Received: 2 March 2022 | Revised: 4 December 2022 | Accepted: 10 January 2023

DOI: 10.1002/qj.4428

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

SPHERA, a new convection-permitting regional reanalysis over Italy: Improving the description of heavy rainfall

Antonio Giordani[1,2] | **Ines Maria Luisa Cerenzia**[2] | **Tiziana Paccagnella**[2] | **Silvana Di Sabatino**[1]

[1]Department of Physics and Astronomy (DIFA) "Augusto Righi", University of Bologna, Bologna, Italy

[2]ARPAE Emilia Romagna, Bologna, Italy

Correspondence
Antonio Giordani, Department of Physics and Astronomy, University of Bologna, viale Berti Pichat 6/2, Bologna, Italy
Email: antonio.giordani3@unibo.it

Funding information
OPERANDUM (OPEn-air laboRAtories for Nature baseD solUtions) funded by the European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 776848; Universita degli Studi di Bologna within the CRUI-CARE Agreement; TRIGGER (SoluTions foR mItiGatinG climate-induced hEalth thReats) funded by the European Union, Grant/Award Number: 101057739

Abstract

Regional reanalyses allow us to better describe weather patterns related to rapidly evolving high-impact events thanks to substantially finer detailing than global datasets. However, most regional datasets still do not permit the explicit representation of deep convection. SPHERA (High rEsolution ReAnalysis over Italy) is a new high-resolution convection-permitting reanalysis centred over Italy. It covers 26 years (1995–2020), is based on the non-hydrostatic limited-area model COSMO, and is produced by dynamically downscaling the global reanalysis ERA5. A nudging data assimilation scheme steers the model toward observations. The fine horizontal grid spacing of 2.2 km allows us to switch off deep-convection parametrization. This study reports the added value of SPHERA over ERA5 in representing rainfall over Italy, particularly for severe precipitation, using rain-gauge observations during 2003–2017 as reference. Concerning the 95th percentile of spatial rainfall distributions, ERA5 presents dry estimates with biases reaching $-12\,\mathrm{mm\cdot day^{-1}}$ over mountainous regions. At the same time, the enhanced locally driven effects of SPHERA produce seasonal biases ranging from wet in JJA (up to $+12\,\mathrm{mm\cdot day^{-1}}$) to dry in DJF (down to $-9\,\mathrm{mm\cdot day^{-1}}$). For daily maximum rates, the regional reanalysis shows better skill in detecting occurred events (with hit rates higher than ERA5 by roughly 0.4 points in the range of 15–80 $\mathrm{mm\cdot day^{-1}}$) and frequency biases closer to 0 at all intensities when coming to daily averages. Similarly, for hourly maximum accumulations, improved adherence to observations is detected for SPHERA at all intensities, conversely to the underprediction of the global driver (with frequency biases <1 starting from 1.5 $\mathrm{mm\cdot hr^{-1}}$). Additionally, the analyses of two specific events reveal the enhancements of SPHERA in simulating extreme precipitation, with a maximum intensity underestimation on the order of 24% versus the 73% detected for ERA5. Further improvements include the spatial detailing, timing, and temporal evolution of the events.

KEYWORDS

convection-permitting, ERA5, fuzzy verification, regional reanalysis, severe precipitation, SPHERA

1 | INTRODUCTION

Deep moist convective processes are characterized by extremely chaotic and nonlinear behaviour, growing from small- to meso-scale, making their simulation particularly challenging. The investigation of past meteorological conditions leading to convection and the associated events is crucial to enhance their understanding and extend this knowledge to the present and future states of the atmosphere. The most common way to investigate past meteorological states in atmospheric science is through reanalysis datasets. Atmospheric reanalyses respond to the necessity of homogeneous spatio-temporal meteorological data by combining numerical weather predictions (NWPs) and observations. In these datasets, data assimilation is used to ingest observations and physically constrain model data. Atmospheric reanalyses extend from several years to a few decades. Among global reanalyses, ERA5 (the fifth-generation reanalysis: Hersbach *et al.*, 2020) produced at the European Centre for Medium-Range Weather Forecasts (ECMWF) is widely considered state-of-the-art. Anyhow, global datasets, including ERA5, are still characterized by coarse horizontal grid spacings spanning from 125 to 31 km and with temporal frequencies varying from 6 to 1 hr. Finer spatial grids are essential to describe small-scale and rapidly evolving features, as for convective processes. For instance, the simulation of deep convection needs parametrization schemes to be accounted for in large-scale models. This constitutes a major source of errors and inaccuracies in the simulations (Prein *et al.*, 2015).

The necessity of a superior level of precision led to the development of higher-resolution limited-area reanalyses. These are usually obtained by downscaling a global dataset to produce a finer displacement of the data, hence having a higher level of detail. In dynamical downscaling, the boundary conditions of the global driving model force a limited-area system to produce atmospheric fields at the desired grid refinement (e.g. Castro *et al.*, 2005). However, the downscaling process could generate relevant small-scale variabilities (constrained to the larger-scale state but not to the observations), which can be detrimental to the finer-scale estimates (Giorgi, 1990; Simon *et al.*, 2013; Desamsetti *et al.*, 2019). The assimilation of regional observations is used to minimize coarse-scale driven variabilities and improve the quality of high-resolution estimates. Regional reanalyses are produced by coupling a downscaling strategy with a data assimilation method, which has proven to be superior to a mere downscaling (i.e. hindcast) of a global reanalysis (e.g. Bollmeyer *et al.*, 2015; Jermey and Renshaw, 2016). Several regional reanalyses have been produced in recent years on the continental or national scale. Some examples are: China Regional Reanalysis project (CNRR: Zhang *et al.*, 2017) having 18 km grid spacing, ASRv2 (Arctic System Reanalysis version 2: Bromwich *et al.*, 2018) with 15 km grid spacing, Bureau of Meteorology Atmospheric high-resolution Regional Reanalyses for Australia (BARRA, New Zealand, and Southeast Asia: Su *et al.*, 2019), or CARRA, the Arctic Regional Reanalysis produced by the Copernicus Climate Change Service (C3S) at a very fine grid spacing of 2.5 km (Køltzow *et al.*, 2022). In Europe, the interest has been particularly high, as demonstrated by the numerous datasets recently developed: European Reanalysis and Observations for Monitoring (EURO4M: Klein Tank, 2010) and its continuation Uncertainties in Ensembles of Regional Reanalyses (UERRA: Unden *et al.*, 2016), the Consortium for Small-scale MOdelling regional reanalysis (COSMO-REA6: Bollmeyer *et al.*, 2015) covering part of the European CORDEX domain (Coordinated Regional Downscaling EXperiment, cordex.org) with a grid spacing of 6 km, the new ensemble-based Copernicus European Regional ReAnalysis (CERRA: El-Said *et al.*, 2021) covering Europe and North Africa at 11 km horizontal grid spacing (while 5.5 km is used for its deterministic counterpart), or MEteorological Reanalysis Italian DAtaset (MERIDA: Bonanno *et al.*, 2019) covering Italy with a 7 km grid. The increased detail of these datasets provides added value, especially in the representation of precipitation (Bollmeyer *et al.*, 2015; Jermey and Renshaw, 2016) and in the case of complex orography (Isotta *et al.*, 2015). However, these grid spacings are still too coarse to permit the explicit representation of deep moist convection, which is paramount for substantially improving the simulations.

In fact, with the explicit representation of deep convection, achieved through the so-called convection-permitting (CP) scales, an improved description of precipitation is obtained in terms of different aspects: enhanced representation of the summer diurnal cycle of precipitation (Fosser *et al.*, 2015; Brisson *et al.*, 2016), better accordance with the observed intensities of the most severe precipitation events (Prein *et al.*, 2013; Fosser *et al.*, 2015), lower biases of average and extreme rainfalls (Pal *et al.*, 2019), and smaller frequency biases of weak precipitation events (Berg *et al.*, 2013). Furthermore, CP models successfully represent organized convective structures and the formation of isolated convective cells and self-regenerating thunderstorms (Clark *et al.*, 2016), as well as orographically driven convection thanks also to the enhanced detailing of the topography and surface heterogeneities (Weusthoff *et al.*, 2010; Kirshbaum *et al.*, 2018). Similarly, CP regional reanalyses better describe the spatial variability on the local scale of precipitation and the frequency of heavy rainfall (Wahl *et al.*, 2017). Anyhow,

CP simulations require km-scale grid spacings (i.e. below 4 km), whose computational cost is still too high to cover extended/global domains (e.g. Schär *et al.*, 2020). For this reason, CP reanalyses are limited to small spatial domains on the order of the national and below the continental scale. Recently produced CP reanalyses over Europe include: COSMO-REA2 (nested in COSMO-REA6: Wahl *et al.*, 2017) covering Central Europe at 2 km grid spacing, Met Éirann ReAnalysis (MERA: Gleeson *et al.*, 2017) over Ireland and the United Kingdom with a 2.5 km grid, and the in-development high-resolution regional reanalysis for Iberian Peninsula and Balearic Islands (IBERA: Calvo Sánchez *et al.*, 2021) developed by Agencia Estatal de METeorología (AEMET) at 2.5 km grid spacing. When coming to Italy, the considerable interest in developing highly resolved re-forecast datasets is demonstrated by the recent production of two CP regional hindcasts. These are obtained by downscaling ERA5 using the COSMO model at 2.2 km grid spacing (Raffa *et al.*, 2021; Reder *et al.*, 2022) or the model MOLOCH at 2.5 km (Capecchi *et al.*, 2022). Anyhow, in neither case is the additional assimilation of regional observations included in the production of the datasets.

This prompted the development of a new CP regional reanalysis, SPHERA (High rEsolution ReAnalysis over Italy), covering Italy, the surrounding seas, and part of the neighbouring states (Figure 1). The details of SPHERA production and the necessary preliminary experimentations to define its optimal set-up have been investigated by Cerenzia *et al.* (2022). SPHERA is nested on the global reanalysis ERA5 and based on the COSMO model, run deterministically with a $0.02° \times 0.02°$ grid spacing (i.e. approximately 2.2×2.2 km), that is coupled with a nudging data assimilation scheme ingesting regional *in situ* observations. This study presents the first extensive validation of the new dataset in terms of precipitation, with a focus on heavy rainfall occurrences, on a daily and sub-daily basis, as well as the comparison with the parent reanalysis ERA5 to quantify the added benefits of SPHERA.

The article is organized as follows: Section 2 reports the characteristics of SPHERA, the observational data employed, and the validation strategy applied; Section 3 describes the preliminary sensitivity analysis for the upscaling needed for the verification; Section 4 reports the results of the performance for precipitation and the comparison with ERA5 on a daily and hourly basis and with particular focus towards extreme precipitation events; Section 5 analyses two particular case-studies of recent severe precipitation over Italy; and the discussions and conclusions are drawn in Section 6.

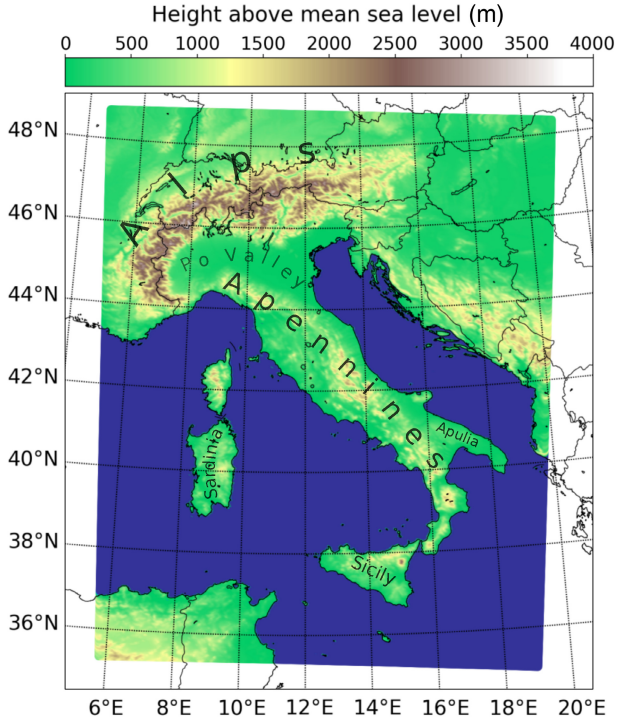

FIGURE 1 The spatial domain and model orography of SPHERA reanalysis [Colour figure can be viewed at wileyonlinelibrary.com]

2 | DATA AND VALIDATION STRATEGY

This section presents the new reanalysis SPHERA and its driver ERA5 (Section 2.1), the observation dataset used for their validation (Section 2.2), and the strategy to assess the performance of the two reanalyses (Section 2.3).

2.1 | The reanalysis dataset SPHERA

SPHERA (High rEsolution ReAnalysis over Italy) is the new high-resolution regional reanalysis dataset produced at ARPAE-SIMC (the hydro-meteo-climate service of Emilia Romagna region, Italy), covering Italy, part of the neighbouring countries and the surrounding seas (Figure 1) at the CP horizontal grid spacing of $0.02° \times 0.02°$ (approximately 2.2×2.2 km). Table 1 summarizes SPHERA main features, including the list of parametrizations employed to account for sub-grid physical processes. The outputs are produced with hourly temporal frequency on 65 vertical levels (0–22 km above sea level (a.s.l.)) and seven soil levels (0–14.58 m). SPHERA spans 26 years (from 1995 to 2020), and further extension of the dataset in the future will be possible. SPHERA is based on the non-hydrostatic limited-area model COSMO (Baldauf *et al.*, 2011; Schättler *et al.*, 2018),

