ranking of interannual variability of annual precipitation and the average position in the ranking of seasonal precipitation. It highlights possible deviations of the performances of the models at different time scales (the higher the deviation, the higher the distance from the bisector). When considering the seasonal scale, the reduced performance of CM5-ALAD is evident, such as the better ranking of MPI-CCLM. In general, the best models both at the interannual and the seasonal scale, are Had-RACM and ECE-CCLM, followed by the two versions of ECE-RACM and two other CCLM models (namely, MPI-CCLM and Had-CCLM, the latter being penalized by the relatively lower ranking in winter).”

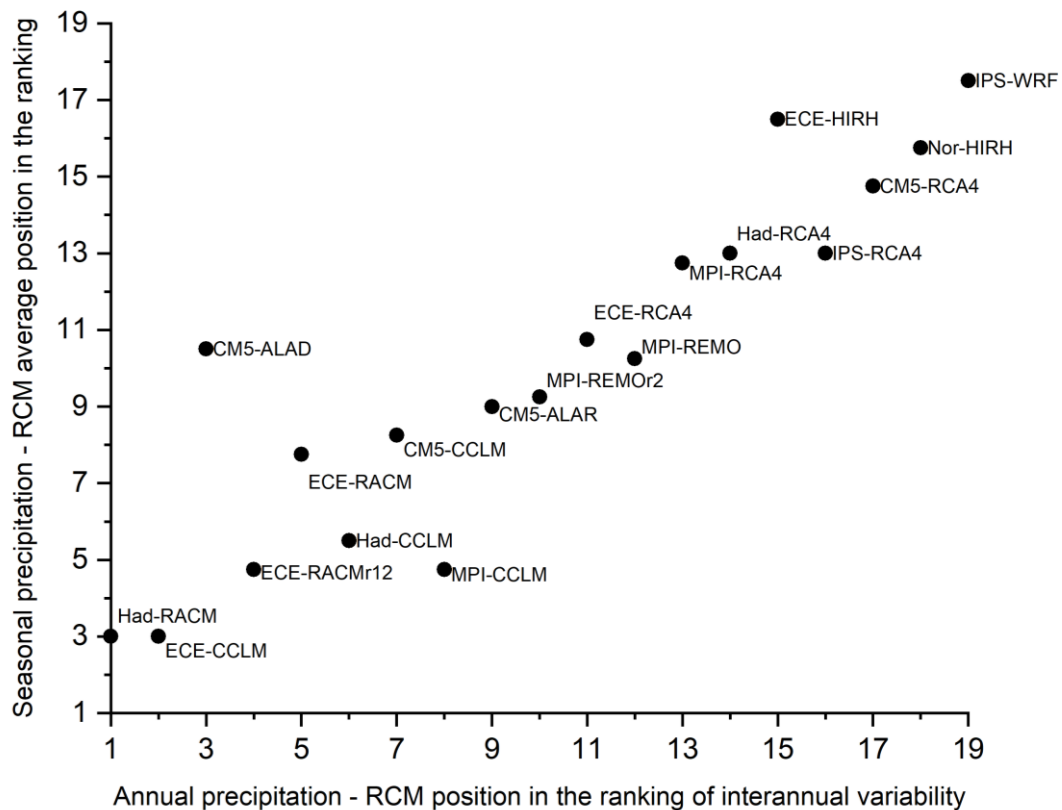


Figure 15. Comparison between the RCM position in the ranking of interannual variability of annual precipitation versus the average position in the ranking of seasonal variability of seasonal precipitation. Data concerns the whole study area (Calabria and Sicily).

RC: Line 45: Extremes occur everywhere. I suggest changing to "... occurrence of particularly intense extreme events, ...". If this is what the authors mean, a reference is also needed.

AC: The sentence was changed according to the suggestion.

Author's changes to the manuscript (LL 48-50): "... and is characterized by the occurrence of particularly intense extreme eventss, such as prolonged droughts and high-intensity storms leading to floods (Bonaccorso et al., 2013; Bonaccorso et al., 2015a and 2015b; Llasat et al., 2016; Senatore et al., 2020)."

RC: Line 55: CMIP5 has been already defined; just use the acronym.

RC: Line 61: I suggest adding "historical" before simulations.

RC: Line 326: it should be "show".

AC: The text was revised accordingly.

Anonymous Referee #2

RC: The paper is well written, the state-of-art well described in the introduction, and the methodology used in this study are clear and can be easily understood from the paper. Overall, the quality of this paper is good, but to be honest I do not think this should be a paper. I mean, I see it more as a technical report or, even better, as the preliminary part of a wider study, maybe from the selection of the best models to dedicated projections of hazard and impacts. I am aware that other studies dealing on the evaluations of newest GCM-RCM simulations do exist, focusing on small regions, e.g. the one cited by authors about Sardinia, but I feel that this is not a research paper, but a (very well performed) study on the performance of models on a test region. Thus, I am questioning myself: once the authors have decided that one combination of GCMs-RCMs performs better than the others, for each quantity analyzed (precipitation, temperature, drought), time scale (annual, seasonal), sub-region (3 for Sicily, 3 for Calabria).. what shall the reader do with this information inserted in a scientific paper? The region is very small, so - as the authors say (see lines 452-453) - the choice of the best model depends on many factors, making this piece of work not conclusive. What shall be really of interest is what the authors plan for further analyses (Lines 454-456). I also have another major point about the possible publication: the paper is not about droughts. Drought is just slightly touched and with very basic metrics, far from the current standard in drought-related analyses, so my final verdict is to reject this submission. However, I see that authors made great efforts, so they might consider to rethink about the paper and try to resubmit, but I would definitely remove the word droughts from title.

AC: *We thank the referee for the attention devoted to our study and his partial appreciation of the manuscript. As the referee mentioned, many papers deal only with the evaluation of RCM historical simulations and do not include assessment of future impacts of climate change, as confirmed by the bibliographic review in Table I. Nonetheless, one of the distinguishing features of our study compared to the literature on the subject, is the high density of temperature and precipitation ground-based stations available in the case-study region. Besides, the target region is representative of one of the main hot-spots for climate change – the Mediterranean Basin. Concerning its spatial extent (about 40,540 km²), it should be pointed out that our interest lies in the implementation of RCMs for climate change impact studies and hydrological applications at small spatial scale regions with a complex topography (see LL 51-54). To this end, it is particularly important to test the RCMs' skills in encompassing surface heterogeneities and mesoscale atmospheric processes at the considered spatial scale. We agree that the choice of the best model depends on many factors. That's exactly why our study intends to provide indications on the best model to choose based on the variable, the time and the spatial scale considered.*

Moreover, it is worth highlighting the novelties introduced by the methodological approach, which adopts both PCA for identifying sub-regions in the analyzed area and proposes hybrid rankings involving precipitation, temperature, and drought characteristics.

Finally, a comprehensive evaluation of RCMs is an important resource for readers and potential users of the RCM data. There are several ways to use this information, and the authors will not surely cover all possible ways. So, we want to provide the readership with a tool that they can use for specific purposes. From our part, we notice that this study could be useful for hydrological applications, where the use of a limited but properly selected set of models can help to avoid unnecessary computational burden, or for other high-temporal resolution applications, where information about models' performance allow the user to narrow down the search domain for the most suitable projections.

Regarding the drought analysis, we agree with the reviewer that more analyses were needed. Therefore, following his/her suggestions, we extended the analysis to the seasonal data, likewise the investigation on

precipitation and temperature, and to the return period of drought duration as well (new Section 4.3 and related discussion in improved Sections 5.1 and 5.2). The effects of the improved drought analysis also influence the overall ranking (Section 5.4)

RC: Why not using also Med-Cordex?

AC: *We could include Med-CORDEX data in our study. However, only a couple of models are currently available at the resolution used in this study (0.11°), thus we decided to focus on EURO-CORDEX only.*

RC: Are the Euro-CORDEX bias-adjusted? Why not using the bias-adjusted runs?

AC: *The EURO-CORDEX data in our study are not bias-adjusted. This is because the bias-adjustment is usually based on observed data (as a calibration procedure) and is particularly useful when RCMs are used for future projections. However, future projections are out of the scope of the present study, which addresses the evaluation of historical climate models simulations. The basic idea behind an evaluation study is to analyze the models' skill in simulating hydro-climatic processes against observations, rather than to correct the simulations as it is, for instance, required in the case of climate impact studies.*

RC: I'd like to see more details on the station data, which could be potentially one of the most interesting parts of the study.

AC: *A Table with the most relevant information about the weather stations used in this study has been added as Supplementary Material.*

RC: Don't include equations in the core manuscript, move them all to supplementary materials.

AC: *As a matter of fact, there are only three equations in the original manuscript. In the revised manuscript, few equations were added regarding the drought analysis, increasing the total number to eight. For the sake of readability, we prefer to keep them in the main text of the manuscript.*

RC: Drought part is very poor. Why not using, at least, the SPI and the SPEI? Also, the choice of quantities related to drought are not enough to justify the publication, I'd expect a lot more (frequency of events, intensity, severity, return periods, spatial aggregation, etc.) especially on monthly basis (not annual).

AC: *We thank the referee for this valuable suggestion. We agree with the referee that our work can benefit from more analyses on droughts, though we only partially agree with carrying out some of the analyses that he/she suggests. In particular, SPI and SPEI, by definition, follow a standard normal distribution. Hence long-term statistics (mean, standard deviation, etc.) are the same for the model and the observations. This feature hinders the possibility to use the considered error metrics and models' ranking to evaluate the models' performances, as in principle differences between the statistics derived from simulated and observed standardized drought index series could be primarily accounted for as sampling variability, rather than the*

actual RCMs' skill in reproducing wet and dry conditions. For this reason we preferred to apply the theory of runs to precipitation data for drought identification.

To extend the drought analysis, as suggested by the referee, drought events were also identified on seasonal precipitation values simulated for the period 1971-2000. Also, the return period of drought events of fixed duration computed on both annual and seasonal precipitation data was included in the revised manuscript. The methodological aspects of these changes were addressed in the revised Section 3.2 of the manuscript (LL 239-329). Concerning results, the following outcomes are described and discussed:

- Frequency distribution of RCMs percentage errors in maximum drought duration, maximum drought accumulated deficit and maximum drought intensity at the annual and seasonal scale (new Fig. 11);*
- Frequency distribution of RCMs percentage errors in the return period of drought event of duration L equal to 1, 3, 5 and 7 years (new Fig. 12);*
- Frequency distribution of RCMs percentage errors in the return period of drought event of duration L equal to 2, 4, 6 and 8 seasons (new Fig. 13);*
- Improved rankings based on drought characteristics merging drought maximum intensities and return periods at the annual and seasonal levels (new Fig. 14);*
- Improved overall ranking embedding the new drought analysis (new Fig. 16).*

All the new findings were introduced and described in the improved sections concerning drought characteristics results (Section 4.3.1 LL 470-489 and the whole new Section 4.3.2 about results at the seasonal scale) and discussion (particularly Section 5.1, LL 566-576, and Section 5.2, LL 601-611). Tables 5 and 6 were also updated accordingly.

Finally, regarding the possibility to include spatial aggregation of drought events (or drought characteristics) in our study, as the current investigations rest upon at-site analysis, for the sake of clarity (and brevity) we preferred to not introduce results at different spatial levels. However, we aim to consider regional droughts in a future evaluation study.

RC: Some conclusions are exactly what one might expect: precipitation is modelled worse than temperature, drought (as computed in this study) is similar to precipitation, RCMs deeply affect the results more than GCMs.

AC: *In light of the changes made to the manuscript, and in particular, of the enhanced drought analysis, conclusions were rewritten, highlighting the main findings of the study.*

Towards a reliable assessment of climate change impact on droughts in Southern Italy: Evaluation of EURO-CORDEX historical simulations by high-quality observational datasets

David J. Peres¹, Alfonso Senatore², Paola Nanni¹, Antonino Cancelliere¹, Giuseppe Mendicino², Brunella Bonaccorso³

¹Department of Civil Engineering and Architecture, University of Catania, Catania, 95123, Italy

²Department of Environmental Engineering, University of Calabria, Arcavacata di Rende (CS), 87036, Italy

³Department of Engineering, University of Messina, St Agata, Messina, 98166, Italy

Correspondence to: Alfonso Senatore (alfonso.senatore@unical.it)

Abstract. Many recent studies indicate climate change as a phenomenon that significantly alters the water cycle in different regions worldwide, also implying new challenges in water resources management and drought risk assessment. To this end, it is of key importance to ascertain the quality of Regional Climate Models (RCMs), which are commonly used for assessing at proper spatial resolutions future impacts of climate change on hydrological events. In this study, we propose a statistical methodological framework to assess the quality of the EURO-CORDEX RCMs concerning their ability to simulate historic climate (temperature and precipitation) and drought characteristics (duration, accumulated deficit, ~~and~~ intensity and return period) determined by the theory of runs, at seasonal and annual time scales, by comparison with high-density and high-quality ground-based observational datasets. In particular, the proposed methodology is applied to Sicily and Calabria regions (Southern Italy), where long historical precipitation and temperature series were recorded by the ground-based monitoring networks operated by the formerly Regional Hydrographic Offices, whose density is considerably greater than observational gridded datasets available at the European level, such as E-OBS or CRU-TS. Results show that among the more skilful models, able to reproduce, overall, precipitation and temperature variability, as well as drought characteristics, ~~are many are~~ based on the COSMO-CLM-Community RCM, ~~with the significant exception of the particularly in~~ combination ~~based on the HadGEM2-ES GCM and the RACMO~~ with the HadGEM2 RCM GCM. Nevertheless, ~~the the ranking of the models may slightly choice of the most appropriate model change~~ dependings on the specific variable analysed, as well as the temporal and spatial scale of interest. From this point of view, the proposed methodology highlights the skills and weaknesses of the different configurations; ~~supporting and can serve as an aid for a selecting the most suitable proper climate model selection for climate projections assessing climate change impacts depending on the examined hydrologic processes on drought processes and the primitive underlying variables on which it depends.~~

1 Introduction

A growing number of scientific studies claims that climate change due to global warming will significantly alter the water cycle, with an increase of the intensity and frequency of extreme hydro-climatic events in several areas around the globe (Arnell et al., 2001; Huntington, 2006; IPCC, 2014; IPCC, 2018). These include the Mediterranean region, which is recognized as one of the major hot spots of climate change due to future projections of temperature increase and annual precipitation decrease (Giorgi, 2006; Kjellström et al., 2013).

Global Circulation and Regional Climate Models (GCMs and RCMs) can play a crucial role in understanding the potential spatiotemporal evolution of climate change in the future, thus improving current monitoring and planning tools (e.g., Mendicino and Versace, 2007; Hart and Halden, 2019) and supporting decision-makers to choose and implement the best solutions to minimize the impact of climate change on human systems and the environment at the regional scale. While GCMs' simulations describe climate evolution at large scale, by using coarse resolution information, RCMs simulations, derived through climate-downscaling techniques, aim at representing regional and local scale weather conditions with grid resolutions lower than 50 km down to about 10 km (Kotlarski et al., 2014; Peres et al., 2019).

Several studies, focused on the use of climate models to simulate future climate scenarios for hydrological analyses, have shown that changes in temperature and precipitation vary in space depending on the future climate scenario, type, and resolution of the models, as well as on spatial heterogeneity of climatic features. This is particularly evident in the Mediterranean region where, for instance, precipitation is partially controlled by orography, shows strong seasonality and large interannual fluctuations, and is characterized by the occurrence of particularly intense extreme eventss, such as prolonged droughts and high-intensity storms leading to floods (Bonaccorso et al., 2013; Bonaccorso et al., 2015a and 2015b; Llasat et al., 2016; Senatore et al., 2020).

Recently, there is a growing interest in the implementation of RCMs derived by dynamical downscaling of GCM outputs for climate change impact studies at small spatial scales. These are high-resolution models able to provide a more realistic representation of important surface heterogeneities (such as topography, coastlines, and land surface characteristics) and mesoscale atmospheric processes.

The Coordinated Regional Climate Downscaling Experiment (CORDEX) initiative is the first international program providing a common framework to simulate both historical and future climate at the regional level, under different Representative Concentration Pathways (RCPs) (van Vuuren et al., 2011), and over different domains which cover all the land areas. More specifically, it provides climate data simulated by an ensemble of RCMs developed by several research centres all over the world which-that are forced by Global Circulation Models (GCMs) from the Coupled Model Intercomparison Project phase 5 (CMIP5); (Taylor et al., 2012). In the present study, we refer to the CORDEX domain centred on the Euro-Mediterranean area, known as EURO-CORDEX (Jacob et al., 2014; (www.euro-cordex.net)). In particular, EURO-CORDEX provides simulations for a historic reference period (baseline) and future projections up to 2100, with a 12.5 km grid resolution, available for four RCPs defined at the international level within the Coupled Model Intercomparison Project—Phase 5 (CMIP-5).

The reliability of individual RCMs in representing climate effects on the hydrological cycle depends on the quality of historical simulations and must be evaluated before using their output for impact assessment. Assessing RCMs performance is essential to either select single models for further applications (e.g., Senatore et al., 2011; Peres et al., 2017; Smiatek and Kunstmann, 2019) or properly weight individual RCMs in multi-model ensembles to predict future impacts of climate change on hydrological processes (e.g., Christensen et al., 2010; Coppola et al., 2010). ~~Indeed, intercomparison and validation studies to evaluate RCMs' performances and to provide a ranking based on some hydrological measures, have demonstrated that no model can be considered optimal for every variable and region.~~ Table 1 provides a broad, although not thorough, list of intercomparison studies within the CORDEX framework available in the literature. Overall, these studies show that CORDEX RCMs can reproduce the most important climatic features at regional scales, but ~~that~~ important biases remain, especially regarding precipitation or climate extremes. As reported by Kotlarski et al. (2014) and references therein, model biases may depend on the analysed region, choices in model configuration, internal variability, and uncertainties of the observational reference data themselves (Gampe et al., 2019). Concerning the latter, a common approach in evaluation exercises consists in comparing models' simulations to observational gridded datasets, from remote sensing or model-derived reanalyses products available at global or continental spatial scales.

In general, statistical measures, such as bias, root mean square error, correlation, and trend analysis, are used to quantify model performance. Regardless of the specific methods used to assess the differences between simulated and observed data, one of the main limitations in this approach is that the considered spatial resolution is too coarse for reliable climate change impact studies at relevant hydrological scales, especially in areas of complex topography. From this point of view, large-scale observational gridded datasets are of poor applicability, since they are built upon low-density hydro-meteorological networks. In principle, more accurate evaluations can be achieved when they rely on gridded reference data sets that are obtained by spatial interpolation of point measurements onto a regular grid. To this end, two main prerequisites are that data coverage well reflects the topography and variables with limited spatiotemporal climatic variability are investigated (Wagner et al., 2007). For example, Mascaro et al. (2018) compared the skill of several EURO-CORDEX RCMs at ~ 50 and 12 km grid spatial resolution in reproducing annual and seasonal precipitation regimens and trends in Sardinia (Italy), against a dense network of rain gauges with long term records. Their analysis revealed that, although the simulated spatial patterns of annual and seasonal means are well correlated with the observations, positive and negative biases up to $\pm 60\%$ in the simulation of annual mean and interannual variability are detected. Furthermore, the majority of RCMs underestimate winter and overestimate summer precipitation.

In this study, ~~we present we propose a similar an enhanced evaluation analysis exercise over~~ a different Mediterranean area with complex topography, namely Sicily and Calabria regions (Southern Italy). ~~In particular, after ,by~~ investigating the ability of the EURO-CORDEX models to simulate the annual and seasonal temperature and precipitation regime, we analysed the skill in reproducing as well as drought event characteristics determined from identified through the run method proposed by (Yevjevich, (1967)). ~~Run method, here defines a drought as a spell of consecutive intervals where the annual precipitation values are continuously below the long term mean, according to the theory of runs proposed by Yevjevich (1967). Within the~~

analysis of drought analysis, we also consider also investigated the return period of drought events of fixed duration at both the annual and seasonal scales. In this case, given the limited number of droughts in a thirty-year long time series, we apply inferential methodologies an analytical framework was applied that allow to compute return period starting from based on probabilistic reasonable assumptions on the probabilistic structure of annual and seasonal precipitation (Bonaccorso et al., 2003; Cancelliere and Salas, 2004). Furthermore, we analysed model skills at a sub-regional level. To this aim, we proposed the use of Principal Component Analysis (PCA) for delimitation of climatically homogeneous areas. The ability of climate models to reproduce observed precipitation, temperature and drought features was analysed both per single characteristic as well as per multiple characteristics (e.g. precipitation and temperature together), by introducing a specific ranking criterion that allows combining any of the performances respect to the single characteristics. Indeed, understanding how well the models can reproduce past droughts is crucial for future effective water resources management in the Mediterranean region. In particular, the performance of Nineteen 19 coupled GCM and RCM simulations within the EURO-CORDEX framework were evaluated against a high-density and high-quality monitoring station-based reference dataset. Monthly temperature and precipitation records were retrieved by two monitoring networks, operated by the former Regional Hydrographic Services, whose density is significantly higher than observational datasets available at the European scale, such as E-OBS (Haylock et al., 2008) or CRU-TS (Harris et al., 2014), allowing for a more accurate evaluation of the models. Beyond the intercomparison analysis of the EURO-CORDEX RCMs, the present study also aims at identifying potential sub-regions where model improvements are particularly advisable. The study is organized as follows: after introducing the study area, the station-based reference dataset, and the GCM and RCM datasets in Section 2, Section 3 outlines the methodology applied for identifying climatically homogeneous zones in the study area through the Principal Component Analysis (PCA), and for evaluating models' performance in both the whole study area and the homogeneous zones; furthermore, the adopted statistical performance metrics and the ranking criteria are introduced. Then, Section 4 presents the evaluation results for each investigated variable over the whole study area and the different zones. The results are further discussed in Section 5, highlighting the basic model capabilities identified, as well as the biases in modelling climate and drought conditions in Southern Italy. Conclusive remarks are drawn in Section 6, together with an outlook on future evaluation and prediction activities in the EURO-CORDEX framework.

2 Study area and datasets

Our analyses were focused on Calabria and Sicily regions in Southern Italy, which respectively have an extension of 15,080 km² and 25,460 km², for a total area of 40,540 km² (Fig. 1). Climate is of Mediterranean type with hot and dry summers and moderately cold winters with peak monthly precipitation occurring mostly in late autumn and winter. About 75% of the total precipitation in the study area occurs from October to March, because of cyclonic storms. These climate features make the area particularly prone to droughts, with the most recent severe episode has occurred in 2017 on (Rossi and Benedini, 2020; Bonaccorso; Senatore). Climate features are also highly variable in space due to a rather complex orography. In particular,

130 the mountain chains close to the coast enhance intense orographic precipitation and lead to relatively cold temperatures at the
highest altitudes.

2.1 Observed data

Within the EURO-CORDEX control period (1951-2005), the comparison with observations was performed ~~ion~~ on the period
from 1971 to 2000. These three decades had the greatest availability of historical series of precipitation and temperature
135 recorded by both the regional monitoring networks of Calabria and Sicily, managed by the Multirisk Operational Centre of
Calabria region (ArpaCal) and the Water Observatory of Sicily region (WOS), respectively. Specifically, 84 thermometers (43
in Sicily and 41 in Calabria [Calabria and near the regional borders](#)) and 335 rain gauges (173 in Sicily and 162 in Calabria and
near the regional borders) were used (Fig. 1). [Details on the monitoring network are given in the Supplementary Material to
this paper.](#)

140 The corresponding data were retrieved by the WOS (www.osservatorioacque.it) and the ArpaCal (www.cfdcalabria.it)
websites. Observations were enough widespread to represent the quite heterogeneous features of the study area. The
temperature stations were located between 2 and 1295 m a.s.l., with annual average values ranging from 9.2 °C to 20.6 °C
(mean value = 16.2 ± 2.4 °C), while the rain gauge elevations varied from 1 to 1369 m a.s.l., with annual accumulated values
ranging from 373 mm to 1736 mm (mean value = 812 ± 287 mm).

145 2.2 Climate models

Monthly precipitation and monthly mean air temperature data from the EURO-CORDEX CMIP5 simulations (Jacob et al.
2014; <https://www.euro-cordex.net/>) were retrieved from the nodes of the Earth System Grid Federation (ESGF, e.g.
<https://esgf.llnl.gov>).

We analysed the data at the finest resolution, 0.11° (~ 12.5 km), EUR-11 and considered the period 1971-2000 as a baseline.
150 In particular, the combination of six GCMs (Tab. 2) and eight RCMs (Tab. 3) leading to 17 datasets, reported in Tab. 4, were
collected for the study. Moreover, for two GCM-RCM combinations, two versions were available from the ESGF portal.
Therefore, an overall ensemble of 19 combined models (CMs) was analysed. The ensemble mean of the 19 CMs was also
evaluated. Even if the CMs have the same spatial resolution, each one is distributed on a specific grid (with slightly different
origin and orientation of the axis). Therefore, the various data sets were resampled on the grid of the ECE-HIRH CM, which
155 is shown in Fig. 1.

We choose EUR-11 rather than EUR-44 simulations as several studies (Torma et al., 2015; Prein et al., 2016) have found that
generally higher resolution CORDEX RCMs have better skills in simulating seasonal precipitation in regions with complex
terrain.

3 Methodology

160 3.1 Data processing and PCA

To allow the comparison between the spatially distributed RCMs data and site-specific observations, the latter were spatially interpolated using the CORDEX 0.11° grid as reference (Fig. 1). In this way, month by month, each cell of the CORDEX grid could be associated with a single temperature or precipitation value derived from the observations network. Specifically, concerning temperature, an Inverse Distance Weighting (IDW) interpolation was applied to the residuals of the values obtained using a regression model with the altitude. For precipitation, whose measurement network is much denser, a simple IDW interpolation was performed. As shown in Fig. 1, the CORDEX grid cells which are not covered by any rain gauge are relatively few (less than 30%) and, except one case, the distance of the closest rain gauge to every grid cell is always less than 10 km.

The precipitation patterns obtained by the interpolation procedure were analyzed adopting a methodology based on the Principal Component Analysis (PCA) to distinguish zones with rather independent climatic variability within the area under investigation. PCA is a well-known statistical tool used to transform an original set of intercorrelated variables into a reduced number of new linearly uncorrelated ones explaining most of the total variance (Rencher, 1998). The latter, derived as linear combinations of the original variables, are the principal components (PCs), while the coefficients of the linear combinations are the loadings, which in turn represent the weight of the original variables in the PCs. From a procedural standpoint, PCA consists of solving an eigenvalue-eigenvector problem applied to the covariance matrix. The eigenvectors, properly normalized, are the loadings of the principal components, while the eigenvalues provide a measure of the total variance explained by each loading (Bordi and Sutera, 2001 and references therein). Under this decomposition, the loadings represent the correlation between the associated PCs and observed time series. Mapping the loading patterns of each PC among those selected, based on the percentage of the total explained variance of the process, largely allow to identify independent climatic areas within the study region. Moreover, it may be useful to apply a rotation operation to the eigenvectors, so that the corresponding loadings are more spatially localized. In other words, the rotation leads to loadings with a high correlation with a smaller set of spatial variables and a low correlation with the remaining variables. Here, only orthogonal rotations were considered, computed by the varimax algorithm in Matlab® R2016. Clearly, each rotated pattern will not explain the same variance of the unrotated one, although the total variance explained remains unchanged.

In the present study, the first nine rotated PCs both at the annual and seasonal (DJF, MAM, JJA, SON) scales were investigated. Regardless of the considered period, the selected PCs always explain more than 78% of the total variance, with a maximum of 85% in the winter season (DJF). The loading patterns of these rotated PCs were mapped for each considered period to identify climatically homogeneous regions. Homogeneous sub-regions were detected at the annual scale and in autumn and winter seasons, whereas a confusing picture arose in spring and summer seasons. Furthermore, since about 75% of the total annual rainfall of the case-study area occurs between autumn and winter (as a result of cyclonic storms), the climatically homogeneous sub-regions identified at the annual scale approximately overlap with those identified at seasons SON and DJF. Isolated grid cells showing a different PC correspondence with respect to the surrounding cells, were manually corrected to simplify the

delimitation of the homogeneous sub-regions. This approach led in dividing the whole area into six climatically homogenous zones, three for Sicily and three for Calabria (Fig. 1), for which separate performance assessments were carried out. Concerning Sicily region, the three identified sub-regions roughly coincide with the ones detected by Bonaccorso et al. (2003), who investigated the spatial variability of droughts in Sicily region based on SPI series computed on monthly precipitation observed in traditional rain gauges and on NCEP/NCAR reanalysis data from 1926 to 1996. In particular, three distinct areas, namely North-Eastern (identified in the PCA as zone 5, Fig. 1b), South-Central Eastern (zone 4), and Central-Western (zone 1), were identified. In Calabria, three main zones were also determined, namely North-Western (zone 2), North-Eastern (zone 3) and South-Eastern (zone 6), broadly corresponding to climatic homogenous areas found in previous studies (e.g., Versace et al., 1989). Interestingly, the South-Western tip of Calabria is identified as a part of a broader area (zone 5) extending over the North-Eastern Sicily. The precipitation patterns obtained by the interpolation procedure were analyzed adopting a methodology based on the Principal Component Analysis (PCA) to distinguish zones with rather independent climatic variability within the area under investigation. PCA is a well known statistical tool used to transform an original set of intercorrelated variables into a reduced number of new linearly uncorrelated ones explaining most of the total variance (Rencher, 1998). The latter, derived as linear combinations of the original variables, are the principal components (PCs), while the coefficients of the linear combinations are the loadings, which in turn represent the weight of the original variables in the PCs. From a procedural standpoint, PCA consists in solving an eigenvalue-eigenvector problem applied to the covariance matrix. The eigenvectors, properly normalized, are the loadings of the principal components, while the eigenvalues provide a measure of the total variance explained by each loading (Bordi and Sutera, 2001 and references therein). Under this decomposition, the loadings represent the correlation between the associated PCs and observed time series. Moreover, it may be useful to apply a rotation operation to the eigenvectors, so that the corresponding loadings are more spatially localized. In other words, the rotation leads to loadings with a high correlation with a smaller set of spatial variables and a low correlation with the remaining variables. Here, only orthogonal rotations are considered, computed by the varimax algorithm in Matlab® R2016. Clearly, each rotated pattern will not explain the same variance of the unrotated one, although the total variance explained remains unchanged. In the present study, PCA led in dividing the whole area into six climatically homogenous zones, three for Sicily and three for Calabria (Fig. 1), for which separate performance assessments were carried out. Concerning Sicily region, the three identified sub-regions roughly coincide with the ones detected by Bonaccorso et al. (2003), who investigated the spatial variability of droughts in Sicily region based on SPI series computed on monthly precipitation observed in traditional rain gauges and on NCEP/NCAR reanalysis data from 1926 to 1996. In particular, three distinct areas, namely North Eastern (identified in the PCA as zone 5, Fig. 1b), South Central Eastern (zone 4), and Central Western (zone 1), were identified. Also in Calabria, three main zones were determined, namely North Western (zone 2), North Eastern (zone 3) and South Eastern (zone 6), broadly corresponding to climatic homogenous areas found in previous studies (e.g., Versace et al., 1989). Interestingly, the South-Western tip of Calabria is identified as a part of a broader area (zone 5) extending over the North Eastern Sicily.

3.2 Performance metrics and models' ranking

225 The CMs were evaluated based on their performances in capturing specific properties, namely: the interannual and seasonal variability of precipitation, temperature and drought characteristics. Such properties were expressed based on some relevant statistics.

Let $X(j)$ and $X_\tau(j)$ be the variable under investigation (precipitation or mean temperature) at grid cell j at the annual and seasonal scale, respectively. For precipitation and mean air temperature, the following statistics were derived for each CM and
230 cell in the area of interest:

- Seasonal mean $\mu_m(X_\tau(j)) = \frac{\sum_{i=1}^N x_{\tau,i,m}(j)}{N}$

where $x_{\tau,i,m}(j)$ is the value of the variable at season τ ($\tau = 1, 2, 3, 4$) and year i ($i = 1, 2, \dots, N$) produced by the m -th CM ($m = 1, 2, \dots, M$) at cell grid j . Seasons are December – February (DJF), March – May (MAM), June – August (JJA), and September – November (SON);

- 235 Seasonal standard deviation $\sigma_m(X_\tau(j)) = \sqrt{\frac{\sum_{i=1}^N (x_{\tau,i,m}(j) - \mu_m(X_\tau(j)))^2}{N-1}}$;

- Annual mean $\mu_m(X(j)) = \frac{\sum_{i=1}^N x_{i,m}(j)}{N}$;

where $x_{i,m}$ is the value of the variable at year i ($i=1, 2, \dots, N$) produced by m -th CM;

- Annual standard deviation $\sigma_m(X(j)) = \sqrt{\frac{\sum_{i=1}^N (x_{i,m}(j) - \mu_m(X(j)))^2}{N-1}}$.

240 Drought events were identified on both annual and seasonal (DJF, MAM, JJA, SON) precipitation values simulated for the period 1971-2000, according to the theory of runs (Yevjevich, 1967). In particular, drought events were selected as the periods during which consecutive annual or seasonal values of precipitation did not exceed a given threshold, here assumed equal to the long term means of annual and seasonal data (i.e. onea-different threshold -for each season). Once drought events were identified, the corresponding drought characteristics in each cell were determined. In particular, the following statistics for drought characteristics are considered hereafter to assess the models' performance:
~~Drought events were identified on annual precipitation values simulated for the period 1971-2000, according to the theory of runs (Yevjevich, 1967). In particular, drought events were selected as the periods during which consecutive annual values of precipitation did not exceed a given threshold, here assumed equal to the long term mean. For further details about the theory of runs, the readers may refer to Bonaccorso et al. (2003, 2013) and reference therein. Once drought events were identified, the corresponding drought characteristics in each cell were determined. In particular, the following statistics for drought characteristics are considered~~
245 ~~hereafter to assess the models' performance:~~
250 ~~hereafter to assess the models' performance:~~

- Maximum drought duration L_{max} : maximum length of periods with consecutive annual precipitation values below the threshold;

- Maximum drought accumulated deficit D_{max} : maximum of the sums of the differences between the threshold and the precipitation values along with the drought duration:-
- Maximum drought intensity I_{max} : maximum of the ratio between drought accumulated deficit and duration:-
- Return period of drought events of fixed duration (at both annual and seasonal scales):-

Concerning the return period of drought events, let E be a critical drought (e.g., a drought with duration L equal to a fixed value). Assuming independence between consecutive drought events, the return period of drought event E can be expressed as (Gonzales and Valdes, 2003; Cancelliere and Salas, 2004; Cancelliere and Salas, 2010; Bonaccorso et al., 2012):

$$T_E = \frac{E[L] + E[L_n]}{P[E]} \quad (1)$$

where $E[L]$ is the expected value of drought duration L and $E[L_n]$ is the expected value of the non-drought duration L_n and $P[E]$ is the probability of occurrence of a critical drought E , which can be determined once that the probability distribution function of the event E is known.

Regarding the probability distribution of drought duration, let us consider a ~~periodic~~ stochastic hydrological variable denoted as $X_{v,\tau}$ where v represents the year and τ represents the season. According to the theory of runs, drought duration L is the number of consecutive time intervals (seasons) where $X_{v,\tau} \leq x_{o,\tau}$ is preceded and followed by at least one season where $X_{v,\tau} > x_{o,\tau}$, where $x_{o,\tau}$ is a threshold level representing water demand. The original variable can be replaced by a Bernoulli variable $Y_{v,\tau}$ such that:

$$\begin{cases} Y_{v,\tau} = 0 \text{ if } X_{v,t} \leq x_{0,\tau} \text{ (deficit)} \\ Y_{v,\tau} = 1 \text{ if } X_{v,t} > x_{0,\tau} \text{ (surplus)} \end{cases}$$

(2)

Assuming that $Y_{v,\tau}$ is a lag-1 Markov stationary process, it can be shown (Sen, 1976; Cancelliere and Salas 2004; Cancelliere and Salas, 2010) that the probability distribution of drought duration L is geometric with parameter p_{01} :

$$f_L(\ell) = P[L = \ell] = P[Lp_{01}]^{\ell-1} p_{01}$$

(3)

The parameter p_{01} represents the transition probability from a deficit to a surplus, namely $p_{01} = [Y_{v,\tau} = 1 | Y_{v,\tau-1} = 0]$. Estimation of transition probabilities can be carried out following a non-parametric approach based on maximum likelihood, which leads to (Bonaccorso et al., 2012):

$$p_{01} = 1 - p_{00} = 1 - \frac{n_{00}}{n_{00} + n_{01}}$$

315

320

325

330

335

340

(4)

where n_{00} is the number of observations $y_{v,\tau} = 0$, for which $y_{v,\tau-1} = 0$, and n_{01} is the number of observations $y_{v,\tau} = 1$, for which $y_{v,\tau-1} = 0$.

For independent stationary series, the probability distribution of drought duration L is geometric with parameter $p_1 = P[Y_\tau = 1]$. The latter can be simply estimated by applying a frequency analysis on Y_τ .

Following previous studies (Bonaccorso et al., 2003; Cancelliere and Salas, 2004), the annual series were assumed independent stationary, whereas the seasonal series as lag-1 stationary Markov.

Models' skills in reproducing the interannual and seasonal variability of precipitation and mean air temperature variables were first assessed through:

- boxplots of the errors and percentage errors of the mean values in all the grid cells of the investigated areas, which allow analysing the spatial variability of the models' bias;
- Taylor diagrams (Taylor, 2001), which show three metrics at the same time, i.e.: coefficient of correlation, standard deviation, and centred root mean square error of the anomalies (i.e., the variables of interest minus the corresponding means). It is noteworthy that standard Taylor diagrams do not provide any information about first-order statistics (i.e., bias).