TABLE 1 Main technical characteristics of SPHERA

Initial conditions	ERA5
Boundary conditions	ERA5, updated every hour
Nesting modality	One-way nested in ERA5
Sea-surface temperature	Interpolated from ERA5 every day
Deep soil temperature	Parametrized from ERA5 soil temperature after Cerenzia *et al.* (2022)
Assimilated observations	SYNOP and SHIP (not temperature at 2 m nor precipitation for either), TEMP, PILOT and AIREP
Code version	INT2LM 2.04 (pre-processing), COSMO 5.05 in double precision
Spatial domain	Approximately 35°N, 5°E; 49°N, 20°E (estimated from the rotated domain, Figure 1)
Spatial resolution	0.02° (~2.2 km) horizontal (576 × 701 grid cells), 65 vertical levels terrain-following (0–22 km), seven soil levels (0–14.58 m)
Temporal frequency	1 hr
Temporal coverage	1995–2020
Radiation scheme	δ two-stream scheme after Ritter and Geleyn (1992)
Turbulence scheme	Prognostic turbulent kinetic energy closure at level 2.5 after Raschendorfer (2001)
Land-surface scheme	Multi-layer soil after Jacobsen and Heise (1982)
Transfer scheme	Surface layer scheme coupled with the turbulence scheme (Doms *et al.*, 2018)
Convection scheme	Only shallow convection with reduced Tiedtke (1989)
Microphysics scheme	Grid-scale cloud and precipitation scheme (three-categories ice scheme) and statistical scheme for sub-grid clouds after Sommeria and Deardorff (1977)
Subgrid-scale orography scheme	Lott and Miller (1997)
Lake scheme	Two-layer bulk model after Mironov (2008)
External parameter: Orography	Global land 1 km Base elevation project (GLOBE task team *et al.*, 1999)
External parameter: Land cover	Global Landcover 2000 Database (Mayaux *et al.*, 2006)
External parameter: Soil type	Digital soil map of the world (UNESCO/FAO: the United Nations' Educational, Scientific and Cultural Organization and the Food and Agriculture Organization)

developed by the European Consortium for Small-Scale Modelling. COSMO is an NWP model used operatively in several European and extra-European countries belonging to the eponymous consortium (i.e. Italy, Marsigli *et al.*, 2005; Germany, Baldauf *et al.*, 2011; Romania, Dumitrache *et al.*, 2011; Russia, Rivin *et al.*, 2015; Poland, Starosta and Wyszogrodzki, 2016; Greece, Avgoustoglou *et al.*, 2018; Israel, Hochman *et al.*, 2018; Switzerland, Klasa *et al.*, 2018). COSMO employs the primitive non-hydrostatic thermo-hydrodynamical equations for a compressible flow in a moist atmosphere. The three-dimensional equations are defined in a rotated geographical coordinate system and include various physical processes through parametrization schemes. A critical process not parametrized in the configuration used in SPHERA is deep moist convection, thanks to the fine grid spacing allowing its explicit representation (or at least for a large part of convective motions). The reader is referred to Cerenzia *et al.* (2022) for a detailed description of the modelling framework adopted. SPHERA is initialized through dynamically downscaling the fifth-generation global reanalysis ERA5 produced at ECMWF (Hersbach *et al.*, 2020). ERA5 is based on the Integrated Forecasting System (IFS) Cy41r2, producing deterministic hourly three-dimensional output at the spectral resolution TL639 (i.e. approximately 31 km grid spacing at midlatitudes). As investigated in Cerenzia *et al.* (2022), the nesting modality to downscale the lateral boundary condition of ERA5 to SPHERA, updated every hour, is chosen as a one-step nesting. This has proven more skilful than nesting passing through an intermediate grid. For computational reasons, the production of SPHERA has been performed through six 4-year streams, each preceded by 6 months of rerun to account for the soil spin-up. The streams are formed by COSMO 24 hr long runs, each of which provides the initial conditions for the subsequent one. The runs are then all stitched together to create a continuous hourly series. The data assimilation technique implemented in COSMO is a continuous nudging of *in situ* observations, similar to COSMO-REA6 (Bollmeyer *et al.*, 2015). The continuous

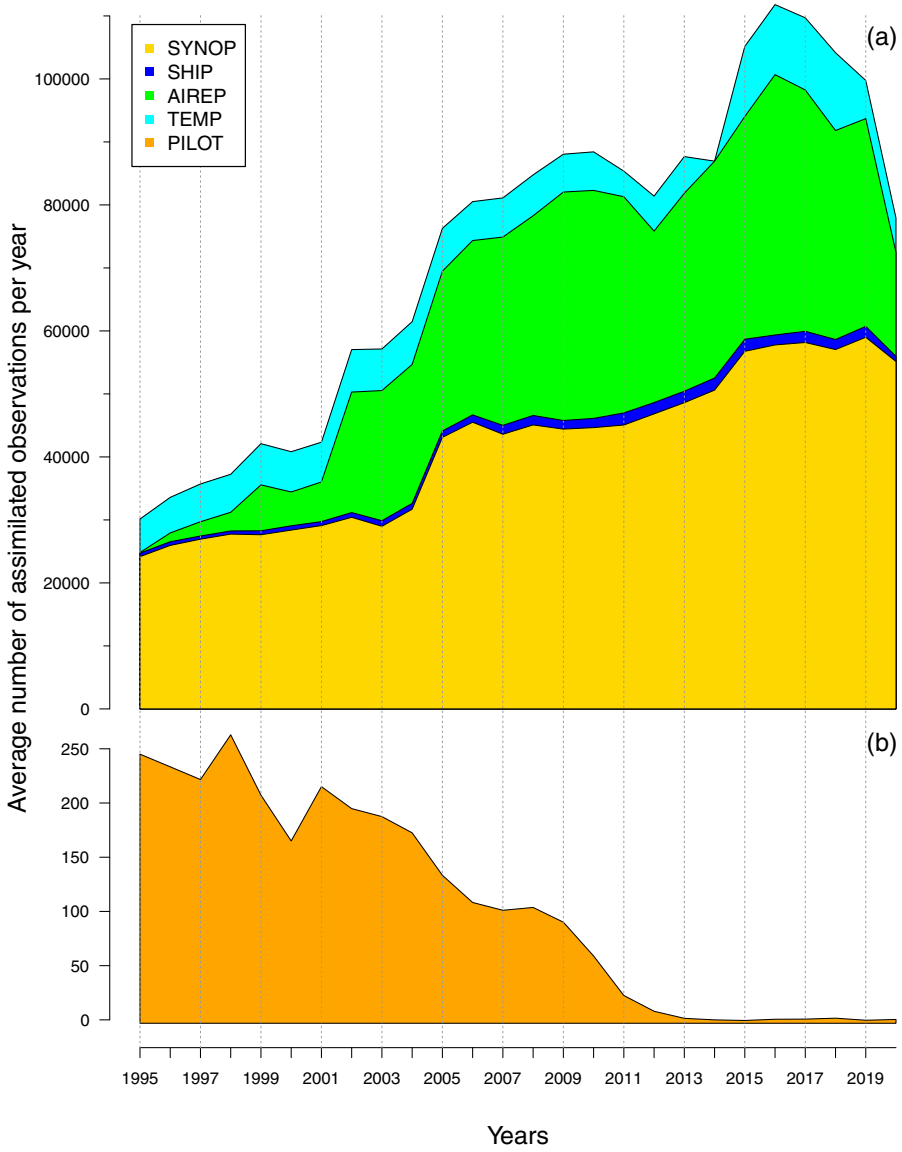

FIGURE 2 Average number of assimilated observations in SPHERA per year from 1995 to 2020. The assimilated observational datasets are reported in different colours. (a) SYNOP (yellow), SHIP (dark blue), AIREP (green), and TEMP (light blue). (b) PILOT (orange), separated from the rest of the assimilated data given its lower numbers [Colour figure can be viewed at wileyonlinelibrary.com]

nudging is based on the Newtonian relaxation principle and aims to dynamically adjust the model towards the prescribed observations within a predetermined time window. This is done by inserting an additional term (so-called nudging term) in the prognostic model equations, which is proportional to the spatio-temporal misfit between the observations and the model, and that continuously adapts the simulation towards the observed values during the forward integration of the model. The nudging term always remains smaller than the largest term of the dynamics in the model equations in order to relax the model fields towards the observations without significantly disturbing their dynamic balance. The reader is referred to Schraff and Hess (2013) for a detailed description of the nudging method implementation in COSMO. The set of observational data nudged in SPHERA comes from the ECMWF catalogue. It comprises near-surface observations over land and sea (SYNOP and SHIP), radiosounding and radar profiler data (TEMP and PILOT), and aircraft reports (AIREP). The meteorological variables assimilated include wind speed components, pressure, air humidity, and temperature, except for temperature at 2 m. All the assimilated observations are operationally quality-checked before the assimilation process and thus accepted or rejected following the standard procedure implemented in COSMO (details are reported in Schraff and Hess, 2013).

Figure 2 reports the average number of accepted observations per year for each assimilated dataset. Notably, their number increased almost constantly during the 26 years of SPHERA coverage, from approximately 30,000 in 1995 up to roughly 100,000 in 2019. As an exception, in 2020 the number of accepted records is remarkably reduced for SYNOP, SHIP, TEMP, and especially AIREP observation types (Figure 2a). This decrease is due to the lower number of available observations (and not to an increased data rejection) which have possibly originated from the globally

reduced observational activity during the first year of COVID-19 pandemic spreading (Riishojgaard, 2020). With respect to PILOT records (Figure 2b), a consistent decrease in the amount of data with time is noted, approaching 0 after 2012. Also in this case, this is not a consequence of an increased rejection ratio in the assimilation process, and is marginal compared to the much larger number of the other types of assimilated data in SPHERA considering the whole 26-year period.

2.2 | The observational dataset

The performance of SPHERA and ERA5 is assessed through a comparison with rainfall rain-gauge observations. These pertain to the national composite of meteorological stations available on Dewetra, the Italian Civil Protection database of pluviometers (Italian Civil Protection Department, CIMA Research Foundation, 2014), containing the entirety of Italian regional networks and covering the period 2003–2017. These are entirely independent of both ERA5 and SPHERA as they are assimilated in neither reanalysis. Dewetra contains a set of ground pluviometers supplying data at hourly temporal frequency. The number of rain-gauges has significantly increased from approximately 1,500 in 2003 to 6,200 in 2017. Pluviometric data have been quality-checked operatively by the department of civil protection. Additional controls have been performed for the most extreme rainfalls recorded (i.e. daily accumulations exceeding 500 mm). Anyhow, it is worth pointing out that rain-gauge data are always affected by limitations, like the spatial inhomogeneity due to the insufficient density of pluviometers or the sensors' inadequacy to properly detect snowfall precipitation. Over the Italian territory, the former is a relevant issue in southern regions (where the situation has improved significantly in recent years), and in mountainous terrain (Crespi *et al.*, 2018). The second issue pertains more to the northern and central regions during winter, particularly over mountains. Nevertheless, given the comparative nature of this analysis, we believe these limitations are not a major issue for assessing the added benefits of one reanalysis over the other when investigating their long-term statistical performance. On the other hand, for the case-studies reported in Section 5, the spatial inhomogeneity of Dewetra's pluviometers may substantially affect the results. For this reason, only for these specific analyses, we used two additional sources of rainfall observations to compare with the simulations, namely: the ARchivio Climatologico per l'Italia centro-Settentrionale (ARCIS, i.e. the high-resolution gridded precipitation analysis of pluviometric data covering north-central Italy at roughly 5 km spacing: Pavan *et al.*, 2019) and surface rainfall intensity (SRI) estimates combining rain-gauge with radar data.

2.3 | Validation strategy

A fuzzy verification approach (Ebert, 2008) is used to compare precipitation performances of SPHERA and ERA5. With a fuzzy method, the matching conditions between the model and observations are relaxed, and the forecast is required to be in approximate agreement with the observation (by being close in space, time, or other quantities). This is opposed to traditional nearest-point verifications, seeking exact matches between forecast/observation pairs. Fuzzy methods are preferred when considering high-resolution models for which matching the observations with absolute precision is too difficult, hence risking "double-penalty" issues (i.e. to correctly forecast a situation but being offset from the observation). Indeed, small displacements in this context doubly penalize the simulation by producing a false alarm and missing the observation. Additionally, given the sparse nature of rain-gauge observations, a spatial aggregation is necessary to objectively assess the quality of the model simulation (Bollmeyer *et al.*, 2015) and to compensate for representativeness limitations that may affect individual point observations (e.g. Weusthoff *et al.*, 2010). These issues are particularly relevant in the representation of highly localized phenomena, such as severe convective precipitation, characterized by low spatio-temporal predictability. Their description needs high-resolution simulations, but the outcomes would be strongly penalized if traditional point-to-point comparisons were used (Lanciani *et al.*, 2008). With fuzzy methods, a spatial neighbourhood to assess the closeness between forecasts and observations is defined around a certain point of interest. This is achieved by systematically upscaling (or boxing) forecasts and observations to a common coarser grid (compared to the respective original ones). The verification domain is subdivided into boxes of the same size, each containing a certain number of reanalysis and observation points (e.g. Weygandt *et al.*, 2004).

The upscaled distributions are then aggregated to provide a single forecast–observation pair for each box. The comparison between the aggregated data pairs is carried out considering both the mean and maximum (or 95th percentile) of their distributions within each box of the upscaled domain (similar to Marsigli *et al.*, 2008). This is needed to evaluate different characteristics of the precipitation distributions. The comparison between the mean values can indicate the average skill of a reanalysis in representing precipitation. In contrast, comparing

the maxima (or the 95th percentiles) can assess the ability to describe intense and localized rainfalls since the most extreme events (or the tail of extremes, respectively) are considered. Particularly, from the latter analysis, a high-resolution system is expected to reveal its added value over its coarser counterpart (e.g. Klasa *et al.*, 2018). These distribution parameters have different impacts on identifying a useful forecast (i.e. a positive match between reanalysis and observations within a box). The matching between maxima is more generous than the averages: for the former case, a useful forecast implies its maximum in the neighbourhood to be close to the observed counterpart, that is, the match relies only on one value of each distribution; for the averages, the closeness criterion is more restrictive since their values are calculated from the whole distributions found in the neighbourhood (Ebert, 2008). For this reason, comparing the performances obtained through different statistical aggregations is also useful to understand whether possible deviations from the observed states are due to the method used for aggregating the meteorological fields. An observational mask is applied to retain only the grid cells where appropriate closeness between reanalysis–observations is met (similarly to Marsigli *et al.*, 2008). This is done to minimize the errors owing to the heterogeneous pluviometers distribution, and the larger spatial domain covered by the reanalysis. For any dimension of the upscaled grid hereafter considered (see Section 3), boxes containing less than five rain-gauges are withdrawn from the analysis to avoid errors caused by insufficient sampling of the observed distribution. This number has been defined for the operational verification of numerical forecasts at ARPAE as the minimum required to allow a representation of precipitation suitable for this work. The upscaled grid neighbourhood (or box) size defines the allowed spatial displacement between forecast and observations. The most appropriate definition of this horizontal scale is not trivial. It depends on various factors (e.g. the particular features investigated through the simulations, the time resolution, the meteorological situation, and the model itself: Gallus, 2002; Ebert, 2008). Therefore, a preliminary sensitivity analysis on the neighbourhood size is performed to investigate how reanalysis skill scores vary with grid size and to find the optimal grid spacing for the upscaling (described in Section 3).

The performance assessment of SPHERA and ERA5 is carried out with a categorical approach. A 2×2 contingency table is obtained by classifying precipitation occurrences in each box as yes/no events based on the exceedance of a rainfall threshold and aggregating the entire spatial domain (see Appendix A). With the contingency table, a complete representation of the joint distribution between the binary observations and the binary outcome of the simulations (i.e. event observed/forecasted or not observed/not forecasted) is gained. From them, a series of dichotomous scores is calculated for each threshold considered, namely: the probability of detection (POD), the false alarm ratio (FAR), which can be expressed as the success ratio (SR = 1 − FAR), the threat score (TS) and the frequency bias. TS measures the correspondence between observed and simulated events when correct negatives are removed from consideration (Wilks, 2019). TS, POD and FAR vary from 0 to 1, while the frequency bias ranges from 0 to infinity, with a perfect simulation obtained when POD, TS and frequency bias are 1, and FAR is 0. The skill scores are calculated for the average and maximum values of the daily precipitation distributions, which are both needed to get a more complete assessment of the performance. Conversely, since the main interest of the present article is investigating precipitation extremes for rapidly evolving cases, only maxima are considered for the hourly precipitation analysis. The resulting skill scores are aggregated over the entire verification domain and different temporal periods and reported through the performance diagram. This visual tool exploits the geometrical relationships between the four dichotomous scores considered (Roebber, 2009). A summary of the indices used is reported in the Appendix A.