Later, to provide synthetic information about each CM starting from the various statistics computed for each property, a method based on Mascaro et al. (2018) was used. Specifically, for each property (i.e. seasonal and interannual variability of precipitation and mean temperature and drought characteristics), a single dimensionless error metric that combines multiple statistics characterizing that property was estimated. The error metrics follows the equation:

$$\varepsilon_m = \sqrt{\sum_{k=1}^S \left(\frac{\sum E_{k,m}(j)}{\sum_{m=1}^M \sum_{j=1}^P E_{k,m}(j)} \right)^2} \quad (45)$$

where $E_{k,m}(j)$ represents an error metric between observed and simulated data of the statistics k ($k = 1, \dots, S$) at grid cell j ($j=1, \dots, P$, where P is the total number of grid cells), whose sum over the whole area was divided by the sum of the error

345 metrics of all models, therefore resulting in a dimensionless indicator for each statistic k of any property. Table 5 summarizes the statistics chosen for each property and describes how the corresponding errors were calculated.

Based on the values of the error metrics in Eq. (1), a ranking of the models, describing the skills in reproducing each property, was obtained. It should be specified that while, for the sake of brevity, the boxplots and the Taylor diagrams illustrated in the next section refer to the whole study area, the ranking of the models for the mean air temperature, precipitation and drought characteristics also refers to the six climatically homogenous zones identified through PCA. This analysis, indeed, can help to highlight whether some models are more suitable than others to simulate certain variables in a given zone.

4 Results

In this section, results are presented and discussed separately for temperature, precipitation and drought characteristics. Results are differentiated for the following temporal and spatial aggregation scales: annual data, seasonal data, the whole case study region and the six climatically homogenous areas identified via PCA.

4.1 Mean air temperature

4.1.1 Interannual variability

The observed and modelled means of the annual mean air temperature values in each of the grid cells within the study area were calculated and compared. More specifically, for each cell j , the error corresponding to the m -th CM was computed as:

$$360 \quad E_{m,j} = \mu_m(T(j)) - \mu_0(T(j)) \quad (26)$$

where $T(j)$ is the mean annual temperature at cell j , whereas $\mu_m(\cdot)$ and $\mu_0(\cdot)$ are the modelled and observed means respectively. For each model, the distribution of the errors computed for all the grid cells of the study area based on Eq. (2), is represented in the form of box-plots in Fig. 2a. In particular, the central line represents the median value and the box is delimited by the first and the third quartile. The width of the box corresponds to the inter-quartile range (IQR), a well-known measure of dispersion. Values outside the whiskers, distant from the box at least 1.5 IQR, can be assumed as outliers.

The overall tendency of the models is to underestimate temperatures, as the medians are negative. Errors are predominantly comprised between the values -5 and -1 °C, thus implying that the models underestimate up to 5 °C. The CMs that produce the most extreme negative errors are the ECE-RACM, ECE-RACMr12 and CM5-ALAD, with the latter showing the broader IQR (e.g. the highest spatial variability of the errors) and the greatest median error. All the CMs with RCA4 show the smallest IQR. The models with the smallest median error are MPI-REMO and MPI-REMOv2.

To extend the CM skill comparison to other statistics, the Taylor diagram for the annual mean air temperature values was developed (Fig. 32b). For the sake of simplicity, standard deviations of the CMs are indicated as σ hereinafter. The diagram allows visualizing if there are clusters of performances related to specific GCMs or RCMs among those considered. In the diagram, GCMs are indicated with different markers, while RCMs with different colors. The value corresponding to the

375 observations is the dot on the x -axis, whose standard deviation is marked through a continuous circular arc. In addition to every single model, the ensemble mean model result is reported in the diagram.

From Fig. 32b, it can be seen that the simulated means are well correlated with the observations, with values larger than 0.8 for all the considered models. Furthermore, the diagram seems to reveal that, on equal GCMs, RCMs play a significant role in determining the performance of the combinations. In general, for most of the models, the best performances are obtained when the RCM RCA4 is used. The only exception is CM5, performing better in combination with CCLM. The worst models are CM5-ALAD and IPS-WRF.

Finally, the ranking analysis described in Section 3.2 yields the results in Fig. 34. The lower the rank, the lower is the error metrics in Eq. (1) and the better is the model. For better readability, ranking values are indicated through a chromatic scale, ranging from dark green (first ranked model) to dark red (last ranked model).

385 The best performing models, in terms of ranking order for the whole study area, are MPI-CCLM, MPI-REMO, and Had-CCLM. ECE-RCA4 and CM5-CCLM are also good models as highlighted by the Taylor diagrams. Figure 4 also shows rankings for each of the six homogeneous areas. As it can be observed, based on the range of colours in each row, MPI-CCLM and MPI-REMO provide the best performance for almost all the zones.

Indeed, some differences exist for Zones 3 and 6 (North and South-Eastern Calabria), whose best CM is IPS-RCA4. Overall, results show that the worst model is CM5-ALAD for entity and dispersion of errors, lower correlations, higher RMSE, greater deviation from the standard deviation of the observed values, both for the whole study area and individual zones. ECE-RACM, ECE-RACMr12, and ECE-RCA4 also show bad performance (the latter mainly because of its relatively strong bias).

4.1.2 Seasonal variability

For the sake of brevity, the box-plots related to the seasonal variability of mean air temperature are not shown since they provide similar results to the case of annual variability.

Figure 45 shows the Taylor diagrams obtained from the analysis of the individual seasons. CM5-ALAD and IPS-WRF (and, to a slightly lesser extent, CM5-ALAR) appear as the worst models regardless of the season, although in summer (JJA) the worst-performing models are MPI-REMO and MPI-REMr2. Summer is also the season with the (slightly) lowest values of correlation coefficients.

400 Regarding the best models, in general, all the combinations with RCA4 and the CM5-CCLM work better, as for the interannual variability analysis. However, in summer better performances are obtained with ECE-RACM and ECE-RACMr12.

Figure 56 represents the rankings of the models for the individual seasons and all the study areas, namely the whole case study and the six zones. There is a certain correspondence on the least performing models between Figs. 54 and 56. Nonetheless, differently from the results in Fig. 34, models' performances may change significantly from season to season and, in the same season, from zone to zone. The best models for most of the zones are ECE-HIRH in winter (DJF), ECE-CCLM in spring (MAM), IPS-RCA4 in summer (JJA) and MPI-REMr2 in autumn (SON). It's worth highlighting that the latter provides the

best performances also for Zones 2 and 4 in spring and Zones 5 and 6 in summer. Conversely, ECE-HIRH, which is the best model in winter, works poorly in summer and autumn. The Zones 1 (Western Sicily) and 2 (Western Calabria) show a uniform behaviour in all seasons, with the only exception of spring, while Zones 5 (North-Eastern Sicily) and 6 (South-Eastern Calabria) show a uniform behaviour in all seasons but autumn. Besides, in summer and autumn, the best performing models for Zones 1, 2 and 4 (South-Eastern Sicily) are the same as for the whole study area. Zone 3 (North-Eastern Calabria) behaves like Zone 4 in winter and like Zones 1, 5 and 6 in spring.

4.2 Precipitation

4.2.1 Interannual variability

Figure 6a7 shows box-plots for the percentage errors in mean annual precipitation, namely:

$$E_{m,j} = \frac{\mu_m(P(j)) - \mu_0(P(j))}{\mu_0(P(j))} \cdot 100 \quad (73)$$

where $P(j)$ is the total annual precipitation at the grid cell j .

In comparison to temperature, the errors are much larger, as well as the differences between the various models. There is a general tendency for the models to underestimate the total annual precipitation, except for some models like IPS-WRF, which also shows the largest IQR. The median value of the relative errors for some models is less than 20%; however, many models have a large dispersion with error values over 100%. The CM with the highest positive error is IPS-WRF, while the ones with the highest negative errors are the IPS-RCA4 and Nor-HIRH models. The GCM-RCM combinations with the smallest IQR of errors are those using CCLM RCMs. The model with the smallest bias is Had-RACM.

The Taylor diagram in Fig. 6b8 confirms that the best combinations are those with CCLM RCMs. In particular, the best one seems ECE-CCLM. However, when used in combination with CM5, the corresponding model provides poor performance. The worst performing models are ECE-HIRH and Nor-HIRH. The diagram confirms that precipitation is modelled with less accuracy than temperature, as correlations are lower (<0.8).

The application of the ranking criteria (see Fig. 97) suggests Had-RACM and ECE-CCLM as the best combinations for the entire area and most of the zones. Also, CM5-ALAD works well for the whole area and almost all the zones, except for Zone 4, where it ranks the 11th. IPS-WRF, IPS-RCA4, Nor-HIR, and CM5-RCA4 are the worst models.

4.2.2 Seasonal variability

The seasonal variability analysis carried out on precipitation shows (Fig. 810) a lower error dispersion in the wet seasons (i.e., autumn and winter) with respect to summer. In summer, several models show broader IQR, such as all the CM5 models and IPS-WRF, with the latter showing the largest median error. On the one hand, these outcomes depend on the poor performance of some models in reproducing the seasonal cycle, and on the other hand, are due to the fact that in the dry season where rainfall is normally low, large errors may result even though the departure from the observed mean is relatively small. These

results are consistent with those obtained by Giorgi and Lionello (2008) in a subdomain of the Mediterranean region and by Mascaro et al. (2018) for the Sardinia region.

The Taylor diagrams in Fig. 944 highlight that NOR-HIRH and ECE-HIRH are the worst models for all the seasons but summer, where the IPS-WRF is the worst-performing.

These indications are confirmed by the ranking results in Fig. 102. Concerning the best models, the following CMs perform the best in their respective seasons: ECE-RACMr12 in winter (DJF), ECE-CCLM in spring (MAM), MPI-REMO_r2 in summer (JJA), MPI-CCLM and Had-RACM in autumn (SON). It is worth highlighting that ECE-RACMr12 provides the best rank also for Zone 2 in autumn; ECE-CCLM is the best performing also for Zone 6 in summer; MPI-CCLM provides the best performances also for Zone 1 in winter and Zone 4 in spring and Had-RACM is the best model for Zone 2 in spring. For summer precipitation, MPI-REMO_r2 is the best performing CM also for Zones 1, 2, 3 and 4. As for the ranking of seasonal mean temperature, once again there is no uniform behaviour of the models between the different seasons and zones.

4.3 Drought characteristics

4.3.1 Annual scale

The models' performance in reproducing historical drought characteristics both at the annual and the seasonal scale was also tested. In particular, the following drought characteristics derived from the theory of runs were analysed: maximum duration (L_{\max}), maximum accumulated deficit (D_{\max}), and maximum intensity (I_{\max}) and return period of drought duration.

With reference to the drought characteristics identified on annual precipitation, Figures 11a, b and c ~~s from 13 to 15~~ represent the boxplots of the errors related to maximum drought duration, accumulated deficit, and intensity, respectively. In particular, for drought duration, the errors were computed through Eq. (2) by simply replacing T with L_{\max} , whereas for maximum drought accumulated deficit and intensity, the percentage errors were calculated through Eq. (3), by replacing P first with D_{\max} and then with I_{\max} .

There is a slight tendency of some models to underestimate drought duration (Fig. 11a3). Overall, the errors span from -3 and +2 years. The broadest IQR is associated with MPI-REMO, while some models, such as CM5-CCLM, CM5-ALAR, ECE-RACM and, Nor-HIRH seem equally reliable.

The boxplots obtained for D_{\max} (Fig. 11b4), shows that the models may yield considerable errors, which can potentially be larger than those for annual precipitation, as the accumulated deficit, given by the sum of precipitation deficits on a time interval lasting several years, can be affected by multiple errors. For some models, the IQRs are not larger than 50%. The most reliable model is Had-CCLM, but comparable performances are given by models CM5-CCLM, CM5-ALAR and ECE-CCLM, while the least dispersed is MPI-CCLM (for this model, however, the median error is larger than others). The least reliable is IPS-WRF, followed by CM5-RCA4 and MPI-REMO_r2. In general, as it can be seen from the box-plots, this feature is underestimated. Concerning I_{\max} , the results indicate Had-RACM as the best model and CM5-RCA4 as the worst, followed by

IPS-WRF (Fig. 11c5). Errors for this feature are less scattered than for accumulated deficit, and there is a general tendency for I_{\max} to be underestimated by models.

Figure 12 shows box-plots of the errors in the return period of drought events of duration L equal to 1, 3, 5 and 7 years, respectively. In particular, the error was calculated as:

$$E_{m,j} = \mu_m(T_y(j)) - \mu_0(T_y(j)) \quad (8)$$

where $T_y(j)$ is the return period of a drought event of fixed duration at the grid cell j .

As expected, on equal model, the error increases as the considered drought duration increases. However, regardless of the drought duration, there is no general tendency of the models towards overestimation or underestimation of the return periods. ECE-CCLM and Had-RACM are the models with the smallest IQR, with ECE-CCLM showing the lowest median error. Overall, the performance of the models looks rather similar, with limited errors until $L=3$ years (± 0.5 years).

Finally, in agreement with the other variables analysed, the models were also ranked according to their ability in reproducing both observed drought maximum intensities and return periods of drought events with duration $L=3$ years (Fig. 14a6). Since

The purpose of this ranking is to provide a general comparison among the models about drought features, only drought intensity was selected is done concerning this feature only, as it merges drought accumulated deficit and duration of each drought event. With reference to Concerning the return period, it is worth pointing out that the choice of the considered drought duration only affects the magnitude of the errors, while the performance of each model with respect to the others does not change (see Fig. 12). As shown in Fig. 14a6, the best models for the whole study area are confirmed to be ECE-CCLM, Had-RACM, ECE-RACM, CM5-ALAR, and CM5Had-CCLM. These models Interestingly, have the highest ranking also for almost all the zones, with the only exception of CM5-ALAR is the best model for Zone 3, which does not but seem unsuitable for Zone 6 and ECE-RACM for Zone 3 the remaining zones. Overall, the worst models are CM5-RCA4 for all the zones is CM5-RCA4, whereas poor performances are associated to ECE-RACMr12 for Zones 1 and 2, Had-RCA4 for Zone 3, MPI-REMO2 for Zones 4 and 6 and IPS-WRF and, Nor-HIRH for all the zones for Zone 5.

Generally speaking, the skills of CMs in reproducing drought characteristics and interannual variability of precipitation are significantly linked. Drought characteristics, derived through the application of theory of runs, are functions of the departure from the thresholds rather than of the modelled precipitation itself. In other words, although a CM could significantly underestimate or overestimate annual precipitation values (i.e. the data in the boxplots in Fig. 7 may look loosely grouped and the medians very far from 0), still it could provide good performance in terms of drought characteristics simulation if it can reproduce time variability. It is interesting to observe that the distribution of the percentage error of drought intensity (Fig. 15) is, in general, less scattered than that related to the accumulated deficit (Fig. 14); therefore, one can conclude that a partial error compensation occurs when the modelled accumulated deficit is divided by the modelled duration. Despite the differences in the percentage errors, there is however a general agreement in the identification of the best and, mainly, the worst models, also confirmed by the ranking of the models in reproducing drought intensity (Fig. 16).

4.3.2 Seasonal scale

Figures 11d, 11e and 11f represent the boxplots of the errors related to maximum drought duration, accumulated deficit, and intensity identified on seasonal precipitation data.

Concerning drought duration (Fig. 11d), several models (9 out of 19) show a median error equal to 0, while the other models tend to underestimate, with the only exception of IPS-WRF. Overall, the errors span from -4 and +3 seasons.

The broadest IQR is associated with CMC5-ALAR and ECE-CCLM, while some models, such as IPS-RCA4, MPI-RCA4, MPI-REMO-r2 and, Nor-HIRH seem equally reliable.

As for D_{\max} (Fig. 11e), some similarities can be observed with respect to concerning the annual time scale (Fig. 11b) in terms of magnitude of percentage errors, although in the seasonal case most of the models tend to overestimate. The most reliable models are CM5-ALAD, ECE-CCLM and Had-RACM. As for the annual scale, the least reliable is IPS-WRF, followed by CM5-RCA4 and Nor-HIRH.

Concerning I_{\max} , also in the seasonal case Had-RACM is confirmed as the best model, while MPI-REMO-r2 and IPS-WRF are the worst (Fig. 11e). Once again, errors for this feature are less scattered than for accumulated deficit. Only four models underestimate I_{\max} while most of the models ~~has or~~ are close to a zero median percentage error.

Figure 13 shows box-plots of the errors in the return period of drought event of duration L equal to 2, 4, 6 and 8 seasons, respectively. In particular, the error was calculated as in Eq. (8) by replacing T_y with T_s , namely the return period of a drought event of fixed duration identified on seasonal data. As for the annual case, ~~even for the seasonal case~~ the performance of the models looks rather similar, with limited errors (± 5 seasons) until $L=4$ seasons, with the exception of CM5-ALAD, CM5-ALAR, CM5-RCA4 and Had-RCA4.

Figure 14b illustrates the ranking of the models in reproducing the drought maximum intensities and return periods of drought events with duration $L=4$ seasons. With respect to the annual scale, there is a certain agreement in identifying the best performance models, that in this case are Had-RACM, Had-CCLM, and ECE-CCLM. In particular, Had-RACM ~~performworks~~ ~~finewell~~ in every zone, while Had-CCLM is the best model for Zones 1, 2, 5 and 6. ~~With reference to t~~ The least performing models, ~~these~~ are CM5-ALAD, CM5-ALAR, albeit it ranks second for Zone 5, CM5-RCA4 and Nor HIRH.

5 Discussion

Table 6 illustrates the best performing models according to the ranking approach for each of the considered variables over the whole area and the six homogeneous zones, respectively. In particular, the three best performing models are reported for the mean temperature and precipitation interannual variability and drought intensity and return period of drought duration, while only the best CM for each season is indicated for seasonal variability.

It is worth underlining that the rankings are aimed to provide straightforward information about the relative accuracies of the models, e.g., for supporting the selection of a single or few models in a specific area, therefore, for the sake of simplicity, they

provide reduced information based on cardinal numbering. However, the actual performance of each CM compared to the others can be highlighted by looking closer at the ϵ_m values, which reflect and summarize the results provided by the box-plots and the Taylor diagrams.

Two kinds of comparison are carried out in this section: 1) on the same variable, across different time scales; 2) on the same time scale, across different variables. Further discussion is provided about relative impacts of different GCMs and RCMs and, finally, an overall ranking is attempted aimed at providing a global evaluation of the CMs performance.

5.1 Analyses across different time scales (interannual and seasonal)

Concerning temperature, the intercomparison between the interannual and seasonal variability is rather straightforward. All the simulations are characterized by a more or less pronounced underestimation (Fig. 2a), together with a usually high correlation with observations (Fig. 2b3 and 45), i.e. both the observed interannual and seasonal variability are well reproduced. This is somehow confirmed by the rankings, where the relative differences among the models' performances are not very marked.

Conversely, in the case of precipitation, the performances of the models change significantly with the time scale. The most interesting case with this variable is CM5-ALAD that, considering the total area, ranked 3rd with the annual precipitation, but provided low performances in most of the seasons (9th in MAM, 11th in DJF and 18th in JJA). Though CM5-ALAD can reproduce relatively well the annual amount of rainfall, it is not as much able to simulate the seasonal variability, therefore the good performance at the annual time scale is due to the counterbalancing effects of the errors in different seasons. This feature of CM5-ALAD is amplified in several of the six zones, e.g., zone 2 (where it is ranked 4th with the mean annual value, but 14th in DJF and 18th in MAM and JJA) or zone 6 (1st with the mean annual value, but 13th on DJF and 18th on JJA). On the other hand, MPI-CCLM in the total area ranked 8th considering the annual precipitation but provided rather good results in single seasons (it is ranked 3rd on MAM and 1st on SON).

However, considering the total area and the annual precipitation, the values of the error metric ϵ_m leading to the rankings are not very different among the first 9 models, being the ϵ_m value of the model ranked 9th (i.e., CM5-ALAD) only 37% higher than the best. The difference with respect to the best ϵ_m value is lower than 50% in DJF for the first 7 models, in MAM for the first 5 models, in JJA for the first 6 models and in SON for the first 7 models. The models providing always (i.e., considering both the annual and the seasonal values) differences lower than 50% with respect to the best ϵ_m value are Had-RACM, ECE-CCLM and Had-CCLM.

Figure 15xxx shows a comparison between the ranking of interannual variability of annual precipitation and the average position in the ranking of seasonal variability of seasonal precipitation. It highlights possible deviations of the performances of the models at different time scales (the higher the deviation, the higher the distance from the bisector). When considering the seasonal scale, the reduced performance of CM5-ALAD is evident, such as the better ranking of MPI-CCLM. In general, the best models, both at the interannual and, averagely, the seasonal scale, are Had-RACM and ECE-CCLM, followed

by the two versions of ECE-RACM and two other CCLM models (namely, MPI-CCLM and Had-CCLM, the latter being penalized by the relatively lower ranking in winter).

~~With reference to~~Focusing on drought analysis, box-plots highlight a relevant variability in the frequency distribution of the error for all the considered drought characteristics. As for drought duration (Fig. 11a and d), the differences among the models appear more evident at the annual scale, while at the seasonal scale the models' behaviour looks rather similar. A general agreement can be observed between the box-plots of drought accumulated deficit at the annual and the seasonal scale (Fig. 11b and e), where the IPS-WRF is confirmed as the worst model. Concerning drought intensity (Fig. 11c and f), CM5-RC4 provides a very poor performance at the annual scale, but a light improvement can be observed at the seasonal scale. ~~With reference to~~As for the return periods (Fig. 12 and 13), the seasonal scale emphasizes the poor quality of CM5-ALAD, which is also confirmed at the annual scale, together with CM5-ALAR, ECE-RACMr12 and MPI-CCLM. Finally, the rankings combining the performance of the models to simulate maximum drought intensity and return period of drought event of fixed duration (Fig. 14a and b) agree in considering Had-RACM and ECE-CCLM as the best models both at the annual and seasonal scale.

5.2 Analyses across different variables

In terms of interannual variability, it's worth observing that, while MPI models appear the most suitable for mean temperature regardless of the area of investigation, especially regarding those in combination with REMO and CCLM RCMs, this is not the case for precipitation, although both the boxplot and the Taylor diagram indicate some potential of the MPI-CCLM for precipitation (Fig. 6a7 and 6b8). The boxplots for both variables displayed a large spatial variability of the errors, suggesting the limited capacity of RCMs to properly capture spatial variations of both temperature and precipitation patterns. Regarding precipitation, a similar result was obtained by Mascaro et al. (2018) for the Sardinia region. To find a possible explanation, we decided to investigate possible relationships between the amount of the errors and the cells' mean altitude. In particular, correlation analyses between the elevation and the mean and the standard deviation of the mean annual air temperature and precipitation errors were carried out. Nonetheless, results, here not shown for the sake of brevity, did not provide significant correlations.

~~Given the methodology adopted for identifying droughts, based on the annual values of precipitation, it is not surprising that the drought intensity ranking fits quite well that of the annual precipitation. However, models' performances in the drought intensity ranking are closer each other: the first 12 models show differences with respect to the best ϵ_m value (provided, once more, by Had RACM) lower or equal to 50%, while only 5 models (IPS RCA4, Nor HIRH, MPI REMO, IPS WRF and, especially, CM5 RCA4) show differences near to or higher than 100%.~~

Turning to seasonal variability, some similarities between mean temperature and precipitation arise in spring, with the ECE-CCLM model looking valuable for both variables. ECE models also perform well in winter but in combination with different RCMs (i.e. HIRH for temperature and RACM for precipitation). In summer, MPI-REMO_{r2} model is the best option for

precipitation but works well also for mean temperature, mainly for Zones 5 and 6. In autumn, MPI-REMO_{r2} is once again the best performing model but for mean temperature only. Alternatively, MPI-CCLM looks valuable for both mean temperature and precipitation during this season, as also confirmed by the Taylor diagrams (Figs. 45 and 94). Finally, the best models for drought intensity broadly recall those identified for annual precipitation, specifically for ECE-CCLM and Had-RACM.

The skills of CMs in reproducing drought characteristics and variability of precipitation are significantly linked. Drought characteristics, derived through the application of theory of runs, are functions of the departure from the thresholds rather than of the modelled precipitation itself. In other words, although a CM could significantly underestimate or overestimate annual and seasonal precipitation values (i.e. the data in the boxplots in Fig. 6a and 8 may look loosely grouped and the medians very far from 0), still it could provide good performance in terms of drought characteristics simulation if it can reproduce time variability. It is interesting to observe that the distribution of the percentage error of drought intensity (Fig. 11c and f) is, in general, less scattered than that related to the accumulated deficit (Fig. 11b and e); therefore, one can conclude that a partial error compensation occurs when the modelled accumulated deficit is divided by the modelled duration. Despite the differences in the percentage errors, however, there is a general agreement in the identification of the best and, mainly, the worst models, also confirmed by the ranking of the models in reproducing drought intensity and return period of drought events with fixed duration (Fig. 14a and b) both at the annual and the seasonal time scale.

Models' performances in the drought ranking (Fig. 14a) are closer each other: the first 12 models show differences with respect to the best c_{int} value (provided, once more, by Had-RACM) lower or equal to 50%, while only 5 models (IPS-RCA4, Nor-HIRH, MPI-REMO, IPS-WRF and, especially, CM5-RCA4) show differences near to or higher than 100%.

5.3 Impact of GCM and RCM choice and different realizations

Overall, no GCM prevails on the others because the RCMs deeply affect the final results. For example, concerning annual precipitation, the simulations relying on the Had GCM provide two high-ranked models (i.e., Had-CCLM and ~~HAD~~Had-RACM) and a low-ranked model (i.e., Had-RCA4). In the case of precipitation, only one among the GCMs used more than once (i.e. IPS) some indications come only from the two less used GCMs, i.e. IPS (two models) and Nor (one model), which providecoherently provides always bad results (IPS).

Concerning the most used RCMs, CCLM seems able to improve performances always with temperature (Fig. 34) and in most cases with precipitation (Fig. 79). Also, RACM usually provides high rankings with precipitation, while lower performances are found with temperature. The five occurrences of RCA4 very seldom provide high rankings with precipitation, as well as the two occurrences of HIRH.

It is of some interest to analyse the behaviour of different realizations of the same CM, which provide insight into the effects of the variability of a multi-member GCM ensemble (von Trentini et al., 2019). In this study, two cases occur, i.e., ECE_RACM and MPI_REMO. Looking at all the box-plots and Taylor diagrams, the two versions of the models behave rather coherently. Nevertheless, because of the variability of the overall model ensemble, usually, they are not ranked in subsequent positions.

630 E.g., considering annual drought intensity-ranking and the total area, ECE-RACM is ranked 2nd-3rd and ECE-RACMr12 17th,
while in the seasonal drought ranking MPI-REMO is ranked 17th-7th and MPI-REMO^r2 12th-15th. This result highlights that,
at least to a certain extent, the variability induced by different driving ensemble members is of the same order of the variability
given by other GCM-RCM combinations. On the other hand, given the similar performances of the different realizations
pointed out by the box-plots and Taylor diagrams, it is confirmed that rather slight differences in models' performance can be
635 found even for distances of 4-5 positions in the rankings.

5.4 Overall ranking and comparison with literature

For a final evaluation of the models, an overall ranking criterion was applied. This ranking takes into consideration both the
skills of the considered GCM-RCMs models to replicate annual precipitation and temperature variability, as well as drought
characteristics. As shown in Fig. 176, the models with the best overall performances, both in the whole case study area and in
640 the six climatically homogeneous zones are those in combination with CCLM RCMs, with the significant exception of Had-
RACM, which is ranked 1st considering the total area and Zones 2 and 4 for the annual time scale. Generally, the worst models
both at the annual and the seasonal scale are Nor-HIRH, IPS-WRF, and CM5-RCA4, although at the seasonal scale also CM5-
ALAD and CM5-ALAR have poor performances.

An attempt can be made to compare the results of our ranking exercise with similar studies. Such a comparison is here limited
645 to the Euro-CORDEX climate models for which, indeed, only a few studies do exist. Perhaps the study from Kotlarski et al.
(2014) allows the most interesting comparisons for our purposes, being focused on both precipitation and temperature at
seasonal and yearly timescales, and covering all areas of Europe, with specific results for the Mediterranean area. Models here
denoted as CCLM (CLMCOM-11 in the mentioned study) perform well in reproducing annual temperature and precipitation
in both studies. Differences arise for precipitation in the MAM season, since CCLM models show poor performances according
650 to Kotlarski et al. (2014), in contrast to our findings. Mascaro et al. (2018), whose study is focused on the Sardinia region
(Italy), also found that the Had-RACM and ECE-CCLM models perform well in reproducing annual precipitation, while there
is no agreement on the CM5-ALAD model. At the seasonal level, ECE-RACMr12, MPI-REMO^r2 and MPI-CCLM perform
well in both studies in the seasons DJF, JJA, SON respectively, while, in contrast to our results, in the MAM season the ECE-
CCLM does not perform well. These differences in the ranking could be partially due to the different observational datasets
655 used, which have found to play a key role in climate model evaluations (Kotlarski et al., 2017).

6 Conclusions

In the presentis study, we compared the skill of nineteenseveral EURO-CORDEX RCMs at 0.11° (~ 12.5 km) grid spatial
resolution in reproducing the annual and seasonal temperature and precipitation regime, as well as several drought
patternsfeatures, observed in the period 1971-2000 in a dense network of rain gauges in Sicily and Calabria regions (Southern
660 Italy).

From our investigation study a few general and specific conclusions can be drawn. From a general point of view, the model combinations are able to simulate temperature better than precipitation^s, even though important biases do exist in both variables. Models which are reliable in simulating precipitation may not be the same respect to temperature. This is the case, for instance, of the ECE-RACM model which is in the top ranks for precipitation, while being in the lower ranks for temperature. Models that perform ~~the~~ best for precipitation do almost the same for ~~all~~ drought features. Differences between the rankings of annual respect to seasonal characteristics do exist, but ~~top-ranking~~top-ranking models at the annual scale ~~do~~ mostly perform well in the single seasons^s, both for precipitation and temperature. -Looking more specifically to the models, the Had-RACM, ECE-CCLM, Had-CCLM and ECE-RACM are those that perform ~~the~~ best for precipitation and drought, while the CM5-RCA4 and IPS-WRF are those that perform worst. For temperature, models that perform best are MPI-CCLM, MPI-REMO and Had-CCLM, while the worst~~te~~ are CM5-ALAD, ECE-RCA4, ECE-RACM and CM-RCA4. As can be seen ~~the~~ Had-CCLM performs well for both precipitation and temperature, while the CM5-RCA4 performs bad for both. The CMs are more capable to simulate both annual and seasonal mean air temperature than precipitation and drought characteristics, with high correlation values. There is a general agreement among the models to underestimate annual precipitation and mainly mean annual temperature. Most of the models show deficiencies in the simulation of seasonal precipitation, especially concerning summer values, requiring further investigation. Overall, our analyses illustrate that the best performing models depend on the specific property of the investigated variable, as well as the temporal and the spatial scale of interest. It provides a general overview of model performance without aiming at ultimately explaining the biases of individual models. We reserve to carry out detailed investigations in follow up studies that will address specific aspects of model performance and investigate the causes leading to the model biases for possible bias ~~correction~~. Results of this study reveal insight on RCMs performances in small-scale regions, which are often targeted by impact studies and have so far received less attention, and provide some guidance to select the best models about the variable and the area under investigation. This is a key issue before addressing projections changes in the evolution of extreme hydro-meteorological events, such as drought characteristics (frequency, duration, and magnitude).

Data availability. Ground-based datasets are provided, upon request, by the “Centro Funzionale Multirischi – ARPACAL” (<http://www.cfd.calabria.it/> - for Calabria) and the “Osservatorio delle Acque – Regione Sicilia” (www.osservatorioacque.it – for Sicily). Climate data are freely available at the EURO-CORDEX WEBSITE (<https://www.euro-cordex.net/>).

Author contribution. A.S., D.J.P. and B.B. designed the experiments. A.S. and B.B. contributed to sample preparation and preliminary data analysis. D.J.P. and P.N. performed the main computations. A.C. and G.M. supervised the research. All authors discussed the results and contributed to the final manuscript.

Competing interests. The paper belongs to the special issue "Recent advances in drought and water scarcity monitoring, modelling and forecasting". Dr. Brunella Bonaccorso, co-author of the paper, is co-editor of the aforementioned special issue.

Therefore, she ~~will~~did not assume the role of handling editor for this paper. The other authors declare that they have no conflict of interest.

Acknowledgements. The authors thank the “Centro Funzionale Multirischi – ARPACAL” and the “Osservatorio delle Acque – Regione Sicilia” for providing the observed precipitation and temperature data.

References

Arnell, N. et al. (2001). Hydrology and Water Resources. In McCarthy, J. J. et al. (eds.), *Climate Change 2001: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, Cambridge, UK, pp. 192-234.

Adeniyi, M. O. and Dilau, K. A.: Assessing the link between Atlantic Niño 1 and drought over West Africa using CORDEX regional climate models, *Theor. Appl. Climatol.*, 131(3–4), 937–949, doi:10.1007/s00704-016-2018-0, 2018.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., Reinhardt, T., Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, *Mon. Weather Rev.*, 139(12), 3887–3905, doi:10.1175/MWR-D-10-05013.1, 2011.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C. and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, *Geosci. Model Dev.*, 6(3), 687–720, doi:10.5194/gmd-6-687-2013, 2013.

Bonaccorso, B., Cancelliere, A. and Rossi, G.: An analytical formulation of return period of drought severity, *Stoch. Environ. Res. Risk Assess.*, 17(3), doi:10.1007/s00477-003-0127-7, 2003.

Bonaccorso, B., Cancelliere, A. and Rossi, G. (2012). *Methods for Drought Analysis and Forecasting*. In: *Methods and Applications of Statistics in the Atmospheric and Earth Sciences*. p. 150-184, Hoboken, John Wiley and Sons, ISBN: 9780470503447.

Bonaccorso, B., Peres, D. J., Cancelliere, A. and Rossi, G.: Large Scale Probabilistic Drought Characterization Over Europe, *Water Resour. Manag.*, 27(6), 1675–1692, doi:10.1007/s11269-012-0177-z, 2013.

Bonaccorso, B., Peres, D.J., Castano, A. and Cancelliere, A.: SPI-Based Probabilistic Analysis of Drought Areal Extent in Sicily, *Water Resour. Manag.*, 29(2), 459-470, doi: 10.1007/s11269-014-0673-4, 2015a.

Bonaccorso, B., Cancelliere, A. and Rossi, G.: Probabilistic forecasting of drought class transitions in Sicily (Italy) using Standardized Precipitation Index and North Atlantic Oscillation Index, *J. Hydrol.*, 526, 136-150, doi: 10.1016/j.jhydrol.2015.01.070, 2015b.

Bordi, I. and Sutera, A.: An analysis of drought in Italy in the least fifty years, *Nuovo Cimento C*, Vol 25, No. 2, pp. 185-206, 2002.

Cancelliere, A. and Salas, J. (2004). *Drought length properties for periodic stochastic hydrological data*. *Water Resour. Res.*, 10, 1–13.

725

Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I. and Rinke, A.: The HIRHAM Regional Climate Model Version 5 (beta), Tech. Rep. 06-17;, 5, 1–22, 2007.

Christensen JH, Kjellström EK, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. *Clim Res* 44:179–194.

730 Coppola E, Giorgi F, Rauscher SA, Piani C (2010) Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate models. *Clim Res* 44:121–134.

Colin, J., Déqué, M., Radu, R. and Somot, S.: Sensitivity study of heavy precipitation in Limited Area Model climate simulations: influence of the size of the domain and the use of the spectral nudging technique, *Tellus A Dyn. Meteorol. Oceanogr.*, 62(5), 591–604, doi:10.1111/j.1600-0870.2010.00467.x, 2010.

735 Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A. and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, *Geosci. Model Dev.*, 4(4), 1051–1075, doi:10.5194/gmd-4-1051-2011, 2011.

De Troch, R., Hamdi, R., Van de Vyver, H., Geleyn, J.-F., Termonia, P., Troch, R. De, Hamdi, R., Vyver, H. Van de, Geleyn, J.-F. and Termonia, P.: Multiscale Performance of the ALARO-0 Model for Simulating Extreme Summer Precipitation Climatology in Belgium, *J. Clim.*, 26(22), 8895–8915, doi:10.1175/JCLI-D-12-00844.1, 2013.

740 Diasso, U. and Abiodun, B. J.: Drought modes in West Africa and how well CORDEX RCMs simulate them, *Theor. Appl. Climatol.*, 128(1–2), 223–240, doi:10.1007/s00704-015-1705-6, 2017.

Di Virgilio, G., Evans, J. P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., Andrys, J., Hoffmann, P., Katzfey, J. J. and Rockel, B.: Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors, *Clim. Dyn.*, 0(0), 0, doi:10.1007/s00382-019-04672-w, 2019.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N. and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, *Clim. Dyn.*, 40(9–10), 2123–2165, doi:10.1007/s00382-012-1636-1, 2013.

755 Endris, H. S., Omondi, P., Jain, S., Lennard, C., Hewitson, B., Chang’a, L., Awange, J. L., Dosio, A., Ketiem, P., Nikulin, G., Panitz, H. J., Büchner, M., Stordal, F. and Tazalika, L.: Assessment of the performance of CORDEX regional climate models in simulating East African precipitation, *J. Clim.*, 26(21), 8453–8475, doi:10.1175/JCLI-D-12-00708.1, 2013.