3 | SENSITIVITY ANALYSIS FOR THE UPSCALING

This section is devoted to the sensitivity analysis of the neighbourhood size to find the optimal upscaling configuration performed on a sample period of 3 years (2015–2017). The maxima of SPHERA and Dewetra daily precipitation distributions are compared within boxes of different sizes before the verification extension to the entire 15-year period. The box dimensions cover a broad range from $0.14°\times0.14°$ to $1.85°\times1.85°$, corresponding approximately to grids from 15×15 km to 200×200 km. Each of these grid spacings allows an adequate sampling of the number of grid points of SPHERA contained in the box, ranging from approximately 46 grid points (for the finest box of 15 km) to roughly 8,264 grid points (for the coarsest box of 200 km). Figure 3 reports the dependency of TS, POD, FAR and frequency bias scores to the upscaled grid-spacing and on the daily precipitation intensity (with thresholds ranging from 1 to 150 mm·day^{-1}). The results pertain to seasonal averages over summer (June–August, JJA) and winter (December–February, DJF).

Two distinct effects contribute to a joint deterioration of the scores: increase of precipitation intensity and refinement of the upscaled grid spacing. The decrease of POD (Figure 3c,d) and TS (a,b) and the increase of FAR (e,f) and frequency bias (g,h) for growing rainfall thresholds reflects the rising difficulty in detecting occurred

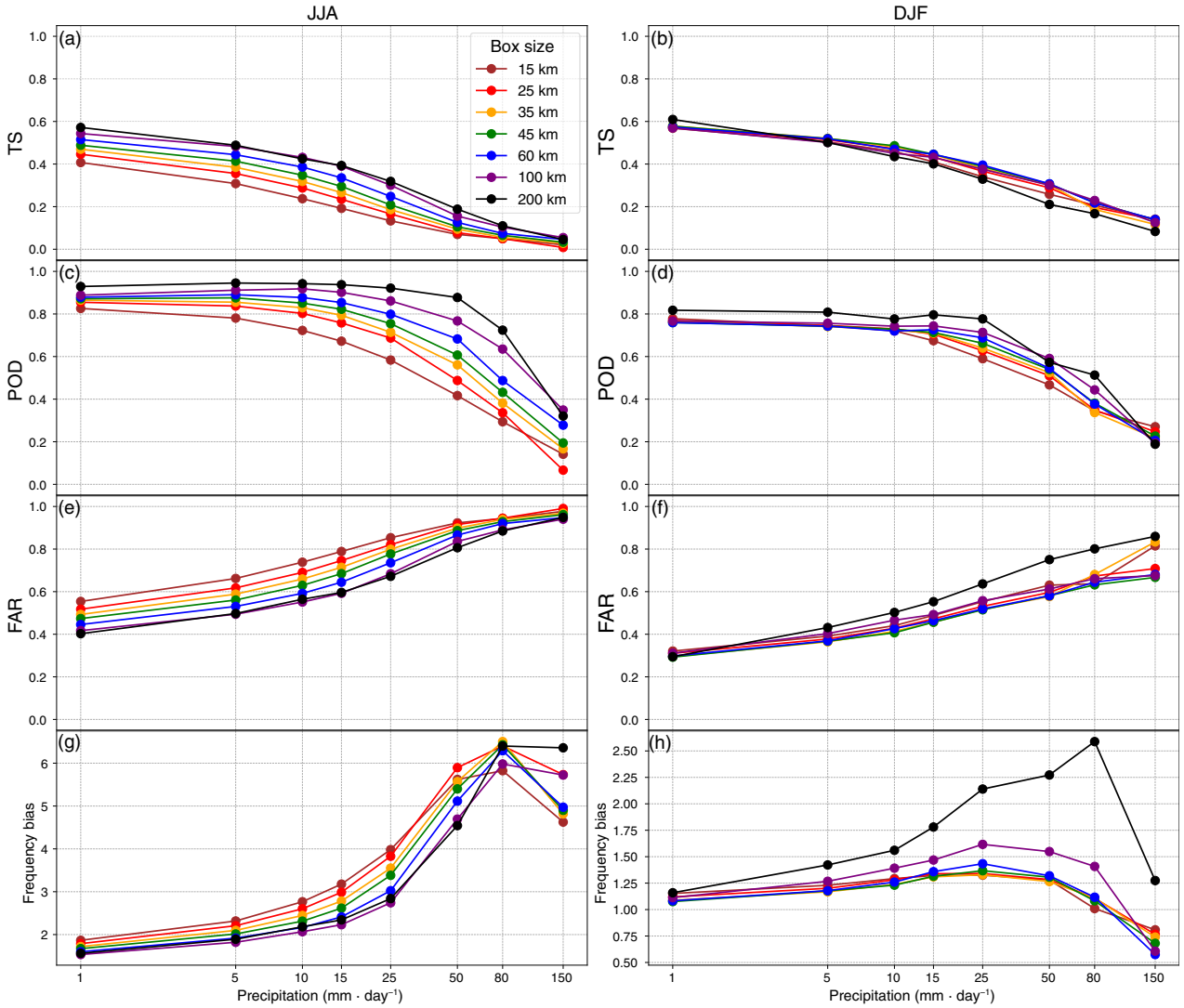

FIGURE 3 Seasonal averages over the period 2015–2017 of dichotomous scores obtained from the maximum values of the upscaled rainfall distributions within boxes of different sizes (indicated with different colours). (a,b) Threat score (TS), (c,d) probability of detection (POD), (e,f) false alarm ratio (FAR) and (g,h) frequency bias (note the different scale on the *y*-axis), as functions of daily precipitation threshold, (a,c,e,g) for the summer season (JJA) and (b,d,f,h) for the winter season (DJF) [Colour figure can be viewed at wileyonlinelibrary.com]

events. As precipitation intensifies, the number of observations correctly simulated reduces, more false detections are produced and the most extreme and rare events are overestimated. This positive frequency bias is reasonably dependent on the inadequate spatial sampling of the sensor networks to detect precipitation extremes. Indeed, if the mean daily precipitation is considered (Figure 8, right panel), the overestimation disappears from rainfalls $\geq$20 mm·day^{-1}. This observation limitation also plausibly explains the large number of false alarms detected, with FAR ranging from 0.4 to 1.0 in JJA and 0.3 to 0.8 in DJF.

The performance degradation owing to the upscaled grid spacing is expected due to the rising impact of short-lived and localized events with low predictability. These are responsible for the majority of spatio-temporal double-penalty errors, particularly in JJA (Marsigli *et al.*, 2008). If the grid spacing is enlarged, the increased rain-gauge quantity per box and the removal of numerous small-scale occurrences affected by phase errors produce a better performance (Weygandt *et al.*, 2004). However, at the same time, the simulations lose sharpness due to the removal of actually occurred heavy rainfall events as well (Roberts and Lean, 2008). The performance improvement for coarser box sizes is enhanced in JJA when convective rainfall is most likely. Figure 3 shows a larger spread among different upscalings for all skill scores in JJA than in DJF. POD in JJA shows the strongest dependence on the grid size especially for intense rainfall accumulations ($\geq$25 mm·day^{-1}, Figure 3c). For example, in the case of 50 mm·day^{-1}, POD ranges from approximately 0.4 (15 km box) to 0.9 (200 km box), while in DJF, the spread never exceeds 0.2 points. This effect is most

likely imputable to different precipitation predictability during the two seasons: the detection of localized summer convective rainfalls is more challenging than the widespread frontal-like winter counterparts (e.g. Crespi *et al.*, 2018). This seasonal dependency is also reflected by the systematic lower risk of committing false alarms in DJF, as expressed by the lower FAR scores (Figure 3f) obtained for each threshold and grid size.

Finally, there is a further effect influencing the performance of the simulations and connected to the upscaling neighbourhood size, which is linked to the typical spatial length of precipitation. This effect causes the opposite behaviour in the performance on a seasonal basis, particularly visible for coarser boxes in terms of FAR and frequency bias (black and purple lines in Figure 3). In DJF the worst performance is obtained with the 200 km grid box, while in JJA with the 25 km one. Indeed, even if the risk of missing events is reduced with coarser grids due to the more generous matching condition, the likelihood of mixing different unrelated events falling within the same box is higher. This is particularly detrimental for the 200 km box case, likely because this spatial resolution may exceed the characteristic length-scale of precipitation: while in JJA, typical convective rainfall is associated with horizontal scales on the order of a few kilometres, in DJF, the horizontal extension of stratiform-like precipitation may vary from tens to a few hundreds of kilometres (Houze, 2014). This length-scale issue does not seem to affect summer precipitation, most likely due to the higher impact of double penalties in localizing convective showers and the under-sampling of the observed state in JJA, which mask this secondary effect.

The deviation between the normalized rainfall distributions of SPHERA and Dewetra as a function of rainfall intensity for different box sizes is also analysed (Figure 4). The aim is to identify the spatial dimension that minimizes the difference between the two distributions, with particular emphasis on severe precipitation, to let the reanalysis maintain its resolution, that is, provide predictions able to correctly distinguish situations with distinctly different frequencies of occurrence (Murphy, 1993; Lewis *et al.*, 2015). For weak to moderate precipitation (up to 50–55 mm·day^{-1}), the grid size best minimizing the difference is the coarsest one (200 km). In comparison, the others worsen almost linearly by reducing the box size. A reversal trend is mostly evident for intense rainfalls exceeding 60 mm·day^{-1}. Indeed, the coarsest horizontal grid spacings (200 and 100 km) are always associated with the highest deterioration of the similarity between SPHERA and Dewetra. This excessive upscaling causes the removal of extreme rainfall occurrences, which are associated with small-scale processes, inevitably smoothed out at these grid spacings. Moving to finer upscalings, the

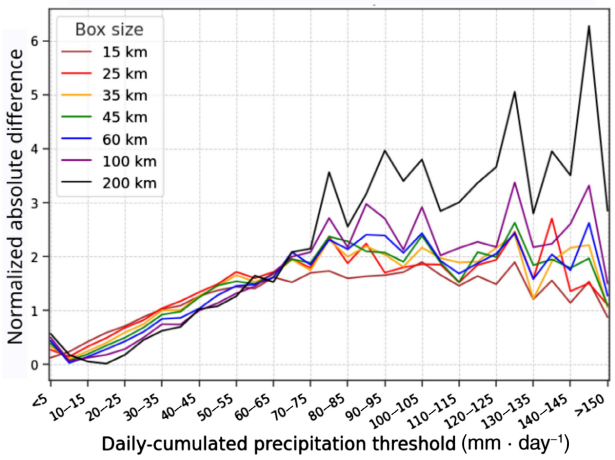

FIGURE 4 Normalized absolute difference between the number of precipitation events per daily threshold relative to the rainfall distributions of SPHERA and Dewetra observations for the period 2015–2017, for different horizontal resolutions of the upscaled domain (in different colours) [Colour figure can be viewed at wileyonlinelibrary.com]

differences in this rainfall range are not prominent, except for the finest 15 km box always producing the most effective minimization. However, this grid spacing is excluded from the present analysis (as well as the 25 km box), as one of the main goals of this article is to perform a long-term comparison between SPHERA and its driver ERA5. ERA5 has a native grid spacing of 31 km. Therefore, upscaling to a 15 km grid (or 25 km grid), would result in a downscaling and it would cause an under-population of the boxes containing less than one native ERA5 point. Considering the remaining box sizes (i.e. 35, 45 and 60 km), no decisive improvements of one over the other emerge at any intensity regime (Figures 3 and 4). However, boxes of 60 km contain enough grid points of ERA5 for its consistent representation (approximately four per box). For this reason, the 60 km grid spacing is chosen to perform the reanalysis evaluation in the next section. It should be noted that for any other applications non-constrained by a comparative analysis, finer grid boxes should be considered (e.g. 15–25 km) for the investigation of severe precipitation events.

4 | EVALUATION OF PRECIPITATION

This section presents the performance verification of SPHERA and ERA5 rainfall estimates against Dewetra pluviometric data over 15 years (2003–2017). Section 4.1 compares daily spatially distributed rainfall fields when focusing on severe precipitation events. In Section 4.2, daily accumulated precipitation is the subject of the

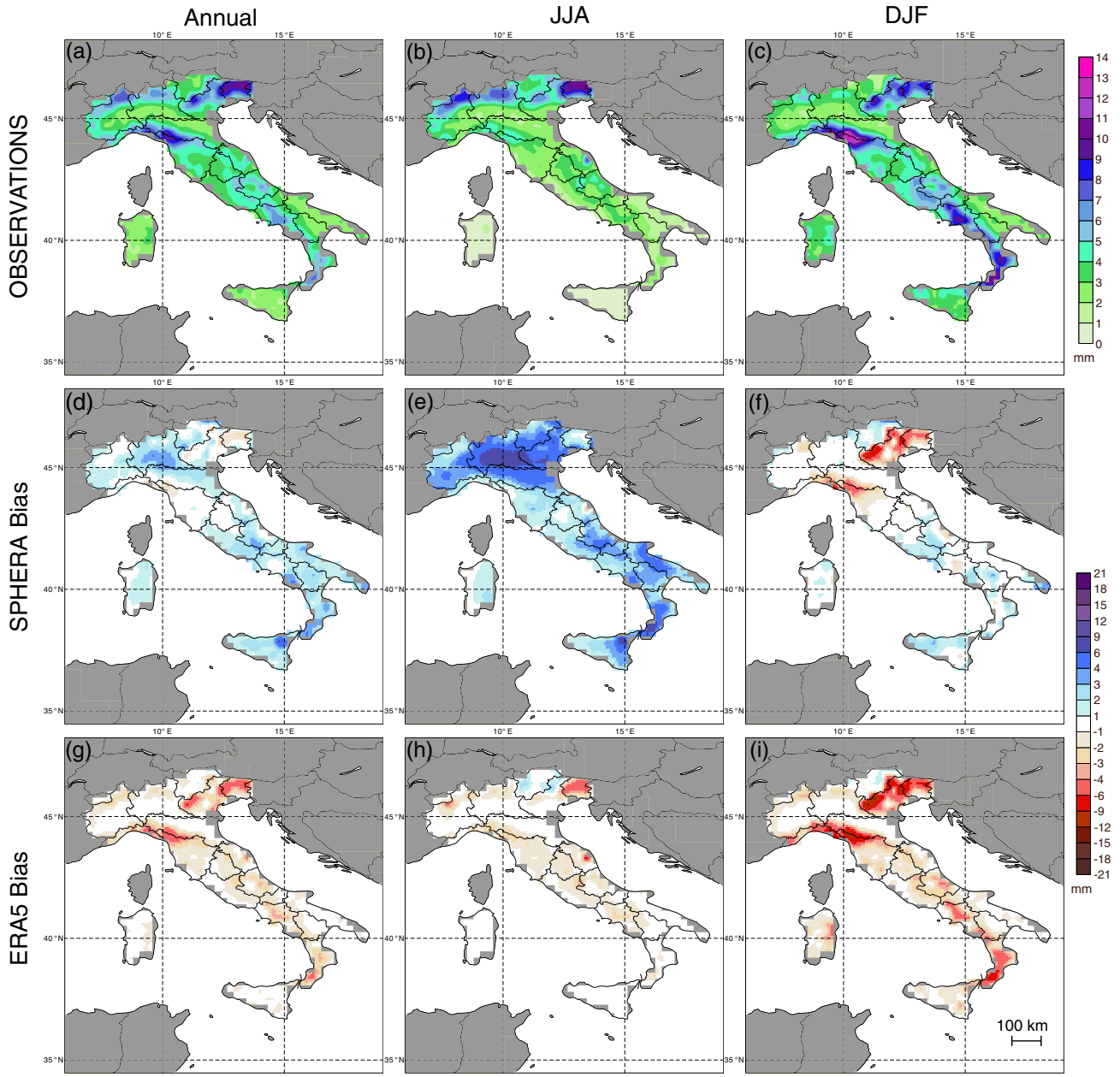

FIGURE 5 (a–c) Observed average of the 95th percentile of daily precipitation distributions (mm) for different temporal aggregations over the years 2003–2017: (a,d,g) annual, (b,e,h) summer (JJA) and (c,f,i) winter (DJF). Daily deviations (mm) from the observed spatial distributions for (d–f) SPHERA and (g–i) ERA5. All three datasets are upscaled over a common grid of approximately 31 km horizontal grid spacing [Colour figure can be viewed at wileyonlinelibrary.com]

fuzzy verification performed with the optimal upscaling resolution of 60 km. This analysis extends to hourly accumulations in Section 4.3. Finally, the mean diurnal precipitation cycle during the summer season is analysed in Section 4.4.

4.1 | Spatial distribution

A fundamental feature to investigate is the reanalysis ability to represent the "correct climate" in terms of spatial distribution of precipitation. The following reports the analysis of the spatially distributed daily observed state aggregated over 2003–2017, on the annual and seasonal terms (JJA and DJF), and the relative deviations of the reanalyses. In this context, the 95th percentile of the boxed distributions is chosen for the spatial intercomparison in order to focus on intense precipitation occurrences. Only in this section, the precipitation fields are upscaled over a common grid of 31 km spacing, roughly corresponding to the ERA5 native grid. Indeed, the loss of detail resulting from the coarser optimal grid of 60 km (defined in

Section 3 when aggregating the precipitation fields over the whole domain) would be too much for this spatial analysis.

Figure 5a–c report the observed 95th percentiles of daily rainfall distribution. Italy's wettest regions are visible from the annual plot: the northern Apennines and the far-eastern Alps, with daily accumulations exceeding 10 mm·day^{-1} (the reader is referred to Figure 1 for geographical references). The western Alps and the south-central Apennines are other areas particularly prone to precipitation, with daily rainfalls of 5–8 mm·day^{-1}. These precipitation patterns are mainly linked to the orographic enhancement of precipitation characterizing the Italian region (Napoli *et al.*, 2019), of which an emblematic case-study is reported in Section 5.1. The drier regions, presenting daily accumulations below 3 mm·day^{-1}, are found over the plains and hilly areas, especially over the Po valley, the Sardinia and Sicily islands, and the Apulian peninsula. The observed patterns are consistent with several recent rainfall climatologies obtained with higher-resolution datasets (Longobardi *et al.*, 2016; Crespi *et al.*, 2018; Pavan *et al.*, 2019) and reflect the impact of the very complex Italian topography on the spatial behaviour of precipitation. Therefore, the upscaling procedure does not affect the observed precipitation patterns estimated with Dewetra.

Considering reanalysis estimates on the annual term, ERA5 (Figure 5g) overall presents a heterogeneous dry bias, mostly non-systematic but peaked over the wettest regions (i.e. primarily mountainous areas) and dampened over the plains, with an average relative bias per box of −1.1 mm·day^{-1}. SPHERA also shows a heterogeneous distribution of the annual bias (Figure 5d), but mainly linked with an overestimation of rainfall intensity, particularly over the Po valley and in southern Italy, with an average bias per box of 1.4 mm·day^{-1}.

The distribution of the observed 95th percentiles over JJA and DJF (Figure 5b,c) reveals the seasonality of precipitation: summer rainfall is mainly enhanced over the Alps, particularly over their far-eastern region, presenting daily accumulations above 10 mm·day^{-1}; winter precipitation is characterized by even more orographic enhancement: rainfall peaks extend across the entire Apennines with daily accumulations larger than 13 and 10 mm·day^{-1} at their northern and southern ends, respectively. These patterns are due to the dominant mesoscale humid flows impinging over mountainous ranges typical of the cold season (e.g. Krichak *et al.*, 2015).

Moving to reanalysis estimates, in JJA ERA5 (Figure 5h) shows a similar bias distribution to that of the annual term with a mean bias per box of −1.7 mm·day^{-1}. SPHERA (panel e) presents a higher wet bias, marked over the Po valley, southern Apennines and eastern Sicily, with a daily average bias per box of 2.9 mm·day^{-1}. During wintertime, ERA5 (panel i) notably underestimates precipitation, especially over the cited mountainous wet spots, with an average bias of −1.7 mm·day^{-1} per box. Also, SPHERA in DJF (panel f) shows a dry bias over part of the northern Italy orography (with an average bias per box of 0.1 mm·day^{-1}). This indicates a difficulty in correctly representing dynamically driven orographic precipitation in specific regions even for a CP model. However, weaker and more spatially limited dry biases are obtained compared to ERA5. Furthermore, ERA5 always underestimates precipitation if the long-term average of the 95th percentile distributions is considered, suggesting a systematic inability to simulate severe rainfall events. This is typical of low-resolution models that employ convection-parametrizing schemes and lack a sufficient level of detail in the representation of topography, which is crucial especially over complex terrain. On the other hand, the substantial overestimation committed by SPHERA in summer should be interpreted as something other than a systematic wet bias of the driving model. Suppose the observed state is not sampled as well as the reanalysis (as potentially could be the case for Dewetra as highlighted in Sections 2.2 and 3, especially in mountainous regions). In that case, an overestimation of the high-resolution rainfall simulations is likely when considering their extreme values of the distribution, such as the 95th percentile. This hypothesized under-sampling of the observed state is supported by the analysis of the distribution for average daily rainfall intensities in the next section: a systematic overprediction of SPHERA for intense rainfall rates is not found in that case. These results confirm and extend the preliminary equivalent analysis over 2015–2016 reported in Cerenzia *et al.* (2022).

4.2 | Daily precipitation

The daily accumulated precipitation of the two reanalyses is assessed through performance diagrams (Figure 6). The performance diagram exploits the geometrical relationship of four dichotomous scores: the probability of detection POD (on the *y*-axis), the false-alarm ratio FAR (expressed on the *x*-axis as the SR = 1 − FAR), the TS (indicated in the diagrams with grey shading contouring), and the frequency bias (shown by dashed diagonal lines, see Appendix A for further details). In the diagrams, the results of an accurate model would lie on the bisector line describing a null frequency bias, and a perfect simulation would lie on the top-right corner of the diagram (i.e. POD, SR and TS = 1).

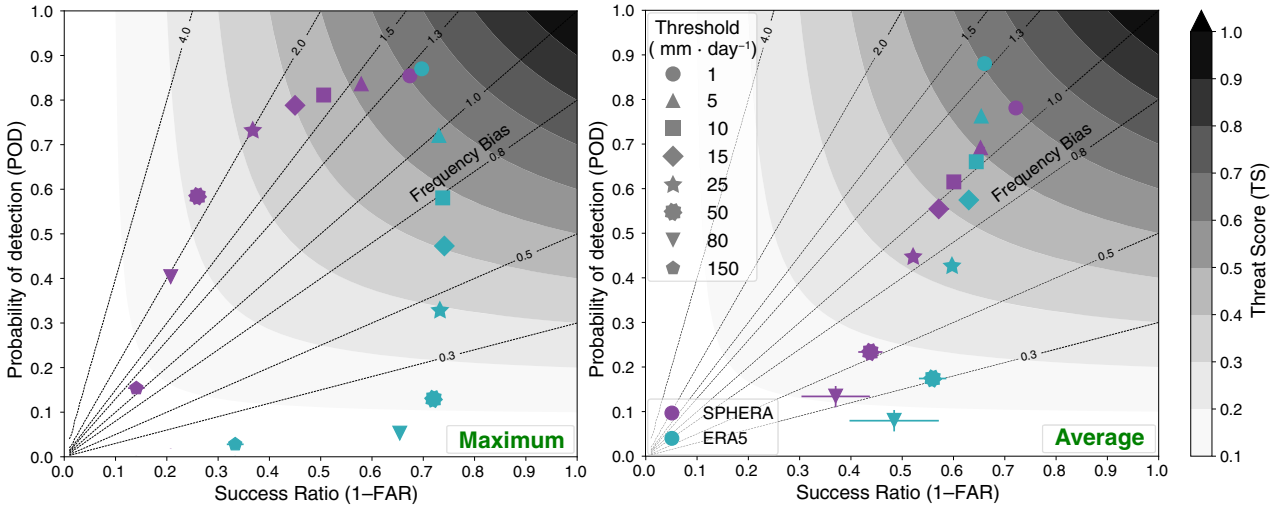

FIGURE 6 Performance diagrams for the aggregation of the scores over 2003 to 2017 for SPHERA (purple symbols) and ERA5 (turquoise symbols) when maximum (left panel) and average (right panel) values over boxes of 60 km are considered. The threat score is indicated in different shades of grey, and the results pertaining to various daily-precipitation intensities (with thresholds ranging from 1 to 150 mm·day^{-1}) are reported with different symbols. The frequency bias can be estimated from the deviations from the 45° black line, indicating an unbiased forecast (i.e. frequency bias = 1). Cross-hairs indicate the uncertainty related to the sampling variability of the data and are calculated from a bootstrap resampling of 1,000 new samples [Colour figure can be viewed at wileyonlinelibrary.com]

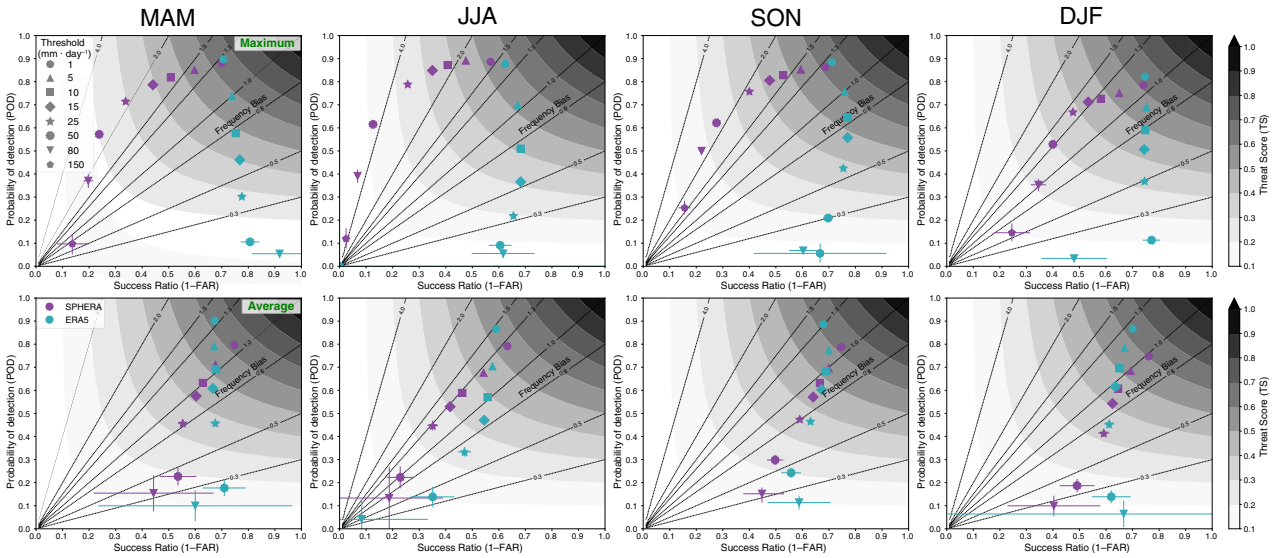

FIGURE 7 As Figure 6 but for seasonal aggregations over 2003 to 2017: MAM, JJA, SON and DJF (columns from left to right) when maximum (upper row) and average (lower row) values over boxes of 60 km are considered [Colour figure can be viewed at wileyonlinelibrary.com]

With respect to maximum precipitation (Figure 6, left plot), ERA5 outperforms SPHERA in the range of weak rainfall (1–10 mm·day^{-1}), especially due to fewer false alarms (i.e. higher SR score). As precipitation intensifies, the SR score stays almost constant for ERA5 (roughly 0.7) and is always higher than the SPHERA counterpart, which gradually decreases from 0.7 to 0.2. Conversely, the hit rate of ERA5 (i.e. POD score), with a difference of more than 0.1 points for each consecutive threshold, decreases more rapidly than for SPHERA, which stays above 0.7 up to 25 mm·day^{-1}. The result is a gradual increase in the score gap between the reanalyses in moderate- to-heavy precipitation (25–80 mm·day^{-1}). This suggests on the one hand a better ability of SPHERA to detect actually occurred heavy rainfall at the expense of a larger number of false detections. On the other hand, ERA5 shows a lower ability to simulate an adequate number of events for increasing rainfall intensity, but it keeps an almost constant

skill to avoid false alarms, due to the smaller sample of events simulated at increasing intensities. These outcomes are in line with the tendency of ERA5 to produce dry biases (as seen in Section 4.1) which are amplified for intense precipitation, as detected from the decreasing frequency bias from roughly 1.3 (1 mm·day^{-1}) to <0.3 (80 mm·day^{-1}). Similarly, SPHERA overestimates the number of events, with a frequency bias larger than 1 for the entire set of intensities, which is maximum (>2.0) for 50 mm·day^{-1}, in line with the weakly wet annual bias found in Figure 5d.