- Foley, A. and Kelman, I.: EURO-CORDEX regional climate model simulation of precipitation on Scottish islands (1971–2000): model performance and implications for decision-making in topographically complex regions, *Int. J. Climatol.*, 38(2), 1087–1095, doi:10.1002/joc.5210, 2018.
- Gampe, D., Schmid, J. and Ludwig, R.: Impact of Reference Dataset Selection on RCM Evaluation, Bias Correction, and Resulting Climate Change Signals of Precipitation. *Journal of Hydrometeorology*, 20(9), 1813–28, doi: 10.1175/JHM-D-18-0108.1, 2019.
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J. and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, *J. Adv. Model. Earth Syst.*, 5(3), 572–597, doi:10.1002/jame.20038, 2013.
- Giorgi, F.: Climate change hot-spots, *Geophys. Res. Lett.*, 33(8), 1–4, doi:10.1029/2006GL025734, 2006.
- Giorgi, F., Lionello, P. Climate change projections for the Mediterranean region. *Global and Planetary Change*, 63, 90–104. <https://doi.org/10.1016/j.gloplacha.2007.09.005>, 2008.
- Gonzalez, J. and Valdes, J. (2003). Bivariate drought recurrence analysis using tree ring reconstructions. *J. Hydrol. Eng.*, 8, 247–258.
- Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, *Int. J. Climatol.*, 34(3), 623–642, doi:10.1002/joc.3711, 2014.
- Hart, O. E. and Halden, R. U.: On the need to integrate uncertainty into U.S. water resource planning. *Sci. Total Environ.*, 691, 1262–1270, <https://doi.org/10.1016/j.scitotenv.2019.07.164>, 2019.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D. and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res. Atmos.*, 113(20), doi:10.1029/2008JD010201, 2008.
- Hazeleger, W., Severijns, C., Semmler, T., Ștefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., Bougeault, P., Caballero, R., Ekman, A. M. L., Christensen, J. H., van den Hurk, B., Jimenez, P., Jones, C., Kållberg, P., Koenigk, T., McGrath, R., Miranda, P., van Noije, T., Palmer, T., Parodi, J. A., Schmith, T., Selten, F., Storelvmo, T., Sterl, A., Tapamo, H., Vancoppenolle, M., Viterbo, P. and Willén, U.: EC-Earth, *Bull. Am. Meteorol. Soc.*, 91(10), 1357–1364, doi:10.1175/2010BAMS2877.1, 2010.
- Huntington, T. G.: Evidence for intensification of the global water cycle: Review and synthesis, *J. Hydrol.*, 319(1–4), 83–95, doi:10.1016/j.jhydrol.2005.07.003, 2006.

- 790 IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp., 2014.
- IPCC, 2018: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield (eds.)].
- 795 Iversen, T., Bentsen, M., Bethke, I., Debernard, J. B., Kirkevåg, A., Seland, Ø., Drange, H., Kristjansson, J. E., Medhaug, I., Sand, M. and Seierstad, I. A.: The Norwegian Earth System Model, NorESM1-M – Part 2: Climate response and scenario projections, *Geosci. Model Dev.*, 6(2), 389–415, doi:10.5194/gmd-6-389-2013, 2013.
- 800 Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B. and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, *Reg. Environ. Chang.*, 14(2), 563–578, doi:10.1007/s10113-013-0499-2, 2014.
- 805 Kjellström, E., Thejll, P., Rummukainen, M., Christensen, J. H., Boberg, F., Christensen, O. B. and Maule, C. F.: Emerging regional climate change signals for Europe under varying large-scale circulation conditions, *Clim. Res.*, 56(2), 103–119, doi:10.3354/cr01146, 2013.
- 810 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.: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble, *Geosci. Model Dev.*, 7(4), 1297–1333, doi:10.5194/gmd-7-1297-2014, 2014.
- 815 Kotlarski, S., Szabó, P., Herrera, S., Rätty, O., Keuler, K., Soares, P.M., Cardoso, R.M., Bosshard, T., Pagé, C., Boberg, F., Gutiérrez, J.M., Isotta, F.A., Jaczewski, A., Kreienkamp, F., Liniger, M.A., Lussana, C. and Pianko-Kluczyńska, K.: Observational uncertainty and regional climate model evaluation: A pan-European perspective, *Int. J. Climatol.*, 39 (9), pp. 3730-3749, 2017.
- 820 [Llasat, M.C., Marcos, R., Turco, M., Gilabert, J., and Llasat-Botija, M.: Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia, *J. Hydrol.*, 541, 24-37, 10.1016/j.jhydrol.2016.05.040, 2016.](#)
- Mascaro, G., White, D.D., Westerhoff, P. [and](#) Bliss, N.: Performance of the CORDEX-Africa regional climate simulations in representing the hydrological cycle of the Niger river basin, *J Geophys Res*, 120 (24), pp. 12,425-12,444, 2015.

- Mascaro, G., Viola, F. and Deidda, R.: Evaluation of Precipitation From EURO-CORDEX Regional Climate Simulations in a Small-Scale Mediterranean Site, *J. Geophys. Res. Atmos.*, 123(3), 1604–1625, doi:10.1002/2017JD027463, 2018.
- Mendicino, G. and Versace P.: Integrated Drought Watch System: A Case Study in Southern Italy. *Water Resour Manage* 21, 1409–1428, <https://doi.org/10.1007/s11269-006-9091-6>, 2007.
- Meque, A. and Abiodun, B. J.: Simulating the link between ENSO and summer drought in Southern Africa using regional climate models, *Clim. Dyn.*, 44(7–8), 1881–1900, doi:10.1007/s00382-014-2143-3, 2015.
- 830 Park, C., Min, S. K., Lee, D., Cha, D. H., Suh, M. S., Kang, H. S., Hong, S. Y., Lee, D. K., Baek, H. J., Boo, K. O. and Kwon, W. T.: Evaluation of multiple regional climate models for summer climate extremes over East Asia, *Clim. Dyn.*, 46(7–8), 2469–2486, doi:10.1007/s00382-015-2713-z, 2016.
- Peres, D. J., Caruso, M. F. and Cancelliere, A.: Assessment of climate-change impacts on precipitation based on selected RCM projections, *European Water*, E.W. Publications, 59:9–15, 2017
- 835 Peres, D.J.; Modica, R.; Cancelliere, A. Assessing Future Impacts of Climate Change on Water Supply System Performance: Application to the Pozzillo Reservoir in Sicily, Italy. *Water* 2019, 11, 2531, doi:10.3390/w11122531
- Prein, A., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., and Nikulin, G.: Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: High resolution, high benefits? *Climate Dynamics*, 46(1-2), 383–412. doi.org/10.1007/s00382-015-2589-y, 2016.
- 840 Rencher, A. C.: *Multivariate Statistical Inference and Applications*, John Wiley and Sons, INC., 559 pp., 1998.
- Rockel, B., Will, A. and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), *Meteorol. Zeitschrift*, 17(4), 347–348, doi:10.1127/0941-2948/2008/0309, 2008.
- Rossi G., Benedini, M.: *Water Resources of Italy. Protection, Use and Control*; Springer International Publishing, 365 pp., 2020.
- 845 Schmidli, J., Goodess, C.M., Frei, C., Haylock, M.R., Hurrel, Y., Ribalaygua, J., Schmith, T.: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps, *J Geophys Res-Atmos*, 112 (4), art. no. D04105, 2007.
- Sen, Z. (1976): Wet and dry periods of annual flow series. *J. Hydraul. Div.*, 102, 1503–1514, 1976.
- Senatore, A., Mendicino, G., Smiatek, G. and Kunstmann, H.: (2011) Regional climate change projections and hydrological impact analysis for a Mediterranean basin in southern Italy. *J Hydrol* 399(1–2):70–92, 2011.
- 850 Senatore, A., Hejabi, S., Mendicino, G., Bazrafshan, J. and Irannejad, P.: Climate conditions and drought assessment with the Palmer Drought Severity Index in Iran: evaluation of CORDEX South Asia climate projections (2070–2099), *Clim. Dyn.*, 52(1–2), 865–891, doi:10.1007/s00382-018-4171-x, 2019.
- Senatore, A., Furnari, L., and Mendicino, G.: Impact of high-resolution sea surface temperature representation on the forecast of small Mediterranean catchments' hydrological responses to heavy precipitation, *Hydrol. Earth Syst. Sci.*, 24, 269–291, <https://doi.org/10.5194/hess-24-269-2020>, 2020.

- Smiatek, G., Kunstmann, H. and Senatore, A.: (2016) EURO-CORDEX regional climate model analysis for the Greater Alpine region: performance and expected future change. *J Geophys Res Atmos* 121:7710–7728.e, 2016.
- Smiatek, G. and Kunstmann, H.: Simulating Future Runoff in a Complex Terrain Alpine Catchment with EURO-CORDEX Data. *Journal of Hydrometeorology*, 20(9), 1925–40, doi: 10.1175/JHM-D-18-0214.1, 2019.
- Strandberg, G., Bärring, L., Hansson, U., Jansson, C., Jones, C., Kjellström, E., Kolax, M., Kupiainen, M., Nikulin, G., Samuelsson, P., Ullerstig, A. and Wang, S.: CORDEX scenarios for Europe from the Rossby Centre regional climate model RCA4, Rep. Meteorol. Climatol., 116(116), 2014.
- Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res- Atmos*, 106(D7), 7183–7192. <https://doi.org/10.1029/2000JD900719>, 2001.
- Taylor, K. E., Stouffer, R. J. and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, 93(4), 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.
- Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., Moseley, C., Pfeifer, S., Rechid, D., Remedio, A. R., Ries, H., Petersen, J., Preuschmann, S., Raub, T., Saeed, F., Sieck, K. and Weber, T.: How does a regional climate model modify the projected climate change signal of the driving GCM: A study over different CORDEX regions using REMO, *Atmosphere (Basel)*, 4(2), 214–236, doi:10.3390/atmos4020214, 2013.
- Torma, C., Giorgi, F., and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain- Precipitation over the Alps. *J Geophys Res-Atmos*, 120, 3957–3972. doi.org/10.1002/2014JD022781, 2015.
- Um, M. J., Kim, Y. and Kim, J.: Evaluating historical drought characteristics simulated in CORDEX East Asia against observations, *Int. J. Climatol.*, 37(13), 4643–4655, doi:10.1002/joc.5112, 2017.
- Versace, P., Ferrari, E., Gabriele, S., and Rossi, F. (1989). “Valutazione delle piene in Calabria.” *Geodata*, 30 (in Italian).
- Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L. and Chauvin, F.: The CNRM-CM5.1 global climate model: description and basic evaluation, *Clim. Dyn.*, 40(9–10), 2091–2121, doi:10.1007/s00382-011-1259-y, 2013.
- van Meijgaard, E., van Ulft, B., van de Berg, W. J., Bosveld, F. C., van den Hurk, B., Lenderink, G. and Siebesma, A. P.: The KNMI regional atmospheric climate model RACMO version 2.1 (KNMI TR-302), Tech. Rep., Technical Report TR-302, 2008.
- van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. and Rose, S. K.: The representative concentration pathways: An overview, *Clim. Change*, 109(1), 5–31, doi:10.1007/s10584-011-0148-z, 2011.
- von Trentini, F., Leduc, M. and Ludwig, R. Assessing natural variability in RCM signals: comparison of a multi-model EURO-CORDEX ensemble with a 50-member single model large ensemble. *Clim Dyn* 53, 1963–1979, <https://doi.org/10.1007/s00382-019-04755-8>, 2019.

- Wagner, S., Kunstmann, H. and Bardossy, A.: Uncertainties in water balance estimations due to scarce meteorological information : a case study for the White Volta catchment in West Africa, *Clim. Res.*, (July), 86–97, 2007.
- Wu, F.T., Wang, S.Y., Fu, C.B., Qian, Y., Gao, Y., Lee, D.K., Cha, D.H., Tang, J.P., Hong, S.Y.: Evaluation and projection of summer extreme precipitation over east Asia in the regional model inter-comparison project, *Clim Res*, 69 (1), pp. 45-58, 2016.
- 895 Yevjevich, V. An objective approach to definitions and investigations of continental hydrologic droughts. Hydrology paper 23, Colorado State University, Fort Collins, Colorado, 1967.

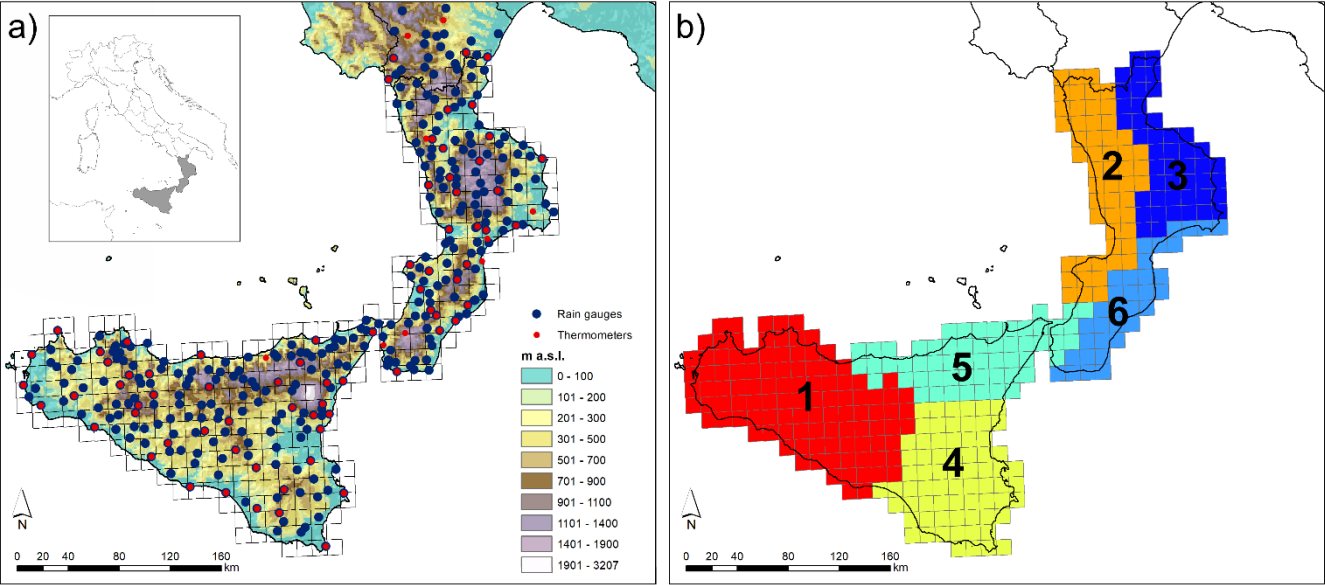


Figure 1. a) Study area (Calabria is the southernmost peninsula of Italy and Sicily is the neighbouring island) with the locations of the gauges of the high-density observational network and the CORDEX reference grid; b) the six homogeneous zones identified through PCA.

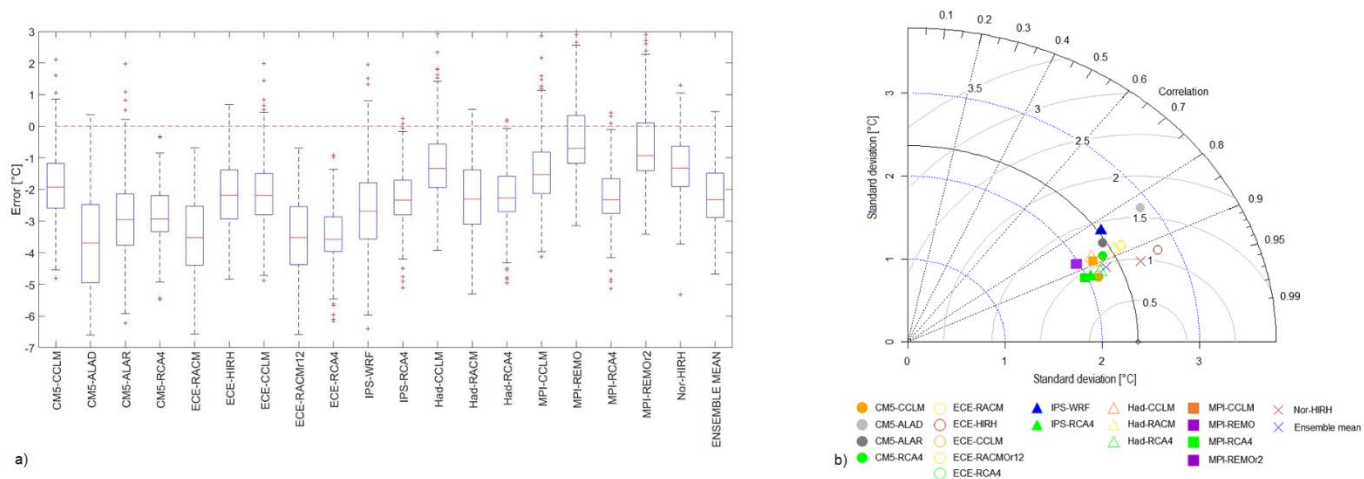


Figure 2. a) Box-plots representing the frequency distribution of RCMs errors in mean annual temperature for the whole study area. b) Taylor diagram comparing models performances in reproducing the interannual variability of mean annual temperature for the whole study area.

	ANNUAL						
	6	7	7	4	6	5	6
CM5-CCLM	19	18	19	19	17	17	18
CM5-ALAD	14	15	13	15	13	13	15
CM5-ALAR	15	14	14	12	16	15	13
CM5-RCA4	16	16	15	17	14	14	16
ECE-RACM	10	10	11	9	10	9	10
ECE-HIRH	8	8	8	8	7	8	9
ECE-CCLM	18	19	18	18	18	18	19
ECE-RACMr12	17	17	17	16	19	19	17
ECE-RCA4	12	13	12	14	12	6	11
IPS-WRF	4	5	4	1	4	7	1
IPS-RCA4	3	2	3	3	3	4	5
Had-CCLM	11	11	10	11	9	10	12
Had-RACM	7	6	6	5	8	11	8
Had-RCA4	1	3	1	2	1	1	3
MPI-CCLM	2	1	2	6	2	2	2
MPI-REMO	9	9	9	7	11	12	7
MPI-RCA4	5	4	5	10	5	3	4
MPI-REMOv2	13	12	16	13	15	16	14
Nor-HIRH							
	Total Area	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6

Figure 3. RCMs ranking with respect to interannual variability of mean annual temperature, for the entire area and the climatically homogenous zones. Taylor diagram comparing models performances in reproducing the interannual variability of mean annual temperature for the whole study area.