The analysis of the distributions for average rainfall intensities (Figure 6, right plot) confirms at first that the oversampling detected for maxima counterparts and for spatially distributed fields is not linked to a systematic wet bias of the model. In this case, a substantial reduction of the positive frequency bias of SPHERA is detected (always <1.3), which turns negative for rainfall >15 mm·day^{-1} and reaches a minimum of roughly 0.3 for 80 mm·day^{-1}. This indicates a better agreement with the number of observed events (revealed by the lower deviation from the null frequency bias line at all intensities). Also, ERA5 produces generally higher skill scores considering the averages. However, an increasing dry bias for intense accumulations, larger than the SPHERA counterpart, is detected starting from 15 mm·day^{-1} and reaches values <0.3 for 80 mm·day^{-1}. This further proves the larger difficulty for ERA5 in producing a sufficient number of events even to match the average observed state. Both reanalyses present similar TS scores, suggesting comparable skills to represent the average daily precipitation for weak to moderate rainfall. At the same time, the improvement of SPHERA is evident for the heaviest accumulations considered.

Figure 7 reports the seasonal analysis of the performance for maximum and average values in the boxes. The main differences compared to the annual aggregations are obtained in JJA and DJF, while March–May (MAM) and September–November (SON) are in line with the annual term performances (hence not discussed in the following). A general detriment of the scores is evident in JJA considering the maximum (upper row): ERA5 produces more false alarms and fewer hits, with SR and POD scores decreasing in the order of 0.1 points for most intensities. A performance deterioration in JJA is also detected for SPHERA in the form of a higher fraction of false detections (generally lower SR scores of 0.1 points than annual). Nevertheless, the ability to detect observed events is preserved and slightly enhanced compared to the annual term (POD scores systematically increased by roughly 0.1 points for most thresholds). This causes an exacerbation of the positive frequency bias at all intensities, which exceeds 4.0 for ≥50 mm·day^{-1}.

As regards DJF, SPHERA performance shows a general opposite behaviour compared to JJA (i.e. higher SR and lower POD and frequency bias at all intensities), as well as ERA5, even if less markedly (higher SR and lower POD). These results highlight the different seasonal predictability of precipitation. During summer, precipitation is mainly localized, short-lived, and linked to intense convective thunderstorms, making it typically more challenging to simulate compared to winter. In wintertime, stratiform rainfall dynamically driven by large-scale frontal activity is more likely (Houze, 2014; Antolini *et al.*, 2016). Interestingly, the performance of SPHERA suggests the added ability of the high-resolution simulations, particularly in JJA when compared to ERA5, to reach heavy rainfall intensities at the expense of a high number of unrequited events (false alarms). As demonstrated previously, the latter is more likely a consequence of the under-sampled observed state when considering extreme values (maxima) of precipitation rather than a deficit owing to the COSMO model. Indeed, moving to seasonal daily averages (Figure 7, lower row), a decisive reduction of the frequency bias of SPHERA is detected for all seasons (in JJA confined between 1.0 and 1.3 for the majority of intensities), suggesting a good agreement with the mean observed precipitation state. For ERA5, similarly to the annual term, the bias (generally wet for intensities ≤10 mm·day^{-1} and dry above) is reduced for the averages, but it always remains stronger than the SPHERA counterpart (i.e. larger deviations from the 1:1 line at all intensities and seasons).

A complementary aspect in comparing simulated and observed precipitation series are the rates of daily rainfall distribution intensities, reported as histograms in Figure 8 for average and maximum values. Focusing on maxima (left plot), in the range below 5 mm·day^{-1}, ERA5 overestimates while SPHERA underestimates the number of events. The trend is reversed for heavier intensities (i.e. ≥10–20 mm·day^{-1}), with SPHERA overpredicting and ERA5 under-sampling Dewetra. The maximum overestimation for SPHERA is detected in the range of 30–50 mm·day^{-1}, decreasing then for higher thresholds but consistently producing more events than observed. The underestimation committed by ERA5 gets larger with rainfall intensity, producing almost no events above 80 mm·day^{-1}. This supports the inadequacy of coarse and convection-parametrizing systems in representing severe precipitation occurrences. Conversely, SPHERA successfully simulates events at all rainfall intensities considered, even for the heaviest precipitation ranges (≥80–100 mm·day^{-1}), confirming the added value of CP models in this sense, as expected (e.g. Klasa *et al.*, 2018). As previously discussed, SPHERA overestimation of moderate to heavy precipitation is more a throwback of the upscaling aggregation based on maxima rather

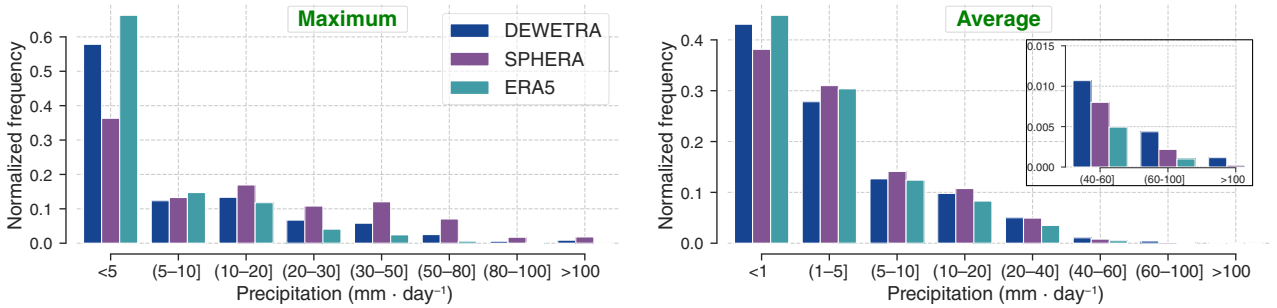

FIGURE 8 Normalized frequency histograms of the distributions of daily rainfall occurrences over 2003–2017 for SPHERA (purple), ERA5 (turquoise) and Dewetra observations (blue), when maximum (left panel) and average (right panel) values over boxes of 60 km are considered. For a better visualization the average values distributions for the highest thresholds are highlighted in a black-framed subplot [Colour figure can be viewed at wileyonlinelibrary.com]

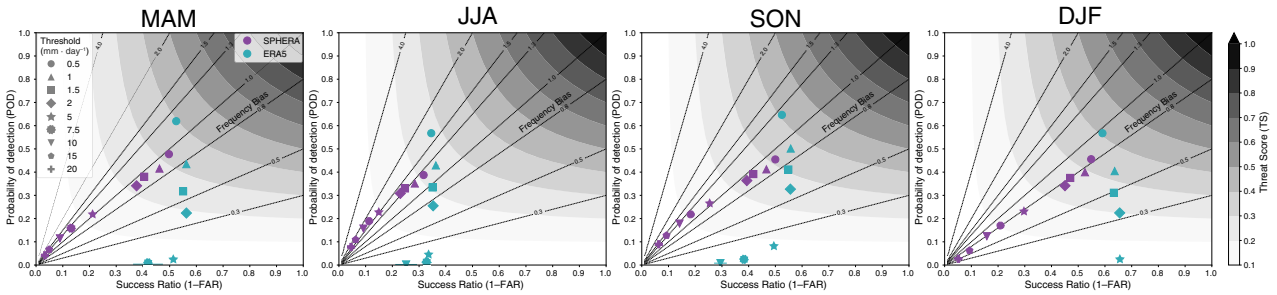

FIGURE 9 As Figure 7 but for hourly-accumulated precipitation, verified against Dewetra, when maximum values over boxes of 60 km are considered. The results pertain to a set of hourly precipitation thresholds ranging from 0.5 to 20 mm·hr^{-1}, reported with different symbols [Colour figure can be viewed at wileyonlinelibrary.com]

than a systematic tendency of the reanalysis. If Dewetra under-samples precipitation maxima, representatively discrepancies may arise in dataset comparisons when considering 60 km boxes. In this case, the number of SPHERA grid points falling within a box may be significantly larger than the corresponding number of rain-gauges. This benefits the reanalysis to the disadvantage of the sparser pluviometric network when coming to detect precipitation peaks. As a proof of this, the frequency distribution analysis performed with the averages (Figure 8, right plot) confirms the absence of the wet systematic bias in SPHERA: both reanalyses better agree with the observations, with SPHERA outperforming ERA5 starting from 5–10 mm·day^{-1}. Moreover, ERA5 underestimates the number of rainfall events starting from 10–20 mm·day^{-1}, even if less strongly than the maxima counterpart. This further confirms its increasing dry bias as precipitation intensifies.

4.3 | Hourly precipitation

The seasonal performance of the sub-daily (hourly) time-scale of precipitation is assessed through performance diagrams for maximum accumulations aggregated from 2003 to 2017 (Figure 9, similar to Figure 7). A systematic worsening of the scores is evident compared to daily accumulation counterparts. As with the daily performance, the added value of SPHERA over its driver emerges particularly in the larger fraction of hit events (starting from 2 mm·hr^{-1} for every season, but in summer from 5 mm·hr^{-1}), as indicated by the higher POD, with gaps as large as 0.2 points more than ERA5. Further, as in the daily analysis, ERA5 systematically produces fewer false detections than SPHERA, as indicated by the larger SR values obtained for every season and threshold, but under-samples the number of hourly rainfalls as the intensity increases. The main difference with daily results is the strongly reduced overestimation of the number of events produced by SPHERA compared to the observations. Indeed, the frequency bias for hourly rainfall maxima is always close to 1 for every intensity and season: it is confined to between 0.8 and 1.3 in MAM and SON, slightly dry in DJF (being arranged along the 0.8 line), and increasingly wet with rainfall intensity in JJA moving from roughly 1.0 to 2.0.

Figure 10 reports histograms of the relative frequency event distributions. The maximum hourly distributions

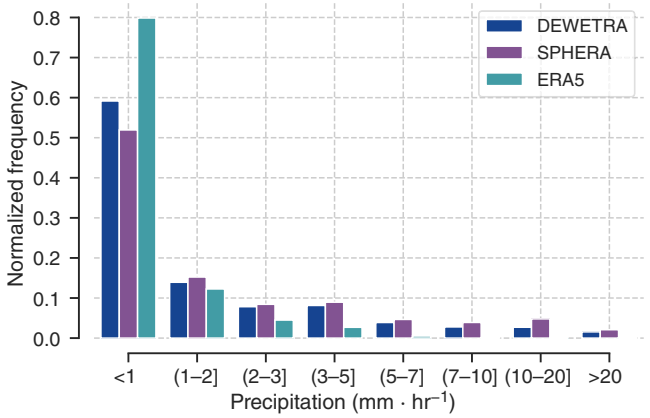

FIGURE 10 As Figure 8, left plot, but for hourly rainfall occurrences [Colour figure can be viewed at wileyonlinelibrary.com]

reflect the behaviour detected for daily frequencies (Figure 8, left plot). The overestimation committed by ERA5 in the number of low-precipitation events ($<1\,mm{\cdot}hr^{-1}$) is more pronounced, with the lowest bin populated by almost 80% of the entire sample of events, which is about 20% more than observed. Starting from $1–2\,mm{\cdot}hr^{-1}$, ERA5 frequency underestimation worsens as rainfall intensifies, producing an almost complete lack of occurrences for precipitation thresholds greater than $5–7\,mm{\cdot}hr^{-1}$. With respect to SPHERA, despite the significant underestimation of roughly 10% in the number of weak rainfalls ($<1\,mm{\cdot}hr^{-1}$), for heavier accumulations the number of events is always in agreement with the observations, even in the range of severe events ($>20\,mm{\cdot}hr^{-1}$). An overestimation with a frequency deviation roughly or below 2% is present, with the maximum in the range of 10 to $20\,mm{\cdot}hr^{-1}$. However, this effect is less pronounced than its daily counterpart in relative terms, indicating an improved skill in representing precipitation maxima at high temporal resolution.

As previously mentioned, when considering the maximum area-aggregated rainfall distributions over 60 km grids, important dynamical aspects of the simulations may be removed from the analysis, which is even more likely at hourly resolution. For this reason, it is worth investigating the frequency of hourly precipitation at the original resolutions of the datasets (i.e. 31 km for ERA5, 2.2 km for SPHERA, and scattered data points for pluviometers). It is worth highlighting the qualitative nature of this comparison given the numerous limitations arising from this approach (which were previously compensated by using a fuzzy verification method), namely, the representativeness limitations of the uneven distribution of individual point observations, the non-fixed number of rain-gauges in Dewetra over the years (increasing by a factor of 4), and the different sample sizes in grid points between the two reanalyses (as SPHERA is roughly 200 times denser than ERA5). The frequency distributions at the respective original resolutions of the datasets are reported in Figure 11. The results pertain to hourly summer rainfalls aggregated over 15 years when distinguishing between weak to moderate (Figure 11, left plot) and severe (right plot) rainfalls to visualize the less frequent severe events better. Since NWPs frequently simulate substantial amounts of low rainfall occurrences, especially when their fields are not spatially aggregated, only wet hours (i.e. precipitation exceeding $0.1\,mm{\cdot}hr^{-1}$, similar to Ban *et al.*, 2014) are considered. The results confirm and strengthen the findings of the upscaled dataset: the substantial overestimation of weak rainfalls committed by ERA5 is evident, as well as its inefficiency in simulating heavier precipitation events, producing less than 1% of its occurrences, versus the 11% of the observed, for precipitation amounts greater than $5\,mm{\cdot}hr^{-1}$. Additionally, SPHERA event distribution suits the observed counterpart well, without any systematic over- or underprediction. Indeed, SPHERA deviation from the observed frequency is always less than 5% in the range of weak to moderate precipitation ($0.1–5.0\,mm{\cdot}hr^{-1}$). Moving to higher intensities, SPHERA generally overestimates Dewetra, with deviations in the frequency below 2%, resulting in a population of occurrences in the $5–50\,mm{\cdot}hr^{-1}$ range equivalent to 14% of the total distribution, which is slightly larger than the observed 11%.