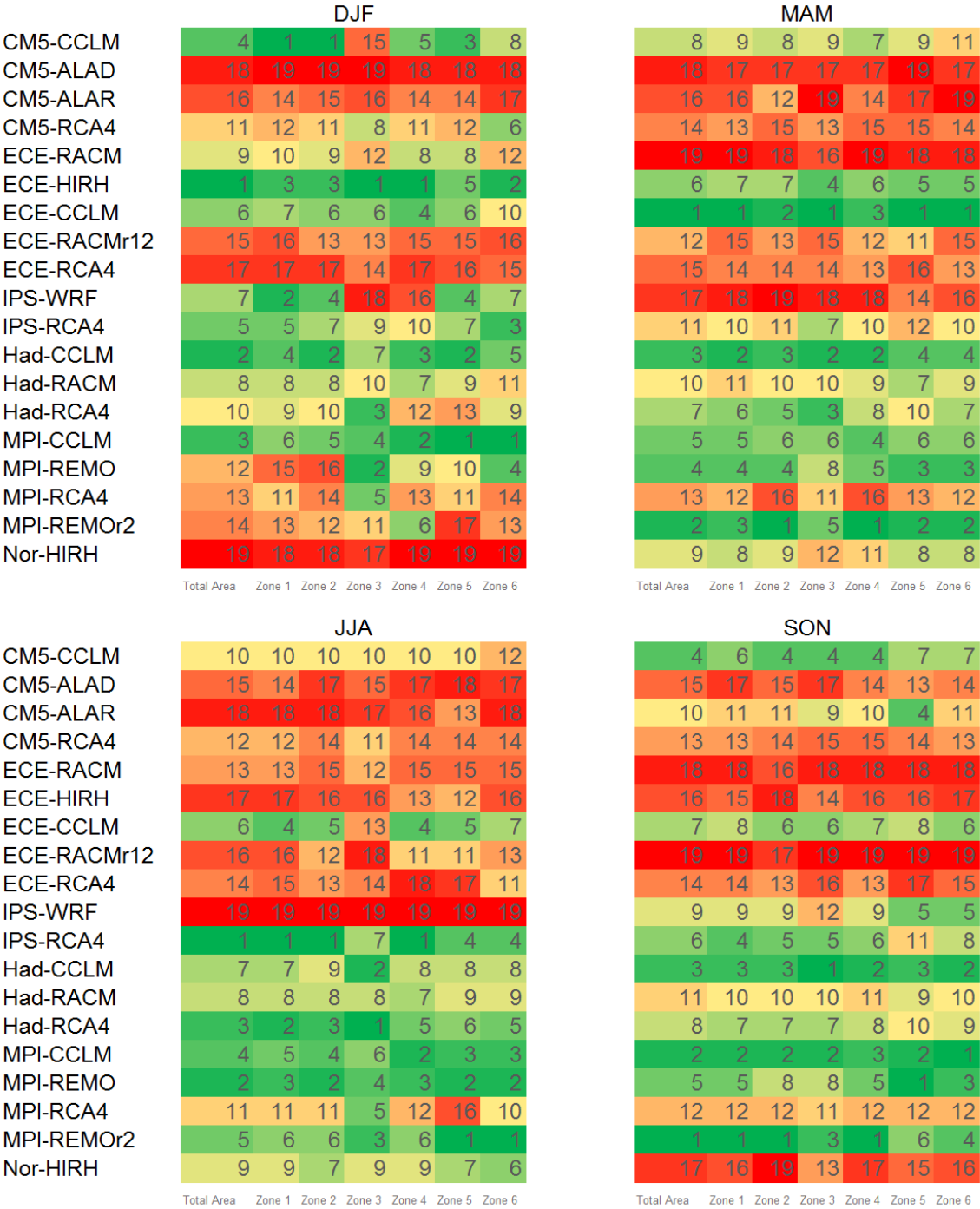
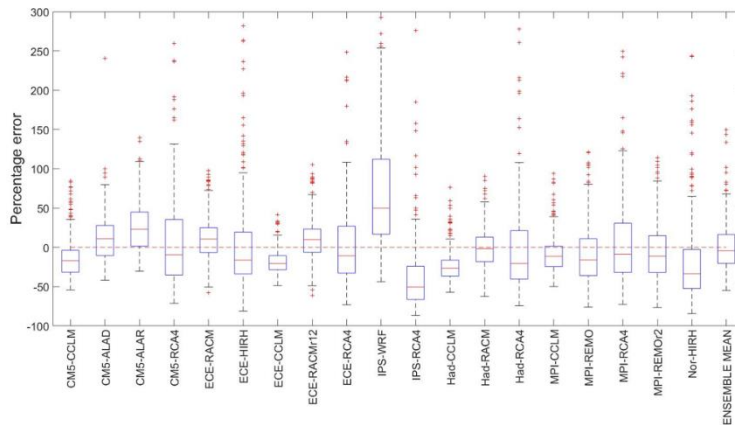
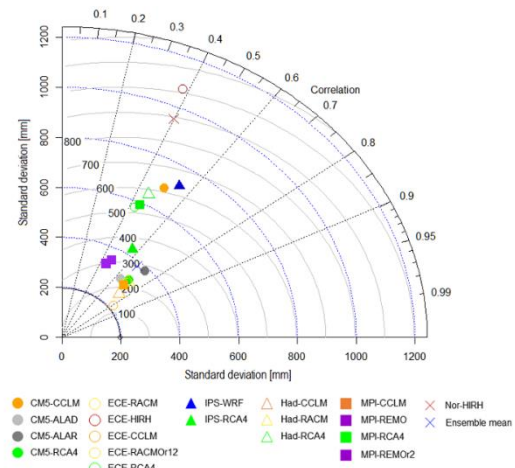


Figure 5. RCMs ranking with respect to seasonal variability of mean annual temperature, for the entire area and the climatically homogenous zones. Taylor diagram comparing models performances in reproducing the seasonal variability of mean annual temperature for the whole study area.



a)



b)

930

Figure 6. a) As Fig. 2a but for annual precipitation. b) As Fig. 2b but for annual precipitation. RCMs ranking with respect to seasonal variability of mean annual temperature, for the entire area and the climatically homogenous zones.

	ANNUAL						
CM5-CCLM	7	2	12	4	6	9	4
CM5-ALAD	3	5	4	5	11	2	1
CM5-ALAR	9	13	3	9	4	7	12
CM5-RCA4	17	16	14	16	15	18	19
ECE-RACM	5	3	1	14	8	5	8
ECE-HIRH	15	11	13	15	18	17	18
ECE-CCLM	2	4	6	1	1	1	6
ECE-RACMr12	4	6	5	6	5	4	3
ECE-RCA4	11	9	10	12	14	11	11
IPS-WRF	19	19	19	11	12	19	15
IPS-RCA4	16	18	17	19	17	10	14
Had-CCLM	6	8	7	2	3	6	7
Had-RACM	1	1	2	3	2	3	2
Had-RCA4	14	12	15	17	16	14	17
MPI-CCLM	8	7	9	8	7	8	5
MPI-REMO	12	14	16	10	9	15	9
MPI-RCA4	13	10	8	13	13	12	13
MPI-REMO _{r2}	10	15	11	7	10	13	10
Nor-HIRH	18	17	18	18	19	16	16

Total Area Zone 1 Zone 2 Zone 3 Zone 4 Zone 5 Zone 6

Figure 7. ~~As Fig. 3 but for annual precipitation. As Fig. 2 but for annual precipitation.~~

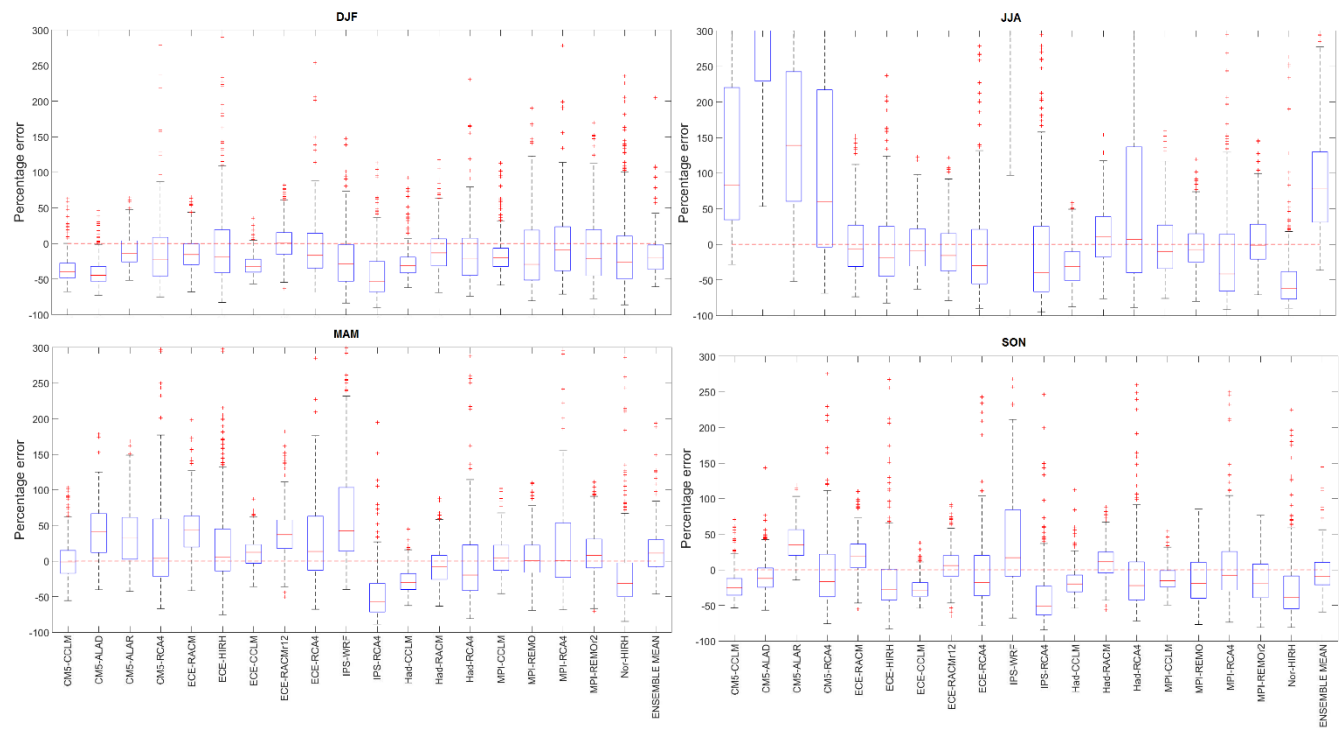


Figure 8. Box-plots representing the frequency distribution of RCMs percentage errors in seasonal precipitation for the whole study area. As Fig. 3 but for annual precipitation.

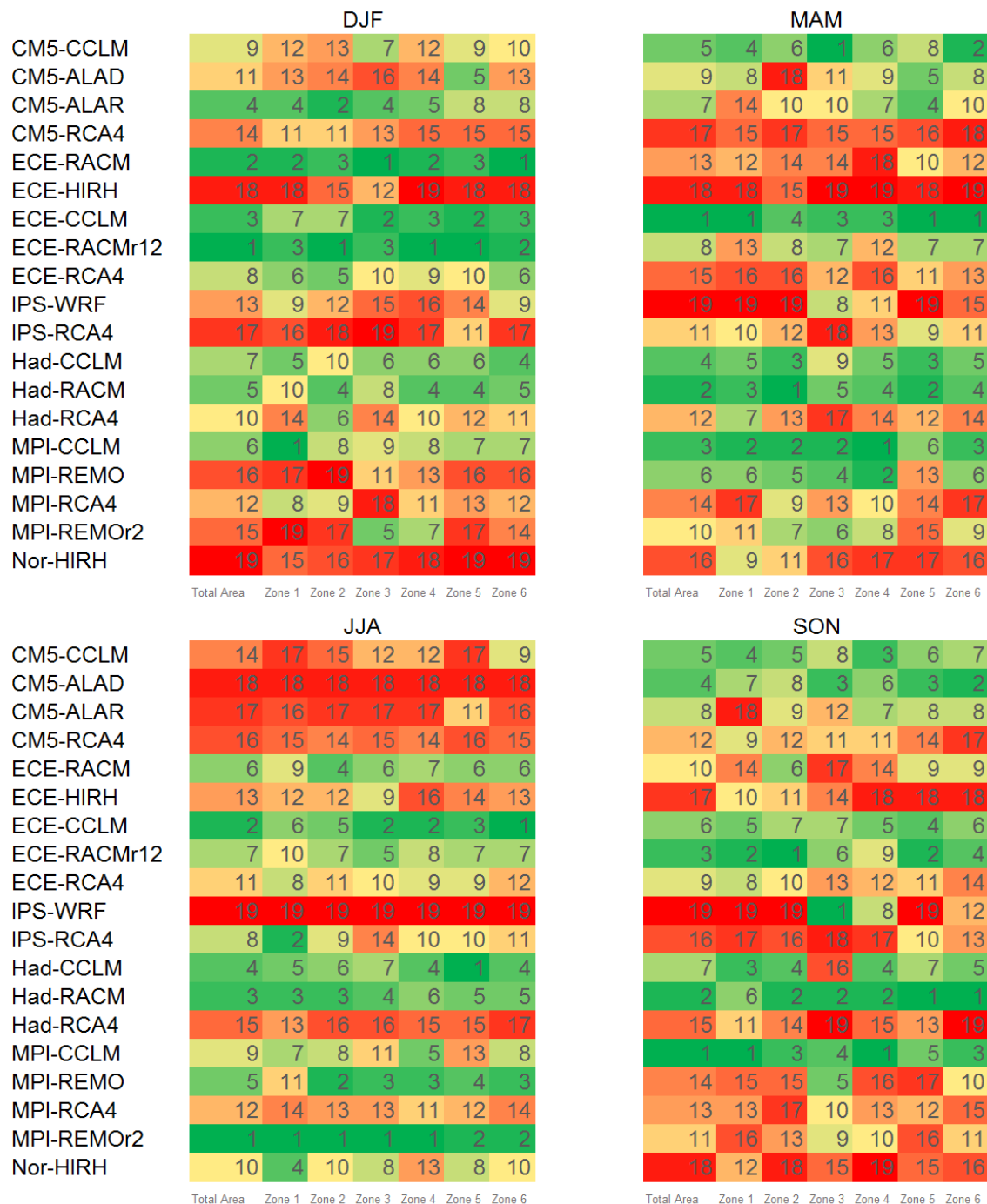


Figure 10. As Fig. 5 but for seasonal precipitation. Box-plots representing the frequency distribution of RCMs percentage errors in seasonal precipitation for the whole study area.

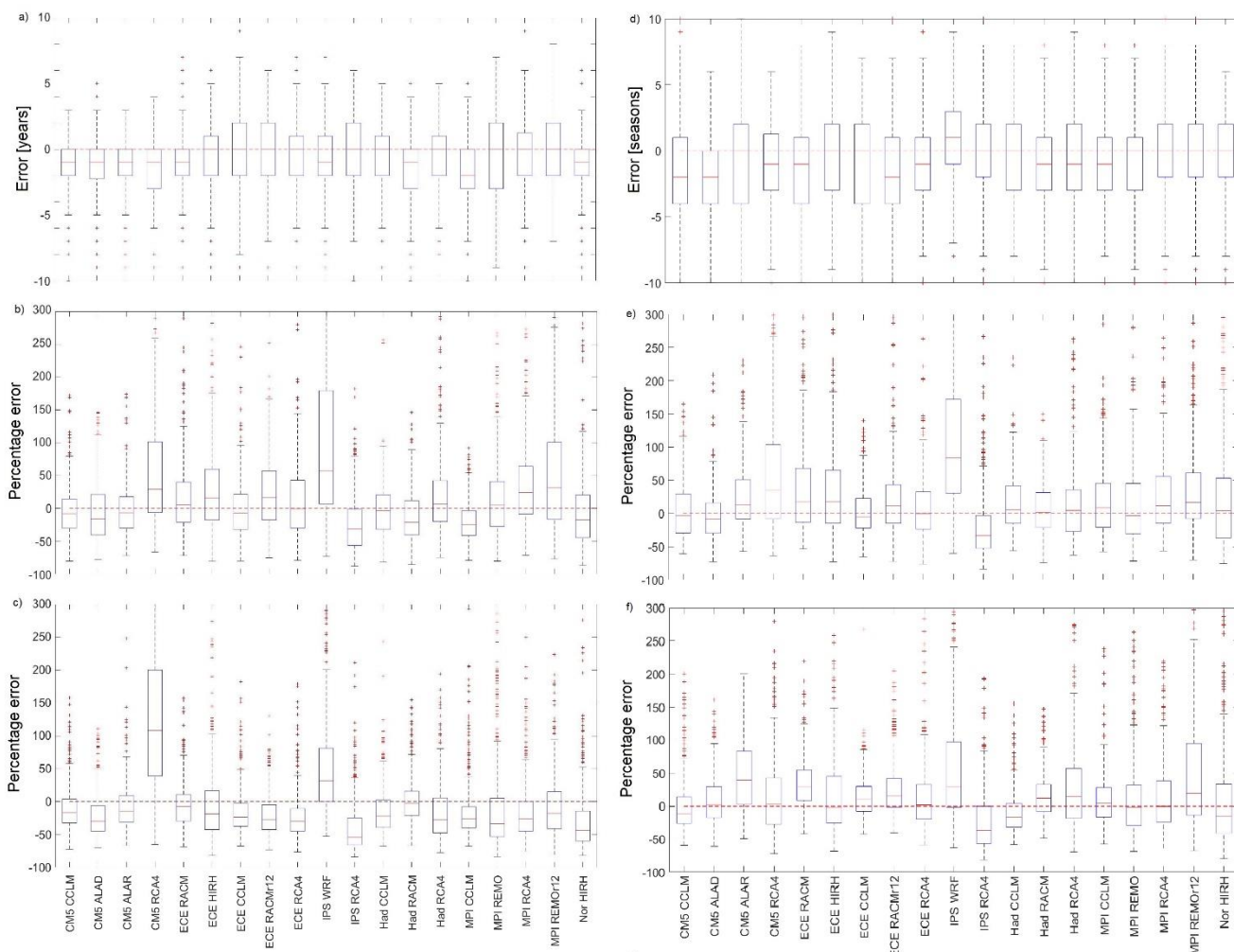


Figure 11. As Fig. 5 but for mean annual precipitation for Sicily and Calabria. a) Box-plots representing the frequency distribution of RCMs percentage errors in: a) maximum drought duration (annual analysis); b) Box-plots representing the frequency distribution of RCMs percentage errors in maximum drought accumulated deficit (annual analysis); c) Box-plots representing the frequency distribution of RCMs percentage errors in maximum drought intensity (annual analysis); d) Box-plots representing the frequency distribution of RCMs percentage errors in maximum drought duration (seasonal analysis); e) Box-plots representing the frequency distribution of RCMs percentage errors in maximum drought accumulated deficit (seasonal analysis); f) Box-plots representing the frequency distribution of RCMs percentage errors in maximum drought intensity (seasonal analysis).

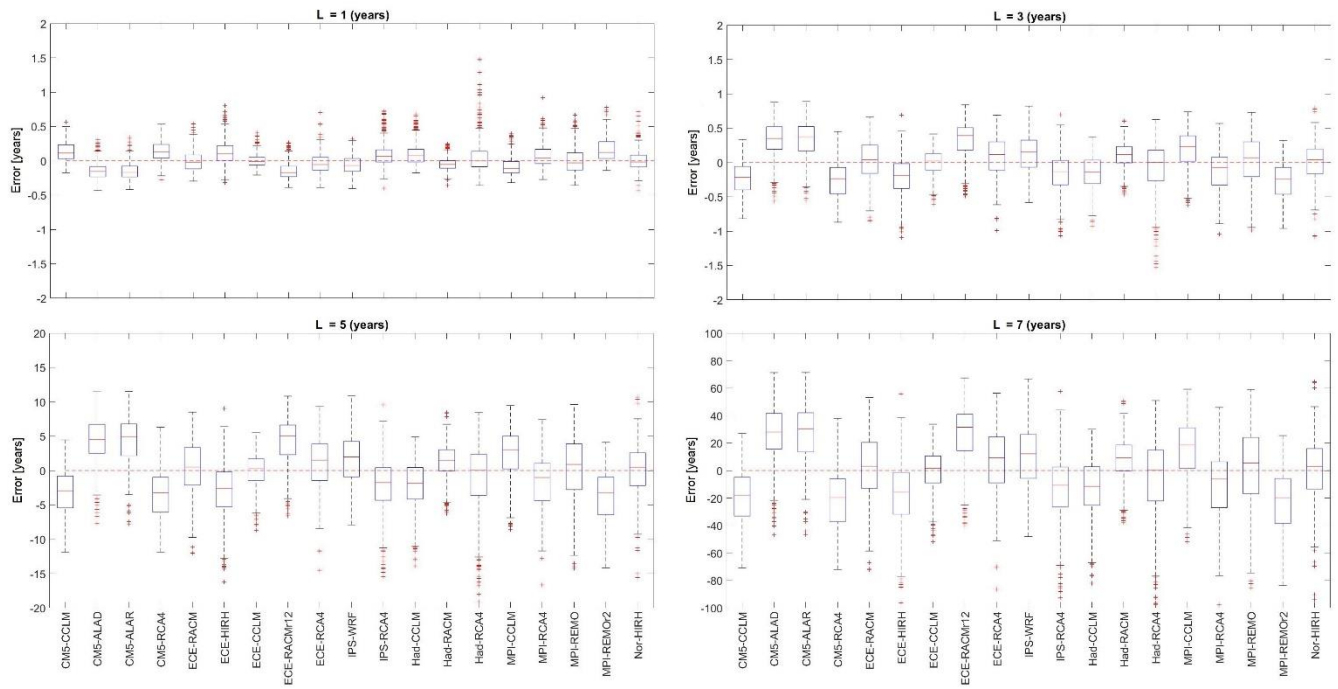


Figure 12. Box-plots representing the frequency distribution of RCMs percentage errors in the return period of drought events of duration L equal to 1, 3, 5 and 7 years. As Fig. 6 but for seasonal precipitation

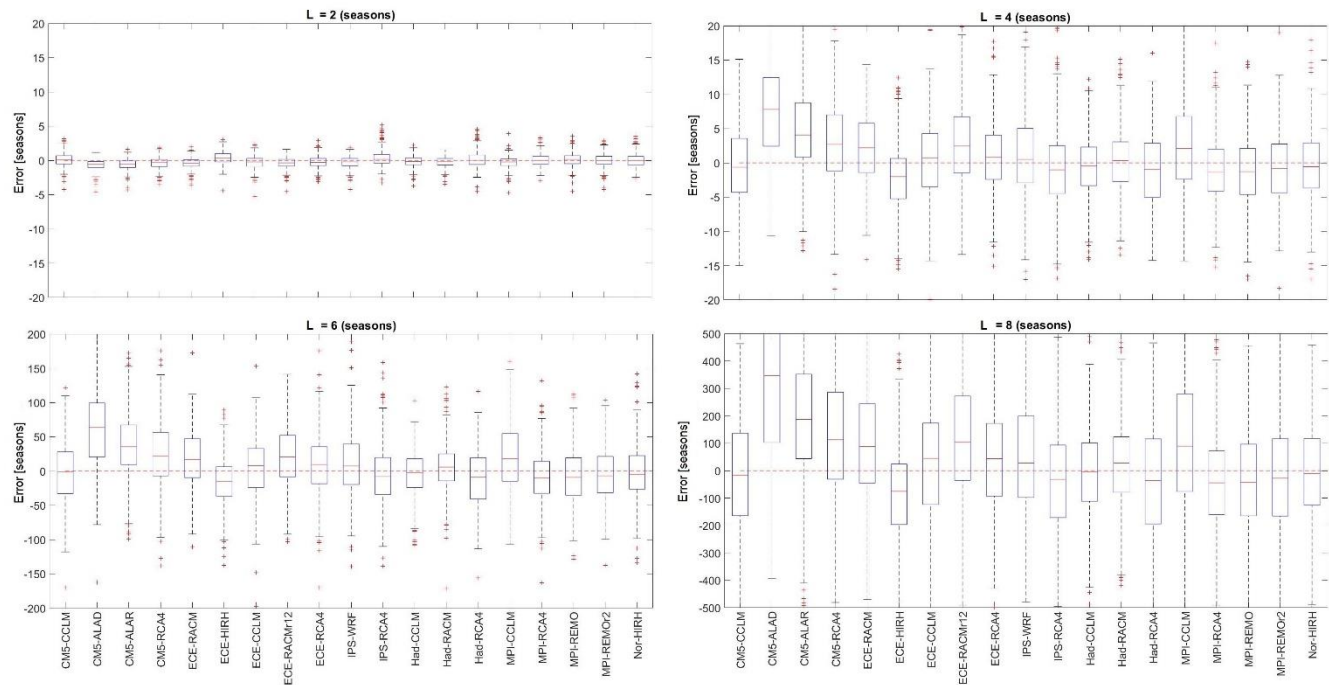


Figure 13. Box-plots representing the frequency distribution of RCMs percentage errors in the return period of drought event of duration L equal to 2, 4, 6 and 8 seasons. As Fig. 2 but maximum drought duration

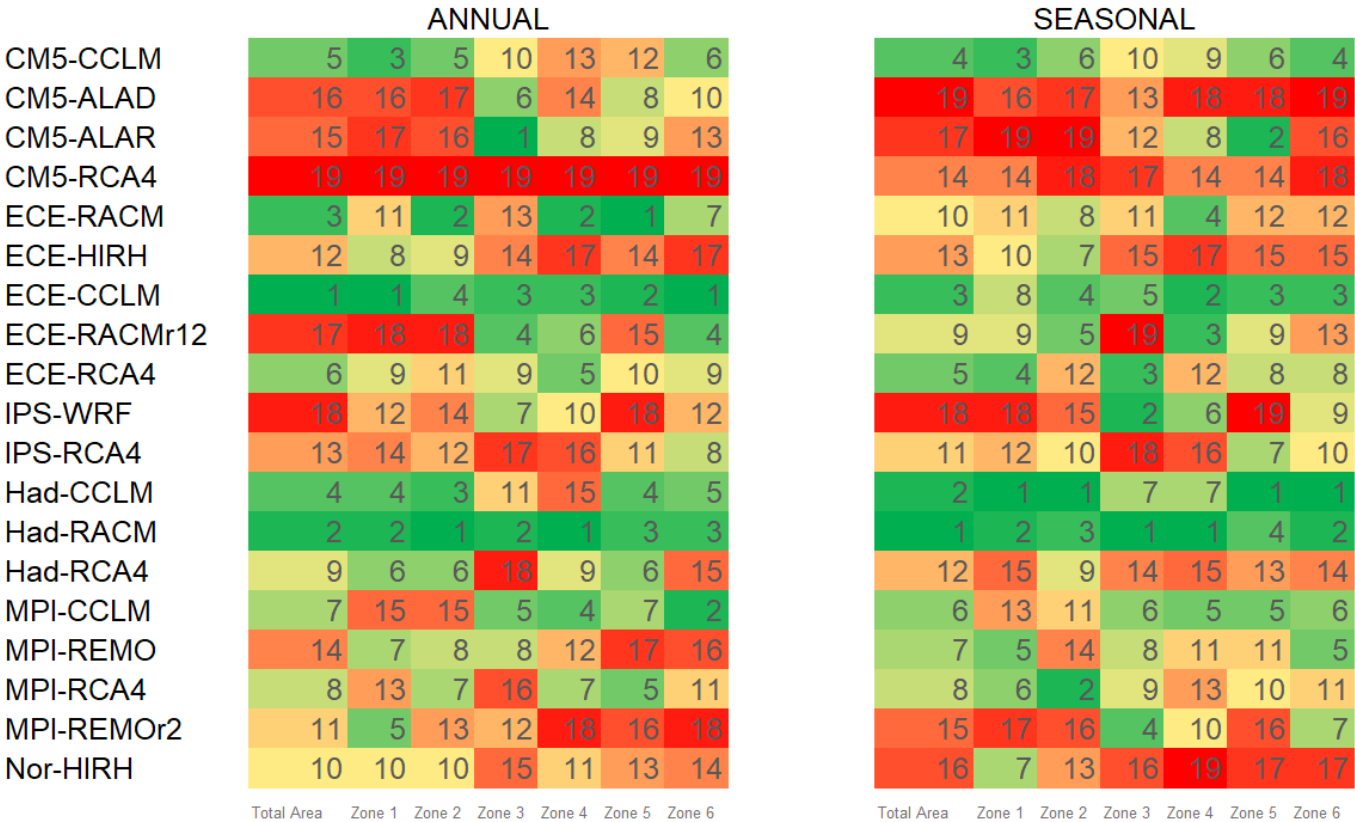
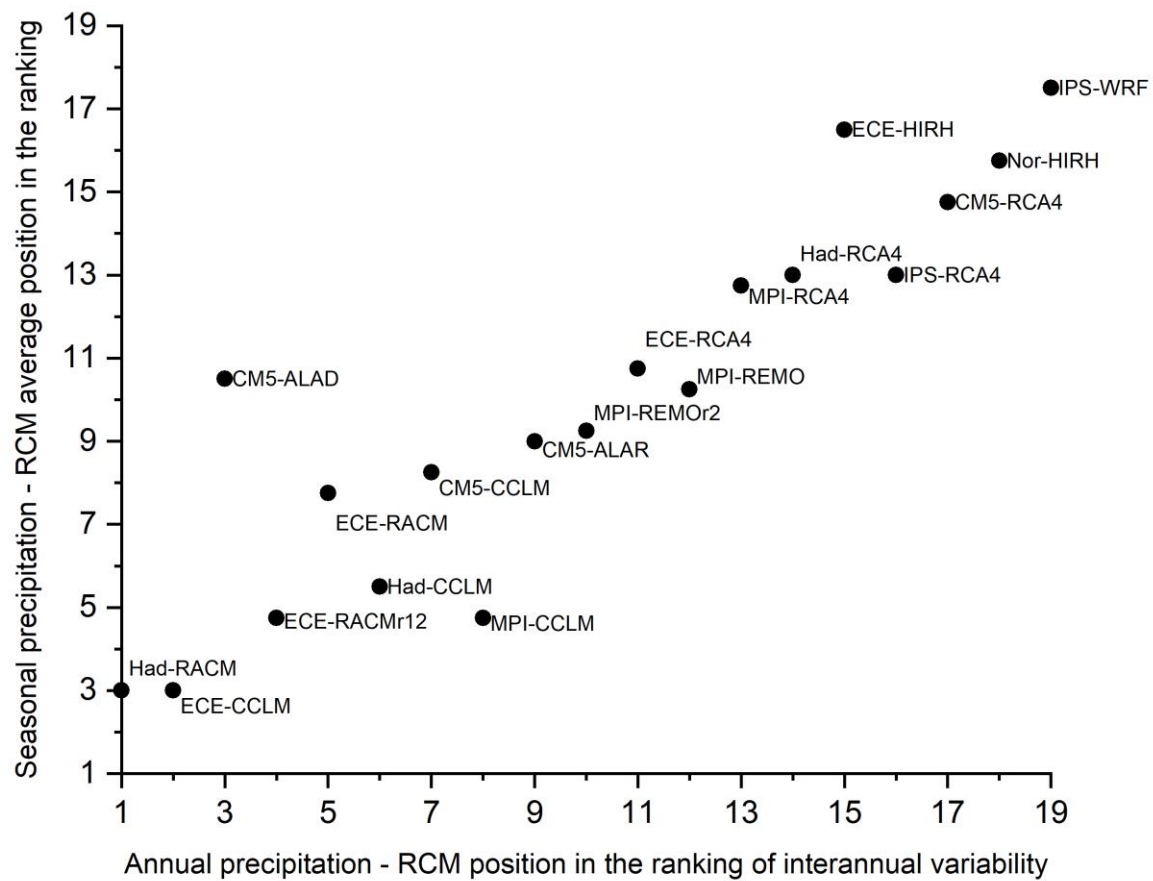


Figure 14. RCMs ranking with respect to their ability in reproducing both observed drought maximum intensities and return periods of drought events with duration L=3 years (left) and L=4 seasons (right). Box plots representing the frequency distribution of RCMs percentage errors in maximum drought accumulated deficit (right).



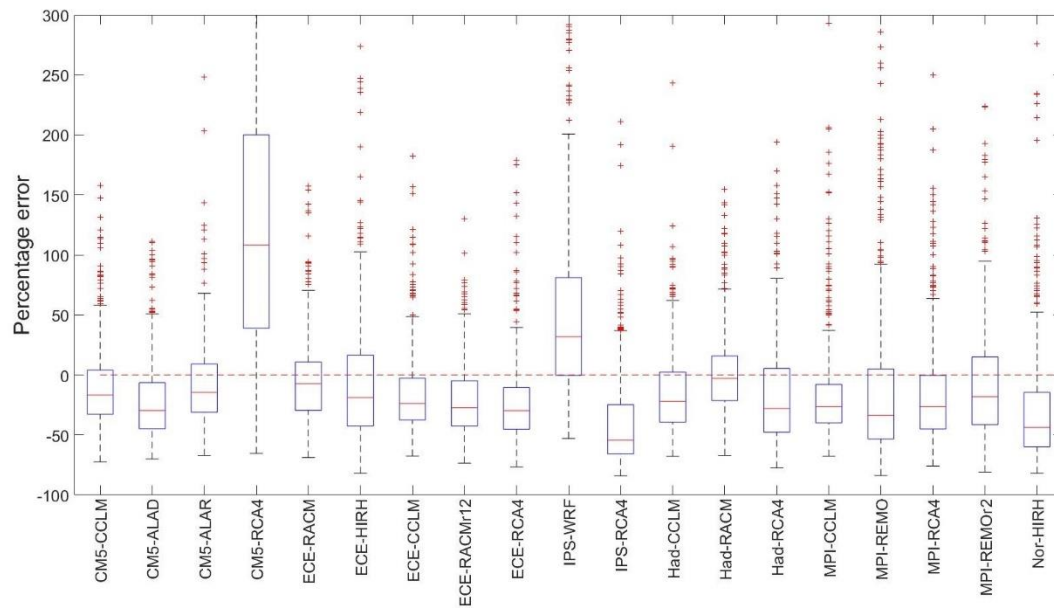


Figure 15. Comparison between the RCM position in the ranking of interannual variability of annual precipitation versus the average position in the ranking of seasonal variability of seasonal precipitation. Data concerns the whole study area (Calabria and Sicily). As Fig. 14 but for maximum drought intensity.

	INTERANNUAL						
CM5-CCLM	4	3	6	5	6	6	4
CM5-ALAD	8	13	13	2	9	3	7
CM5-ALAR	3	4	2	1	2	5	11
CM5-RCA4	19	19	19	19	19	19	19
ECE-RACM	2	5	3	10	4	1	3
ECE-HIRH	14	14	12	11	18	15	14
ECE-CCLM	5	2	9	7	7	4	8
ECE-RACMr12	7	12	14	3	5	2	1
ECE-RCA4	9	10	11	14	13	7	5
IPS-WRF	18	16	15	6	15	18	18
IPS-RCA4	15	18	17	18	16	9	13
Had-CCLM	6	7	4	9	3	8	2
Had-RACM	1	1	1	4	1	10	9
Had-RCA4	13	11	7	13	14	11	15
MPI-CCLM	10	9	5	8	8	16	6
MPI-REMO	17	15	16	16	11	17	17
MPI-RCA4	11	6	8	12	12	12	10
MPI-REMO _r 2	12	8	10	15	10	13	12
Nor-HIRH	16	17	18	17	17	14	16

Total Area Zone 1 Zone 2 Zone 3 Zone 4 Zone 5 Zone 6

Figure 16. Ranking of models in reproducing maximum drought intensity for the whole area and the six climatic zones

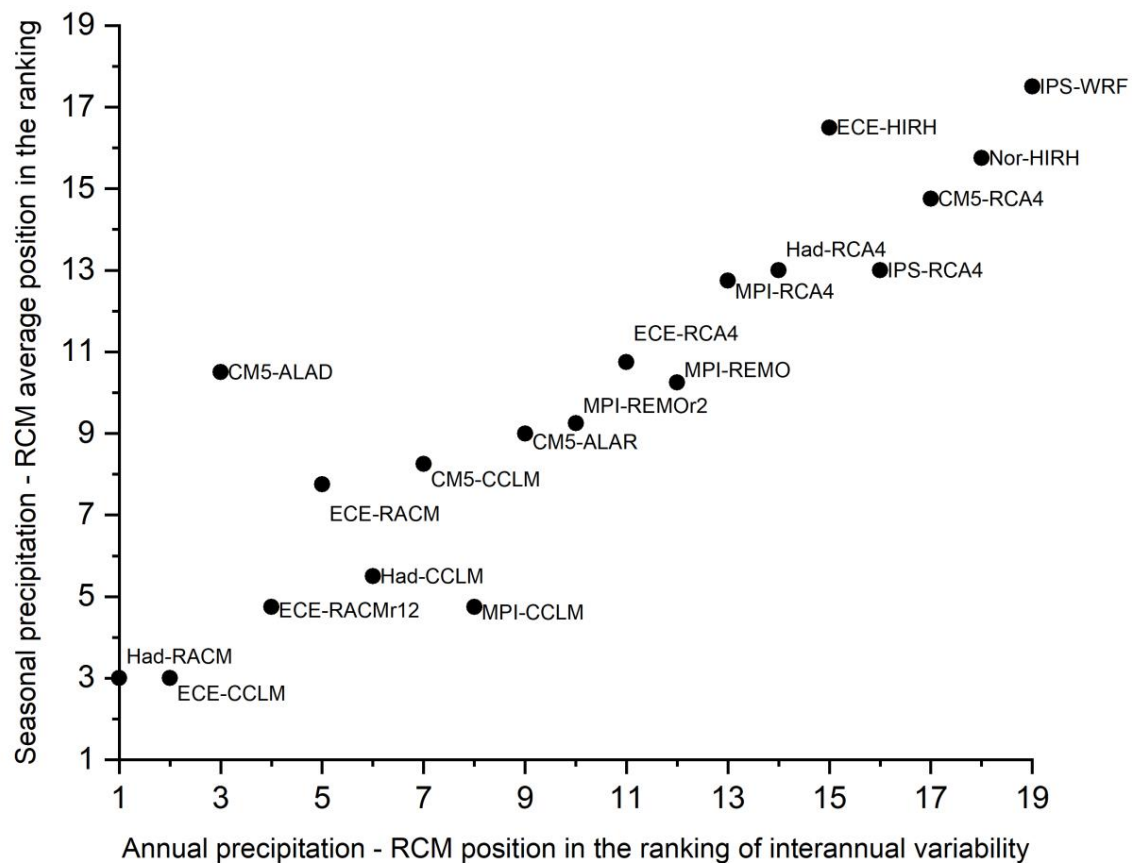


Figure xxx. Comparison between the RCM position in the ranking of interannual variability of annual precipitation versus the average position in the ranking of seasonal variability of seasonal precipitation. Data concerns the whole study area (Calabria and Sicily).