4.4 | Diurnal cycle

A further relevant feature to assess is the ability to represent the diurnal cycle of summer precipitation. In JJA, daily variations of convective activity strongly control weather dynamics over Italy. Hence, we expect an enhanced description when employing CP simulations (Fosser *et al.*, 2015; Brisson *et al.*, 2016). The diurnal time series of hourly averaged rainfall during JJA for SPHERA, ERA5 and Dewetra, aggregated over 2003–2017, are reported in Figure 12. Land grid points only are selected for the reanalyses and averaged over Italy. At the same time, hourly unboxed rain-gauge data are aggregated with the same premises described at the end of Section 4.3. Only wet hours are considered also in this comparison. The observed cycle is characterized by higher rainfall intensities between 0900 and 1700 UTC, with the peak at 1300 UTC, and weaker rainfall during the night and early morning, with a second lower peak between 2300 and 0000 UTC. ERA5 is unable to simulate the observed precipitation rates, implying strong underestimations of the overall intensity and amplitude variation as a natural consequence of the coarse horizontal resolution (Bollmeyer *et al.*, 2015), while the timing of the precipitation peak

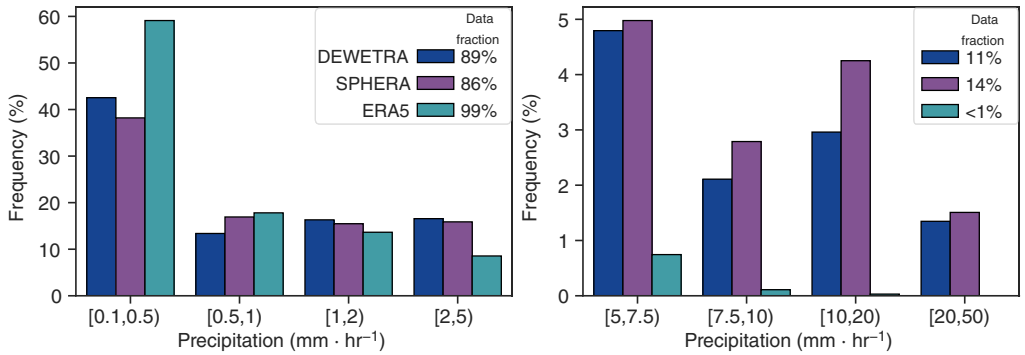

FIGURE 11 Relative frequency distributions of hourly rainfall occurrences for the summer months (JJA) aggregated over 2003 to 2017 for SPHERA (purple), ERA5 (turquoise) and Dewetra (blue). The distributions are obtained from the original unboxed horizontal resolutions of the datasets, and are divided in two plots to better highlight less frequent heavy precipitation events [Colour figure can be viewed at wileyonlinelibrary.com]

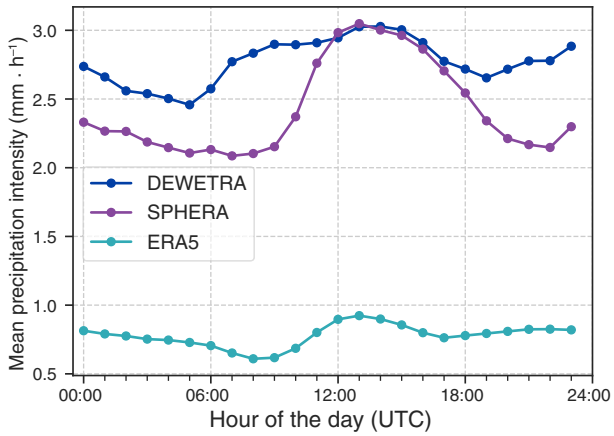

FIGURE 12 Mean diurnal cycle of hourly summer (JJA) precipitation intensity averaged over Italy during 2003–2017, and considering only wet hours of the day. The time series are calculated from the datasets at their original resolutions [Colour figure can be viewed at wileyonlinelibrary.com]

(1300 UTC) is well captured. The SPHERA diurnal cycle is more in line with the observations, with an optimal agreement during the rainiest hours of the day of the early afternoon (from 1200 to 1700 UTC) and a slight underestimation of the intensity (always $\leq 1\,\text{mm}\cdot\text{hr}^{-1}$) for the rest. SPHERA well replicates rainfall temporal evolution. The main difference with the observed cycle is evident in the morning from 0600 to 1100 UTC when observations exhibit higher precipitation intensities. No substantial temporal shifts in the timing of the wettest and driest hours of the day are found for both reanalyses.

5 | CASE-STUDIES

Besides the global statistical assessment for quantifying the overall ability of SPHERA and ERA5 to represent precipitation, it is relevant to investigate their simulation of specific meteorological conditions leading to extreme rainfall. For this reason, this section analyses two relevant case-studies of severe rainfall events in Italy: the precipitation occurrence leading to the flood of the Secchia river in January 2014 (Section 5.1), and the extreme precipitation associated with multiple mesoscale convective systems (MCS) over Sardinia in November 2013 (Section 5.2).

5.1 | Flood of the Secchia river (17–19 January 2014)

From 17 to 19 January 2014, the passage of an upper-level trough over north-central Italy produced moderate to very intense rainfall over the whole Po river valley. A large amount of rainfall fell over the Emilian catchments of the Secchia and Panaro rivers, two right tributaries of the Po river originating from the Apennines, thus leading to their flood. This event caused one fatality, the evacuation of approximately 10,000 people, and damage of roughly €500 million due to the failure of the Secchia river embankment (D'Alpaos *et al.*, 2014; Porcù and Aragão, 2019).

On the 17th, the low-level flow bringing maritime moist air, forced by the favourable synoptic configuration, impacted the northern Apennines and was forced to uplift. The result was the formation of intense and long-lasting orographic rainfall starting over the mountains and extending on their leeward side over a large sector of the southern Po valley. The associated precipitation event was active from the early morning of the 17th over the Apennine ridges, reaching heavy accumulations in a few hours and intermittently persisting up to the evening of the 19th, with a more widespread extension.

The accumulated precipitation in 72 hr simulated by SPHERA and by ERA5 are compared to the ARCIS

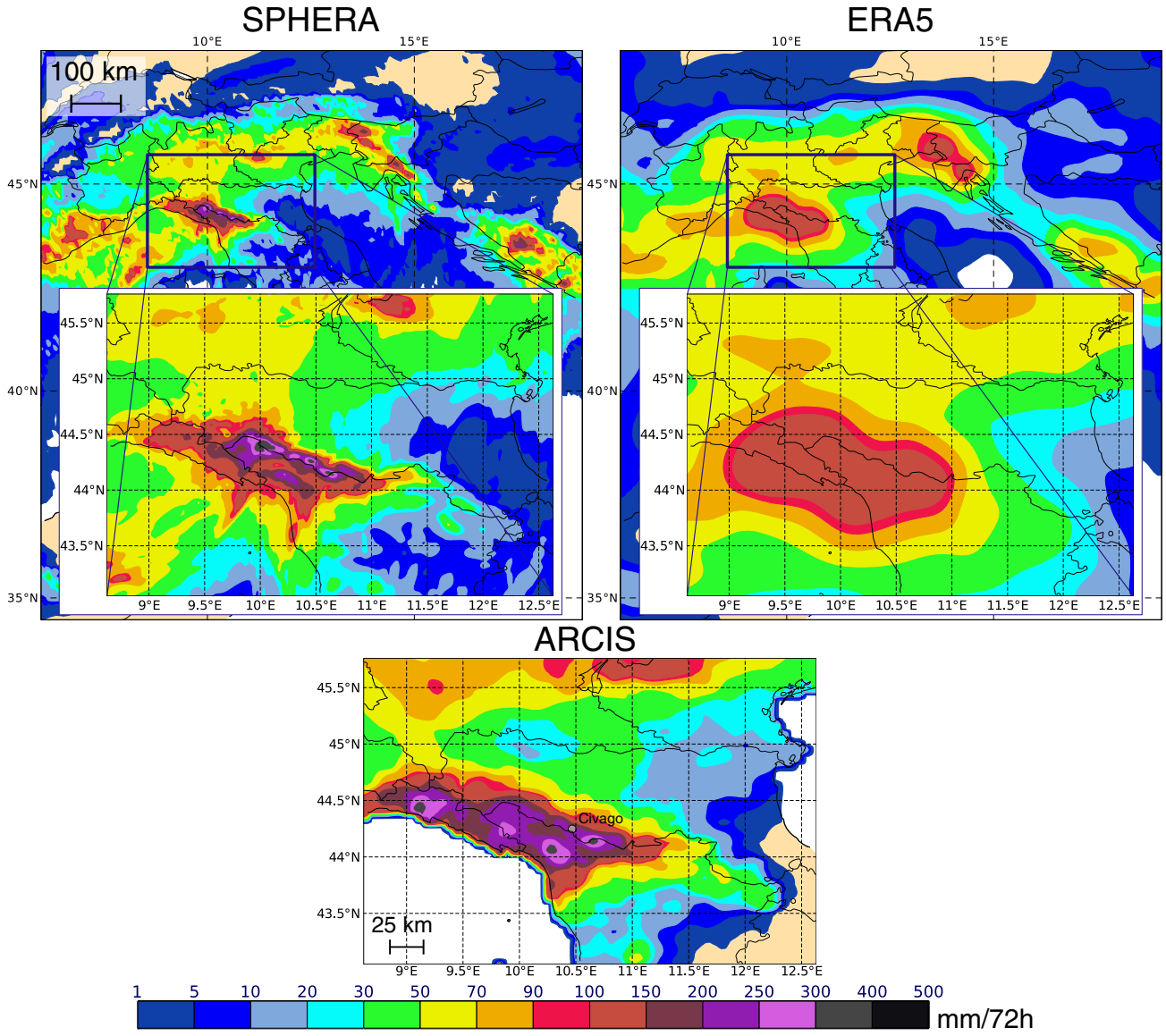

FIGURE 13 Maps of 72 hr accumulated rainfall fields (0000 UTC 17 January to 2300 UTC 19 January 2014), for SPHERA (upper-left panel), ERA5 (upper-right panel) over Italy, and ARCIS observative analysis (lower panel) over north-central Italy (the same sub-domain is reported for the reanalyses maps on their respective lower-left corners). The grey solid dot represents the location of Civago where the ARPA rain-gauge detected the maximum hourly amount of rainfall during the event (14.2 mm·hr^{-1}) [Colour figure can be viewed at wileyonlinelibrary.com]

analysis in Figure 13. The highest amounts of precipitation are reached over the Apennine sectors, especially over the Secchia river catchment (approximately around 44.3°N, 10.5°E). Here, the accumulations upstream exceed 250 mm/72 hr, affecting the overall river embankment stability and triggering the subsequent failure of the embankment downstream. SPHERA simulates rainfall peaks reaching values as high as 300 mm/72 hr over the river basin, in good agreement with the observed magnitude. In comparison, the rainfall field simulated by ERA5 does not exceed maximum accumulations of 150 mm/72 hr in the area. Furthermore, SPHERA spatial distribution presents complex details as a consequence also of the highly resolved topography in the area. As a result, the localization of the most prominent rainfall peaks is sharp and in approximate (but non-exact) agreement with the respective observed pattern, with a tendency for displacing the wettest spots towards the northeast. On the other hand, the insufficient spatial detail of ERA5 produces a blunted, widespread and smoothed rainfall field over the event area.

The observed maximum rainfall rates are recorded on the late morning of the 17th by a few isolated mountain rain-gauges: the absolute maximum is detected in

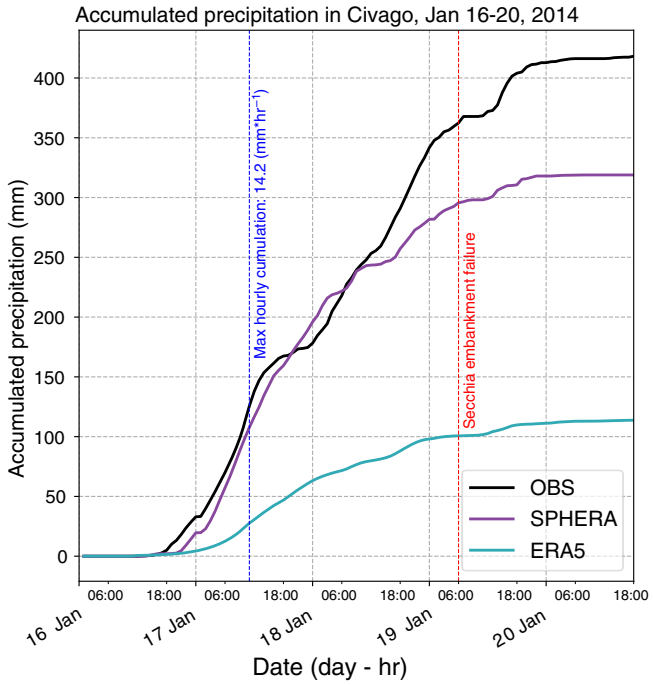

FIGURE 14 Accumulated precipitation for the period 16–20 January 2014, observed (black line) and simulated by SPHERA (purple line) and ERA5 (turquoise line) at the station of Civago. The vertical blue dashed line reports the timing of the observed hourly-cumulated precipitation peak, while the red dashed line indicates the timing of the Secchia river embankment failure downstream [Colour figure can be viewed at wileyonlinelibrary.com]

Civago (1,011 m a.s.l), located upstream in the Secchia river catchment, with 14.2 mm·hr^{-1} over 1100 UTC. The most widespread and long-lived phenomena occurred during the evening and night of the 18th when almost the entirety of pluviometers located in the Panaro and Secchia rivers basins recorded more than 5 mm·hr^{-1} for at least three consecutive hours (Porcù and Aragão, 2019). Figure 14 reports the temporal evolution of the heavy accumulation over Civago. A bilinear interpolation is applied to the reanalyses to perform a point-to-point comparison with pluviometric data. A total of 174.6 mm fell by the end of the first day of the event (17th). SPHERA well represents this steep accumulation trend by matching the observed peak time (between 0900 and 1200 UTC) and simulating 189.1 mm by the end of the 17th (i.e. roughly 8% more than observed). By the end of the third day, characterized by less intense and more spatially distributed rainfall, SPHERA underestimates the total accumulation by 24%, producing 318.9 mm versus the 419.0 mm observed. The high-resolution simulation maintains a good match with rainfall timing throughout the event (as evident by the black and purple profiles in Figure 14, particularly during the intermittent phase of precipitation on 19 January). Moving to ERA5, its simulated field largely underestimates rainfall accumulation for the entire duration of the event. The underprediction is 113.8 mm (i.e. 65% less than observed) by the end of the 17th, increasing to 305.3 mm (i.e. 73%) by the end of the 19th. ERA5 also delays rainfall timing throughout the 3 days, indicating a lower ability to predict the correct temporal sequence during all the phases of the orographically driven event.

This kind of extreme precipitation event, generated by the interaction with the orography of a strong and humid confluent flow ahead of a polar cold front, is frequent in the Mediterranean region during the winter season (Krichak *et al.*, 2015). Evidence is provided by the observed spatial distribution in DJF (Figure 5c) and recent investigations (e.g. Grazzini *et al.*, 2020a, 2020b).

5.2 | Cyclone *Cleopatra* and MCSs over Sardinia (18 November 2013)

During the autumn season over the Mediterranean Sea, a series of factors may favour the formation and organization of convective activity, namely, the strong temperature gradients between the warm subtropical air and the colder northern air masses, the still-warm sea surface releasing large amounts of moisture in the lower troposphere, and atmospheric instabilities caused by the more frequent changes in the baric configurations at these latitudes following the end of the warm season (Jansa *et al.*, 2000, 2001; Nieto *et al.*, 2005; Caillaud *et al.*, 2021).