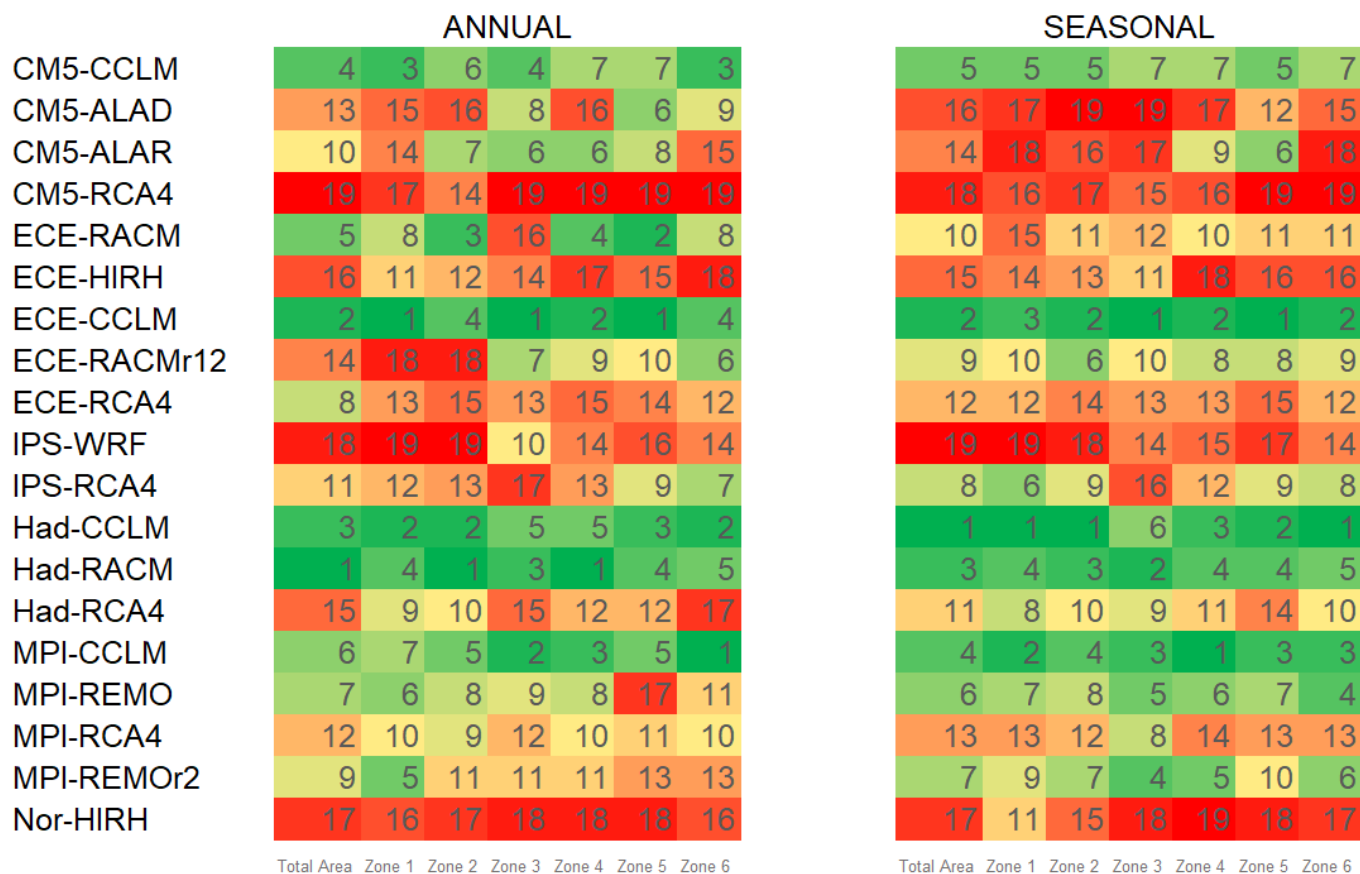


Figure 17. Overall Ranking.

995 **Table 1.** Intercomparison studies of RCMs' performances within the CORDEX framework

References	Models	Variables	Region	Main conclusions
Schmidli et al. (2007)	6 statistical downscaling models (SDMs) and 3 RCMs	Daily precipitation	European Alps	SDMs and RCMs tend to have similar biases but differ with respect to interannual variations, with SDMs strongly underestimate the magnitude of the year-to-year variations, mainly in winter. RCMs indicate a strong trend toward drier conditions including longer periods of drought. The SDMs, on the other hand, show mostly non-significant or even opposite changes. RCMs reasonably simulate the main features of the precipitation climatology. However significant biases are detected in individual models depending on sub-region and season. The ensemble mean has better agreement with observation than individual models. The analysis confirms the ability of RCMs to capture the basic features of the European climate. Seasonally and regionally averaged temperature biases are mostly smaller than 1.5 °C, while precipitation biases are typically located in the $\pm 40\%$ range. ARPEGE model shows the best simulation, while CRCM shows the worst.
Endris et al., (2013)	10 RCMs from CORDEX Africa domain	Seasonal and annual precipitation	Eastern Africa and 3 sub-regions	
Kotlarski et al. (2014)	9 EURO-CORDEX RCMs	Spatiotemporal patterns of the European climate	Europe	
Meque and Abiodun (2015)	10 RCMs from CORDEX Africa domain	Link between El Niño Southern Oscillation (ENSO) and Southern African droughts expressed by the Standardized Precipitation and Evapotranspiration Index (SPEI)	Southern Africa	Most RCMs overestimate (order of +10% to +400%, depending on model and subbasin) the mean annual difference between precipitation (P) and evaporation (E), All models can adequately reproduce the spatial distribution of extremely heavy precipitation. However, they do not perform well in simulating summer consecutive dry days. The ensemble average of multi-RCMs substantially improve model capability to simulate summer precipitation in both total and extreme categories when compared to each individual RCM. RCMs show systematic bias patterns in both seasonal means and extremes. The models simulate temperature means more accurately compared to extremes because of the higher spatial correlation, whereas precipitation extremes are simulated better than their means because of the higher spatial variability.
Mascaro et al. (2015)	6 RCMs driven by 10 GCMs from CORDEX Africa domain	Properties of the hydrological cycle	Niger River basin (West Africa)	
Wu et al. (2016)	4 RCMs from RMIP Project and their regional multi-model ensemble, and their driving GCM ECHAM5	Summer extreme precipitation	East Asia	
Park et al. (2016)	5 RCMs form the CORDEX East Asia domain	Climatology of summer extremes (seasonal maxima of daily mean temperature and precipitation)	East Asia	

Table 1. Continues

References	Models	Variables	Region	Main conclusions
Smiatek et al. (2016)	13 EURO-CORDEX RCMs	mean temperature and precipitation, frequency distribution of precipitation intensity, maximum number of consecutive dry days	Greater Alpine Region (GAR)	Though the models reproduce spatial seasonal precipitation patterns, the seasonal mean temperature is underestimated (from -0.8 °C to -1.9 °C) and mean precipitation is overestimated (from +14.8% in summer to +41.5% in winter). Larger errors are found for further statistics and various GAR sub-regions.
Diasso and Abiodun (2017)	10 RCMs from CORDEX Africa domain	Drought characteristics evaluated through 4 Principal Components of the SPEI	West Africa	Only two models (REMO and CNRM) reproduce all the four drought modes. REMO and WRF give the best simulation of the seasonal variation of the drought mode over the Sahel in March-May and June-August seasons, while CNRM gives the best simulation of seasonal variation in the drought pattern over the Savanna.
Um et al. (2017)	4 RCMs from CORDEX East Asia domain, their ensemble mean and a driving GCM	Drought characteristics based on the SPEI	East Asia	Drought severity diverges markedly among the RCMs. Estimates of drought spatial extent are generally accurate in wet regions but inaccurate in dry regions. In general, the spatial extents of the droughts diverge among the RCMs, and the models fail to accurately capture droughts with large spatial scales.
Foley and Kelman (2018)	7 EURO-CORDEX RCMs and 5 driving GCMs	Several precipitation indices (accumulated precipitation amount, mean daily precipitation amount, max 1-day and 5-day precipitation amounts, simple daily intensity, number of heavy and very heavy precipitation days)	Scottish islands	While no models perform skilfully across all the metrics studied, some models capture aspects of the precipitation climate at each location particularly well.
Adeniyi and Dilau (2018)	10 RCMs from CORDEX Africa domain	Precipitation, temperature and drought	West Africa	ARPEGE has the highest skill at Guinea Coast, while PRECIS is the most skilful over Savannah and RCA over the Sahel.
Senatore et al. (2019)	8 RCMs from CORDEX South Asia domain	Annual and seasonal precipitation and temperature	Iran and 6 sub-regions	No model is significantly better than others for every season and zone. Some enhancements are obtained by a weighting approach to take into account useful information from every RCM in the sub-zones. More reliable models show a strong precipitation decrease.
Di Virgilio et al. (2019)	6 RCMS from CORDEX Australasia domain	Near-surface max and min temperature and precipitation at annual, seasonal, and daily time scales	Australia	All RCMs showed widespread, statistically significant cold biases in maximum temperature and overestimated precipitation, especially over Australia's populous eastern seaboard.

Table 2. List of GCMs, together with the abbreviations used in this paper, included at least once in the EURO-CORDEX ensemble

Model name	Abbreviation	Reference	Institution
CNRM-CERFACS- CNRM-CM5	CM5	Voltaire et al. (2013)	Centre National de Recherches Météorologiques
ICHEC-EC-EARTH	ECE	Hazeleger et al. (2010)	Irish Centre for High-End Computing EC-Earth Consortium, Europe
IPSL-IPSL-CM5A-MR	IPS	Dufresne et al. (2013)	Institut Pierre Simon Laplace
MOHC-HadGEM2-ES	Had	Collins et al. (2011)	Met Office Hadley Centre
MPI-M-MPI-ESM-LR	MPI	Giorgetta et al. (2013)	Max-Planck-Institute für Meteorologie
NCC-NorESM1-M	Nor	Bentsen et al. (2013), Iversen et al. (2013)	Norwegian Earth System Model

1000

Table 3. List of RCMs, together with the abbreviations used in this paper, included at least once in the EURO-CORDEX ensemble

Model name	Abbreviation	Reference	Institution
CNRM-ALADIN53	ALAD	Colin et al. (2010)	Météo-France / Centre National de Recherches Météorologiques
RMIB-UGent-ALARO-0	ALAR	De Troch et al. (2013)	Royal Meteorological Institute of Belgium and Ghent University
CLMcom-CCLM4-8-17	CCLM	Baldauf et al. (2011), Rockel et al. (2008)	Climate Limited-area Modelling Community (CLM-Community)
		Baldauf et al. (2011), Rockel et al. (2008)	
DMI-HIRHAM5	HIRH	Christensen et al. (2007)	Danish Meteorological Institute
KNMI-RACMO22E	RACM	van Meijgaard et al. (2008)	Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
SMHI-RCA4	RCA4	Strandberg et al. (2014)	Swedish Meteorological and Hydrological Institute, Rossby Centre
MPI-CSC-REMO2009	REMO	Teichmann et al. (2013)	Helmholtz-Zentrum Geesthacht, Climate Service Center, Max Planck Institute for Meteorology
IPSL-INERIS-WRF331F	WRF3	-	Institut Pierre-Simon Laplace and French National Institute for Industrial Environment and Risks (Ineris)

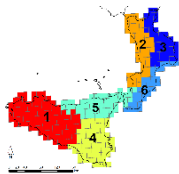
Table 4. List and acronyms of climate models (GCM-RCM combinations) included at least once in the EURO-CORDEX ensemble. The asterisk * means that two versions of the GCM-RCM combination are available

	CNRM- CERFACS- CNRM-CM5	ICHEC-EC- EARTH	IPSL-IPSL- CM5A-MR	MOHC- HadGEM2-ES	MPI-M-MPI- ESM-LR	NCC- NorESM1-M
CNRM- ALADIN53	CM5-ALAD	-	-	-	-	-
RMIB-UGent- ALARO-0	CM5-ALAR	-	-	-	-	-
CLMcom- CCLM4-8-17	CM5-CCLM	ECE-CCLM	-	Had-CCLM	MPI-CCLM	-
DMI- HIRHAM5	-	ECE-HIRH	-	-	-	Nor-HIRH
KNMI- RACMO22E	-	ECE-RACM*	-	Had-RACM	-	-
SMHI-RCA4	CM5-RCA4	ECE-RCA4	IPS-RCA4	Had-RCA4	MPI-RCA4	-
MPI-CSC- REMO2009	-	-	-	-	MPI-REMO*	-
IPSL-INERIS- WRF331F	-	-	IPS-WRF	-	-	-

1010 **Table 5.** Summary of the statistics involved in the ranking process. Statistics with subscript ₀ refer to observed values.

Property	Statistics k	Error $E_{k,m}(j)$
Seasonal variability	Seasonal mean	$ \mu_0(X_\tau(j)) - \mu_m(X_\tau(j)) $
	Seasonal standard deviation	$ \sigma_0(X_\tau(j)) - \sigma_m(X_\tau(j)) $
Interannual variability	Annual mean	$ \mu_0(X(j)) - \mu_m(X(j)) $
	Annual standard deviation	$ \sigma_0(X(j)) - \sigma_m(X(j)) $
Drought characteristics	Maximum drought duration	$ L_{max,0}(j) - L_{max,m}(j) $
	Maximum drought accumulated deficit	$ D_{max,0}(j) - D_{max,m}(j) $
	Maximum drought intensity	$ I_{max,0}(j) - I_{max,m}(j) $
	<u>Return period</u>	$ T_{r,0}(j) - T_{r,m}(j) $

Table 6. Best performing RCMs according to the ranking at the annual and seasonal scale.

		Whole area	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
T interannual variability		MPI-REMO	MPI-REMO	MPI-CCLM	IPS-RCA4	MPI-CCLM	MPI-CCLM	IPS-RCA4
		MPI-CCLM	Had-CCLM	MPI-REMO	MPI-CCLM	MPI-REMO	MPI-REMO	MPI-REMO
		Had-CCLM	MPI-CCLM	Had-CCLM	Had-CCLM	Had-CCLM	Had-CCLM	MPI-CCLM
T seasonal variability	DJF	ECE-HIRH	CM5-CCLM	CM5-CCLM	ECE-HIRH	ECE-HIRH	MPI-CCLM	MPI-CCLM
	MAM	ECE-CCLM	ECE-CCLM	MPI-REMO _{r2}	ECE-CCLM	MPI-REMO _{r2}	ECE-CCLM	ECE-CCLM
	JJA	IPS-RCA4	IPS-RCA4	IPS-RCA4	Had-RCA4	IPS-RCA4	MPI-REMO _{r2}	MPI-REMO _{r2}
	SON	MPI-REMO _{r2}	MPI-REMO _{r2}	MPI-REMO _{r2}	Had-CCLM	MPI-REMO _{r2}	MPI-REMO _{r12}	MPI-CCLM
P interannual variability		Had-RACM	Had-RACM	ECE-RACM	ECE-CCLM	ECE-CCLM	ECE-CCLM	CM5-ALAD
		ECE-CCLM	CM5-CCLM	Had-RACM	Had-CCLM	Had-RACM	CM5-ALAD	Had-RACM
		CM5-ALAD	CM5-ALAD	CM5-ALAR	Had-RACM	Had-CCLM	Had-RACM	ECE-RACM _{r12}
P seasonal variability	DJF	ECE-RACM _{r12}	MPI-CCLM	ECE-RACM _{r12}	ECE-RACM	ECE-RACM _{r12}	ECE-RACM _{r12}	ECE-RACM
	MAM	ECE-CCLM	ECE-CCLM	Had-RACM	CM5-CCLM	MPI-CCLM	ECE-CCLM	ECE-CCLM
	JJA	MPI-REMO _{r2}	MPI-REMO _{r2}	MPI-REMO _{r2}	MPI-REMO _{r2}	MPI-REMO _{r2}	Had-CCLM	ECE-CCLM
	SON	MPI-CCLM	MPI-CCLM	ECE-RACM _{r12}	IPS-WRF	MPI-CCLM	Had-RACM	Had-RACM
Drought-intensity _I + <u>T_y(L=3 years)</u> (annual scale)		<u>ECE-CCLM</u> <u>ECE-CCLM</u>	<u>ECE-CCLM</u> <u>Had-CCLM</u>	<u>Had-RACM</u> <u>CM5-ALAR</u>	<u>CM5-ALAR</u> <u>ECE-CCLM</u>	<u>Had-RACM</u> <u>CM5-ALAR</u>	<u>ECE-RACM</u> <u>ECE-CCLM</u>	<u>ECE-CCLM</u> <u>Had-CCLM</u>
		<u>Had-RACM</u> <u>Had-CCLM</u>	<u>Had-RACM</u> <u>ECE-RACM_{r12}</u>	<u>ECE-RACM</u> <u>Had-RACM</u>	<u>Had-RACM</u> <u>MPI-CCLM</u>	<u>ECE-RACM</u> <u>ECE-CCLM</u>	<u>ECE-CCLM</u> <u>ECE-RACM_{r12}</u>	<u>MPI-CCLM</u> <u>CM5-CCLM</u>
		<u>ECE-RACM</u> <u>Had-RACM</u>	<u>CM5-CCLM</u> <u>ECE-CCLM</u>	<u>Had-CCLM</u> <u>ECE-CCLM</u>	<u>ECE-CCLM</u> <u>Had-CCLM</u>	<u>ECE-CCLM</u> <u>MPI-REMO</u>	<u>Had-RACM</u> <u>MPI-REMO</u>	<u>Had-RACM</u> <u>Had-RACM</u>
<u>I+T_y(L=4 seasons)</u> (seasonal scale)		<u>Had-RACM</u>	<u>Had-CCLM</u>	<u>Had-CCLM</u>	<u>Had-RACM</u>	<u>Had-RACM</u>	<u>Had-CCLM</u>	<u>Had-CCLM</u>
		<u>Had-CCLM</u>	<u>Had-RACM</u>	<u>MPI-RCA4</u>	<u>IPS-WRF</u>	<u>ECE-CCLM</u>	<u>CM5-ALAR</u>	<u>Had-RACM</u>
		<u>ECE-CCLM</u>	<u>CM5-CCLM</u>	<u>Had-RACM</u>	<u>ECE-RCA4</u>	<u>ECE-RACM_{r12}</u>	<u>ECE-CCLM</u>	<u>ECE-CCLM</u>