In this framework, the extratropical cyclone *Cleopatra* formed in November 2013, starting from a deep low-pressure trough in the westerlies and evolving into an upper-air cut-off low-pressure system centred in southern Europe. The result was extreme rainfall causing extensive damage, loss of lives, and record-breaking accumulations (ARPAS, 2014). The dynamical evolution of *Cleopatra*, and the interaction with the complex orography of Sardinia led to the development on 18 November of multiple stationary MCSs over the island producing extreme precipitation. These are visible from the daily accumulation estimated through SRI data obtained by combining radar with rain-gauge observations (Figure 15, right plot). The central and left plots of Figure 15 report the respective daily accumulated rainfall fields simulated by SPHERA and ERA5. The ability of the high-resolution reanalysis to simulate the convective band structures leading to severe rainfall is evident. ERA5 produces a well-localized precipitation field. However, its results are excessively smoothed, with low spatial detail, presenting daily accumulations below 90 mm. The accumulated intensities of SPHERA show lower underestimation than ERA5: the simulated values reach up to 150 mm versus the >400 mm detected, which can be expected when comparing reanalysis simulations

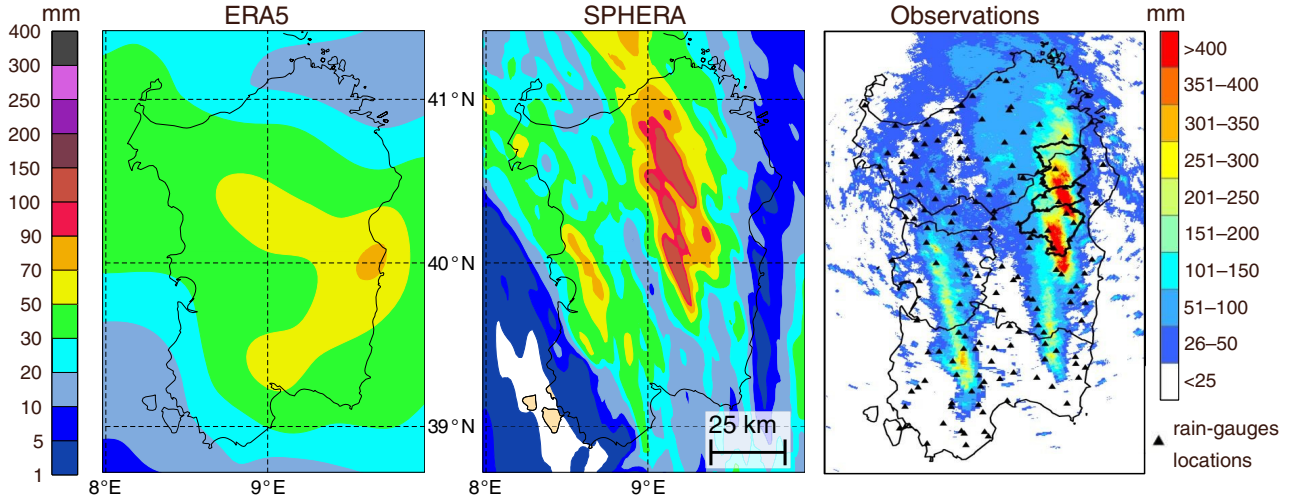

FIGURE 15 Daily accumulated precipitation for 18 November 2013, on the left panel simulated by ERA5, on the central panel simulated by SPHERA, on the right panel obtained from the combination of radar estimates with pluviometric observations (surface rainfall intensity [SRI]) (taken from Niedda *et al.*, 2014). On the right plot the rain-gauges over the territory are indicated with black solid triangles [Colour figure can be viewed at wileyonlinelibrary.com]

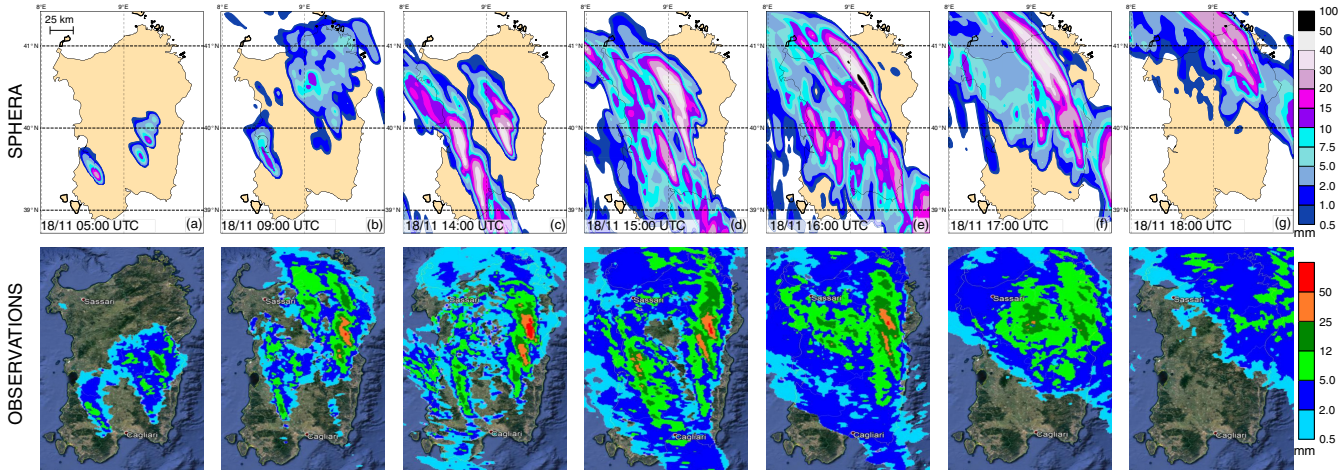

FIGURE 16 (a–g) Hourly accumulated precipitation for progressive hours of 18 November 2013 simulated by SPHERA (upper row) and estimated as surface rainfall intensity (SRI) from radar/rain-gauge data (lower row) [Colour figure can be viewed at wileyonlinelibrary.com]

to observational datasets in case of extreme convective rainfall (Hu and Franzke, 2020).

To get better insights into the temporal evolution of the event, a comparison between SPHERA accumulated rainfall and radar SRI estimates at hourly temporal resolution is reported (Figure 16). During the early morning (at 0500 UTC) of the 18th (panel a), SRI fields reveal the initiation of moderate and localized rainfalls exceeding 12 $mm \cdot hr^{-1}$ in central-southern Sardinia. These are associated with the formation of the first convective cells starting to get organized along narrow bands aligned with the low-level flow direction (i.e. northwest) and resulting from the orographic lift of the advected subtropical air. The radar-based imagery then shows the permanence and development of the convective bands during the rest of the morning, gradually producing more sustained rainfall with values larger than 25 $mm \cdot hr^{-1}$ (panel b). Convective precipitation significantly intensifies and extends starting from the early afternoon (1400 UTC, panel c), owing to the upscale growth of the MCSs, as a result of the gradual aggregation of the isolated convective cells, with particular emphasis over central-eastern Sardinia. The peaked extension and strengthening of the MCSs derive from the uninterrupted orographic lift, coupled with the gradual approach of the strong thermal and pressure gradients linked with the arrival of the cyclonic front (not shown). This configuration triggers a decisive increase in the convective activity and organization, besides favouring strong

stationarity and then regeneration of the convective cells. The marked stationarity of the MCSs is visible from the comparison of SRI fields at 0900 UTC (panel b) with 1400 UTC (panel c), showing no significant spatial shifts of the heavy precipitation bands in 5 hr. The two most intense band structures located in central-eastern Sardinia are characterized by very similar geometries and dimensions, with cross-sections in the range of 5 to 15 km, and presenting extreme spatial gradients in precipitation at their borders (approximately 200 mm/5 km), as further proof of their sharpened stationarity (Niedda *et al.*, 2014). SPHERA successfully simulates the correct timing of hourly accumulations throughout 18 November, producing a remarkable agreement with the observed precipitation field (Figure 16c,g). Already from the early morning, the initiation of isolated and moderate rainfall is satisfactorily simulated, forming along short bands that gradually organize and extend to larger areas. From 1400 UTC, the simulated intensification and extension of the MCSs associated with the cold-front passage is evident by the resulting heavier precipitation rates. However, SPHERA rainfall estimates are spatially shifted compared to the observations: the former are located more westward (by ~20 km) and have their axis slightly tilted northwestward compared to the more northward-oriented radar-based structures. The cause is possibly a slightly different localization of the simulated surface low-pressure minimum of the cyclone (not shown): being further south and west than observed it causes a slight tilt in the convective band axis orientation and a westward shift of their localization, respectively.

6 | DISCUSSION AND CONCLUSION

Rainfall is one of the most critical meteorological quantities to be estimated in NWPs and climate monitoring contexts. Severe rainfall represents one of the most important causes of extended damage and societal costs in Europe (Rebora *et al.*, 2013; Spekkers *et al.*, 2017), hence the improvement in its representation is crucial (Feng *et al.*, 2021). The assessment of the potential added value in the representation of extreme rainfalls as simulated by SPHERA, a new CP regional reanalysis over Italy, is the subject of the present investigation. SPHERA covers 1995–2020 with hourly temporal frequency at 2.2 km horizontal resolution on 65 vertical levels (0–22 km) and seven soil levels (0–14.58 m). The validation of precipitation simulated by SPHERA, and the comparison with its driver ERA5, is performed against the national rain-gauge network Dewetra for 2003–2017. The methodology consists of a spatial-neighbourhood technique after assessing the optimal scale size to operate the upscaling aggregation. An excessive loss of detail in the representation of precipitation fields is detected, especially for heavy rainfall occurrences, when using grid boxes excessively wide (i.e. 100 or 200 km), as expected (Marsigli *et al.*, 2008), while finer resolutions have proven to perform similarly to each other. Hence, a grid spacing permitting an adequate sampling of both SPHERA and ERA5 is chosen (i.e. 60 km).

The statistical analysis of precipitation permits assessing the added value of the CP system, particularly in terms of precipitation extremes for which higher benefits are expected. Concerning heavy-rainfall geographical distributions, ERA5 locally underestimates precipitation intensity throughout the year, especially over the mountainous wettest regions (with a bias as large as $-12\,\mathrm{mm\cdot day^{-1}}$ in DJF). SPHERA shows wet biases up to $12\,\mathrm{mm\cdot day^{-1}}$ in JJA over the plains and dry biases of $-9\,\mathrm{mm\cdot day^{-1}}$ in DJF. Moving to the analysis of daily-accumulated rainfall, ERA5 better represents weak precipitation intensities with generally higher skill scores and weak wet biases. The benefits of the CP setting are evident for moderate and heavy accumulations in all seasons. In these cases, SPHERA successfully simulates severe rainfall occurrences, mainly related to warm-season locally driven convective events, which ERA5 strongly underrepresents. The CP enhancement in skill scores shows a tendency to overpredict the number of events when considering daily distribution maxima. This is most likely due to under-sampling of the observed state rather than a systematic deficit of the model, as proved by the almost null frequency bias obtained for daily average precipitation. In terms of hourly accumulation maxima, the accordance with the observed frequency stays stronger for SPHERA at all rainfall intensities. At the same time, ERA5 produces an insufficient number of events worsening with accumulation intensity. A further proof of the added value of SPHERA is the improved accordance with the observed average daily cycle of summer precipitation, compared to the consistent amplitude underestimation proper of ERA5.

Two case-studies are considered to assess the ability of SPHERA to represent extreme precipitation and to further compare the reanalysis systems in describing specific severe-rainfall events. Both cases relate to extreme accumulations over a few hours resulting from orographic lifting and MCSs producing floods and extensive damage. ERA5 simulates blunted and smooth rainfall fields for both events, underpredicting the observed intense accumulations by 73% (when looking at the rainiest location for one of the events). SPHERA successfully represents the precipitation fields with a higher level of detail in their spatial distribution, hourly-frequency timing, and intensity of extreme precipitation. Anyhow, an underestimation of

24% is detected in the same location (simulating an accumulated rainfall peak of 318 mm/72 hr versus the observed 419 mm/72 hr). This result is expected when considering the simulation of heavy rainfall accumulations, even when performed with very fine spatial grid spacings (e.g. Buzzi *et al.* (2014) using 1.5 km CP simulations obtained an underestimation of 38% for a severe-precipitation event exceeding 500 mm/24 hr). The results suggest the potential of the CP system to reach high precipitation rates. In comparison, the stronger underestimation of the coarser ERA5 is most likely linked to the convection parametrization included in the driver model. In this case, smooth rainfall fields and a decreasing agreement with the observed state with increasing rainfall intensity is obtained, as expected when tailing the analysis towards the extremes (Rivoire *et al.*, 2021; Bandhauer *et al.*, 2022). The underestimation detected with SPHERA may potentially be linked to the spatial shifts in the localization of the rainiest areas and is acceptable in the context of high-resolution weather simulations. Indeed, these mismatches fall into the double-penalty class of occurrences, for which it is challenging to match the observations with absolute precision. The cause is the intrinsic chaotic behaviour controlling deep moist convection, which is responsible for the low predictability of the exact localization of the associated convective processes, justifying the use of fuzzy verification techniques for quantifying the performance of their simulations (Marsigli *et al.*, 2021). This issue could be relevant in downstream modelling applications that require high spatial precision, such as in hydrological modelling (e.g. Lobligeois *et al.*, 2014). Possible strategies to overcome spatial phase errors in high-resolution simulations when dealing with localized rainfall are the additional assimilation of radar observations. This can be done by employing a latent-heat nudging scheme (Wahl *et al.*, 2017) or post-processing reanalysis data with optimal interpolation techniques (Bonanno *et al.*, 2019).

These results prove the added value of SPHERA for describing medium-to-severe local precipitation events, owing to several improvements compared to its global driver. These are: the finer grid spacing allowing for an enhanced physical and microphysical description of the processes, primarily due to not needing deep convection parametrization, the resulting higher level of topography detailing allowing for a better representation of atmosphere–land interactions, and the better adherence to assimilated regional observations. Previous studies reported multiple benefits related to increased grid resolutions in numerical simulations, allowing an adequate representation of local dynamical features and forcings leading to or intensifying precipitation events (Buzzi *et al.*, 2014; Cassola *et al.*, 2015; Clark *et al.*, 2016; Wahl *et al.*, 2017; Klasa *et al.*, 2018; Cerenzia *et al.*, 2020; Capecchi, 2021). Furthermore, the results obtained are in line with those of recently produced CP hindcasts over Italy, sharing similar characteristics with SPHERA, and obtained by downscaling ERA5 with the BOLAM/-MOLOCH model (Capecchi *et al.*, 2022) or COSMO model (Raffa *et al.*, 2021; Reder *et al.*, 2022). In Capecchi *et al.* (2022), similar wet frequency biases in reproducing the 90th percentiles of annual, daily and hourly rainfalls are detected for MOLOCH simulations at 2.5 km grid spacing, as opposed to the dry biases detected with the BOLAM run at 7 km. Further, from the analysis of two severe-precipitation events, a higher level of detail in the spatial characterization and less deviation from maximum intensities is maintained with the CP hindcast, despite underestimating the most extreme rainfall observations (by 64% in one case). As regards the hindcast obtained by downscaling ERA5 at 2.2 km over 20 European cities (Reder *et al.*, 2022), enhancements over its driver are obtained in terms of the spatial patterns of annual precipitation and the relative annual maxima, as well as the monthly and hourly cycles of precipitation. In fact, ERA5 tends to simulate precipitation fields that are too smooth to give an appropriate representation of rainfall extremes, in agreement with the results of the present article. A significant difference between SPHERA and the CP-reforecasts mentioned above is the lack of the additional assimilation of regional observations in the latter. The assimilation procedure in the production phase of the dataset is known to substantially enhance the representation of the simulated atmospheric state (Bollmeyer *et al.*, 2015; Bonanno *et al.*, 2019). In SPHERA, this is accomplished by including several conventional observations on the near-surface (SYNOP and SHIP) or upper-air (TEMP, PILOT and AIREP) through a continuous nudging scheme, permitting a better adherence of the past-weather simulations to the observed atmospheric state. The same observations are also assimilated in the global driver ERA5. However, the much finer grid spacing at which they are ingested in SPHERA (2.2 km vs. 31 km of ERA5) substantially impacts their representativeness. Furthermore, it is worth noting that neither reanalysis assimilates precipitation observations.

Concerning the observational data considered as the reference state, a crucial aspect emerging is the critical dependency of the performance of SPHERA and ERA5 on the quality of rain-gauge measurements. In fact, as highlighted several times in the article, sparse pluviometric data such as those of Dewetra, despite the enhanced spatial density over the years, may be affected by various representativity issues, namely: measurement errors, spatial inhomogeneities due to lower station coverage in some areas (i.e. especially in southern Italy or in mountainous regions, where additionally slightly different sensor positioning

may cause significant underestimations in the recorded data: Crespi *et al.*, 2018), and the non-constant number of the sensors during the years examined which increased by roughly 400% (from 1,500 in 2003 to 6,200 in 2017). For these reasons, the results should be interpreted with caution, as the reference state is far from ideal. A possible way to partly overcome spatial inhomogeneity issues may be to consider multiple observation datasets, especially in the form of high-resolution gridded analyses such as GRIPHO (Fantini, 2019) or ARCIS (Pavan *et al.*, 2019) over Italy. The combined use of these datasets could enhance the uncertainty quantification in describing the actual atmospheric state. Anyhow, the data interpolation required to construct these gridded datasets may be an additional source of error due to the resulting intrinsic smoothing of the rainfall fields, which constitutes a major limit, particularly for the representation of extreme precipitation.

The encouraging enhancements revealed by simulating heavy rainfall and convection with CP settings still leave room for improvement. Even if reduced, relevant uncertainties in the simulations are still present at these scales. A major one is the inability to explicitly resolve shallow convection at the CP scales (Khairoutdinov and Randall, 2006), which may also be detrimental in the simulation of deep convection given the intrinsic bond between these two processes (Teixeira *et al.*, 2008). Consequently, CP simulations are highly model-dependent, posing the problem of the robustness of the results. A way to overcome this issue is through multi-model ensemble-based approaches with which it would be possible to "advance parameterizations of unresolved physics and to assess the full potential of CP models" (Prein *et al.*, 2015). In this context, several recent European efforts demonstrated the potential that multi-model CP regional climate simulations have for numerous aspects: better understanding the response of convection extremes to human-induced climate change and providing critical added value to decision-makers due to the enhanced confidence in simulating convection extremes (Coppola *et al.*, 2020), more realistic representation of heavy precipitation with a significant reduction of the hourly summer bias and a reduction of their temporal uncertainty (Ban *et al.*, 2021), and improved representation of fine-scale details of seasonal, daily and hourly heavy rainfall (Pichelli *et al.*, 2021). In light of this, we believe that the combination of the efforts leading to the recent development of similar CP reanalysis/hindcast datasets over Italy (such as SPHERA or Capecchi *et al.* (2022) and Reder *et al.*, 2022) is of paramount importance. Hence, the proposal is to jointly develop the first Italian multi-model high-resolution reanalysis/hindcast ensemble. With this tool, we expect to better assess the uncertainty of past climate, with a particular focus on high-impact convective events. Further, it would be possible to quantify the role of data assimilation in the production of SPHERA given the lack of this component in its similar datasets.

To conclude from a broader perspective, numerous European achievements have been recently made to improve our understanding of hydro-meteorological hazards, assess their risk, and mitigate their effects through disaster-risk reduction strategies (Shah *et al.*, 2020). Among them, nature-based solutions play a central role. Nature-based solutions, conversely to traditionally engineered risk-reduction strategies, are designed to be long-lasting, cost-effective, and environmentally sustainable (Sahani *et al.*, 2019). Several European projects aim at the investigation and enhancement of nature-based solutions, such as OPERANDUM (OPEn-air laboRAtories for Nature baseD solUtions to Manage hydro-meteo risks Mapping: Debele *et al.*, 2020), PHUSICOS ("According to nature" in Greek: Baills *et al.*, 2021), or RECONECT (Regenerating ECOsystems with Nature-based solutions for hydro-meteorological risk rEduCTion: Ruangpan *et al.*, 2020). The effectiveness of nature-based solutions can be improved with a finer knowledge of the small-scale heterogeneities characterizing the spatio-temporal meteorological configurations (Qiu *et al.*, 2019). This is possible through CP simulations able to better describe high-impact weather events, such as severe precipitation and wind gusts associated with deep convection activity (Weisman *et al.*, 1997). Finally, applications in climate monitoring can also benefit from high-resolution datasets, which can improve the understanding of climate change impacts on a regional basis, help identify the mechanisms accountable for local climatic features, and add information on the climatology of severe weather-related phenomena.

AUTHOR CONTRIBUTIONS

Antonio Giordani: Data curation; formal analysis; investigation; methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Ines Maria Luisa Cerenzia:** Data curation; formal analysis; methodology; resources; software; validation; writing – review and editing. **Tiziana Paccagnella:** Conceptualization; funding acquisition; project administration; resources; supervision. **Silvana Di Sabatino:** Conceptualization; funding acquisition; project administration; supervision.

ACKNOWLEDGMENT

This work is carried out under the framework of the OPERANDUM (OPEn-air laboRAtories for Nature baseD solUtions to Manage hydro-meteo risks) project, which is funded by the European Union's Horizon 2020 research

and innovation programme under the Grant Agreement No. 776848, and of the project TRIGGER (solu-Tions foR mitiGatinG climate-induced hEalth thReats) funded by the European Union under the Grant Agreement No. 101057739. The authors thank two anonymous reviewers whose suggestions led to a substantial improvement of the manuscript, the colleagues of the modelling group at ARPAE-SIMC for the useful discussions, and ECMWF and Copernicus for making the ERA5 dataset available. Open Access Funding provided by University of Bologna within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ORCID

Antonio Giordani https://orcid.org/0000-0001-7480-6283
Ines Maria Luisa Cerenzia https://orcid.org/0000-0001-9902-3701
Tiziana Paccagnella https://orcid.org/0000-0001-6422-7341
Silvana Di Sabatino https://orcid.org/0000-0003-2716-9247

REFERENCES

Antolini, G., Auteri, L., Pavan, V., Tomei, F., Tomozeiu, R. and Marletto, V. (2016) A daily high-resolution gridded climatic data set for Emilia-Romagna, Italy, during 1961–2010. *International Journal of Climatology*, 36(4), 1970–1986.

ARPAS, Regional agency for environmental protection of Sardinia (2014). *Analysis of the weather event of November 18, 2013. Technical Report*. http://www.sar.sardegna.it/pubblicazioni/miscellanea/analisi_evento_18nov2013.pdf [Accessed 10th February 2023].

Avgoustoglou, E., Matsangouras, I.T., Pytharoulis, I., Kamperakis, N., Mylonas, M., Nastos, P.T. and Bluestein, H.W. (2018) Numerical modeling analysis of the mesoscale environment conducive to two tornado events using the COSMO. Gr model over Greece. *Atmospheric Research*, 208, 148–155.

Bailles, A., Garcin, M. and Bernardie, S. (2021) Platform dedicated to nature-based solutions for risk reduction and environmental issues in hilly and mountainous lands. *Sustainability*, 13(3), 1094.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T. (2011) Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Monthly Weather Review*, 139(12), 3887–3905.

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., Ahrens, B., Alias, A., Anders, I., Bastin, S., Belušić, D., Berthou, S., Brisson, E., Cardoso, R.M., Chan, S.C., Christensen, O.B., Fernández, J., Fita, L., Frisius, T., Gašparac, G., Giorgi, F., Goergen, K., Haugen, J.E., Hodnebrog, Ø., Kartsios, S., Katragkou, E., Kendon, E.J., Keuler, K., Lavin-Gullon, A., Lenderink, G., Leutwyler, D., Lorenz, T., Maraun, D., Mercogliano, P., Milovac, J., Panitz, H.J., Raffa, M., Remedio, A.R., Schär, C., Soares, P.M.M., Srnec, L., Steensen, B.M., Stocchi, P., Tölle, M.H., Truhetz, H., Vergara-Temprado, J., de Vries, H., Warrach-Sagi, K., Wulfmeyer, V. and Zander, M.J. (2021) The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Climate Dynamics*, 57(1), 275–302.

Ban, N., Schmidli, J. and Schär, C. (2014) Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, 119(13), 7889–7907.

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O.E. and Frei, C. (2022) Evaluation of daily precipitation analyses in E-OBS (v19. 0 e) and ERA5 by comparison to regional high-resolution datasets in European regions. *International Journal of Climatology*, 42(2), 727–747.

Berg, P., Wagner, S., Kunstmann, H. and Schädler, G. (2013) High resolution regional climate model simulations for Germany: part I—validation. *Climate Dynamics*, 40(1), 401–414.

Bollmeyer, C., Keller, J.D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., Redl, S. and Steinke, S. (2015) Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society*, 141(686), 1–15.

Bonanno, R., Lacavalla, M. and Sperati, S. (2019) A new high-resolution meteorological reanalysis Italian dataset: MERIDA. *Quarterly Journal of the Royal Meteorological Society*, 145(721), 1756–1779.

Brisson, E., Van Weverberg, K., Demuzere, M., Devis, A., Saeed, S., Stengel, M. and van Lipzig, N.P. (2016) How well can a convection-permitting climate model reproduce decadal statistics of precipitation, temperature and cloud characteristics? *Climate Dynamics*, 47(9), 3043–3061.

Bromwich, D.H., Wilson, A.B., Bai, L., Liu, Z., Barlage, M., Shih, C.F., Maldonado, S., Hines, K.M., Wang, S.H., Woollen, J., Kuo, B., Lin, H.C., Wee, T.K., Serreze, M.C. and Walsh, J.E. (2018) The Arctic system reanalysis, version 2. *Bulletin of the American Meteorological Society*, 99(4), 805–828.

Buzzi, A., Davolio, S., Malguzzi, P., Drofa, O. and Mastrangelo, D. (2014) Heavy rainfall episodes over Liguria in autumn 2011: numerical forecasting experiments. *Natural Hazards and Earth System Sciences*, 14(5), 1325–1340.

Caillaud, C., Somot, S., Alias, A., Bernard-Bouissières, I., Fumière, Q., Laurantin, O., Seity, Y. and Ducrocq, V. (2021) Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. *Climate Dynamics*, 56(5), 1717–1752.

Calvo Sánchez, F. J., Díez Muyo, M. V., Escribá, P., Jiménez de Mingo, A., Martín Pérez, D., Morales Martín, G., Navascués, B., Peral García, M.C., Sánchez Arriola, J. and Viana Jiménez, S., (2021). Prototype of a high resolution regional reanalysis for Iberian Peninsula an Balearic Islands (IBERA), Agencia Estatal de Meteorología, 1st ACCORD All Staff Workshop, 12–16 April 2021. http://hdl.handle.net/20.500.11765/12840.

Capecchi, V. (2021) Reforecasting two heavy-precipitation events with three convection-permitting ensembles. *Weather and Forecasting*, 36(3), 769–790.

Capecchi, V., Pasi, F., Gozzini, B., & Brandini, C. (2022). A convection-permitting and limited-area model hindcast driven by

ERA5 data: precipitation performances in Italy. Climate Dynamics, 1–27. https://doi.org/10.1007/s00382-022-06633-2.

Cassola, F., Ferrari, F. and Mazzino, A. (2015) Numerical simulations of Mediterranean heavy precipitation events with the WRF model: a verification exercise using different approaches. *Atmospheric Research*, 164, 210–225.

Castro, C.L., Pielke, R.A., Sr. and Leoncini, G. (2005) Dynamical downscaling: assessment of value retained and added using the regional atmospheric modeling system (RAMS). *Journal of Geophysical Research: Atmospheres*, 110, D05108. https://doi.org/10.1029/2004JD004721.

Cerenzia, I.M.L., Giordani, A., Paccagnella, T. and Montani, A. (2022) Towards a convection-permitting regional reanalysis over the Italian domain. *Meteorological Applications*, 29(5), e2092.

Cerenzia, I.M.L., Pincini, G., Paccagnella, T., Minguzzi, E., Gastaldo, T., Poli, V., Tesini, M.S., Patruno, P. and Cesari, D. (2020) Forecast of precipitation for the 1994 flood in Piedmont: performance of an ensemble system at convection-permitting resolution. *Bulletin of Atmospheric Science and Technology*, 1(3), 319–338.

Clark, P., Roberts, N., Lean, H., Ballard, S.P. and Charlton-Perez, C. (2016) Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23(2), 165–181.

Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., Ban, N., Bastin, S., Belda, M., Belusic, D., Caldas-Alvarez, A., Cardoso, R.M., Davolio, S., Dobler, A., Fernandez, J., Fita, L., Fumiere, Q., Giorgi, F., Goergen, K., Güttler, I., Halenka, T., Heinzeller, D., Hodnebrog, Ø., Jacob, D., Kartsios, S., Katragkou, E., Kendon, E., Khodayar, S., Kunstmann, H., Knist, S., Lavín-Gullón, A., Lind, P., Lorenz, T., Maraun, D., Marelle, L., van Meijgaard, E., Milovac, J., Myhre, G., Panitz, H.J., Piazza, M., Raffa, M., Raub, T., Rockel, B., Schär, C., Sieck, K., Soares, P.M.M., Somot, S., Srnec, L., Stocchi, P., Tölle, M.H., Truhetz, H., Vautard, R., de Vries, H. and Warrach-Sagi, K. (2020) A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dynamics*, 55(1), 3–34.

Crespi, A., Brunetti, M., Lentini, G. and Maugeri, M. (2018) 1961–1990 high-resolution monthly precipitation climatologies for Italy. *International Journal of Climatology*, 38(2), 878–895.

D'Alpaos, L., Brath, A., Fioravante, V., Gottardi, G., Mignosa, P., & Orlandini, S. (2014). Relazione tecnico-scientifica sulle cause del collasso dell'argine del fiume Secchia avvenuto il giorno 19 gennaio 2014 presso la frazione San Matteo. Report of the Emilia-Romagna Region. https://93.46.106.184:90/doc_relazione_Secchia_lug14.pdf [Accessed 10th February 2023].

Debele, S.E., Kumar, P., Sahani, J., Bowyear, P., Pröll, J., Preuschmann, S., Mickovski, S.B., Ukonmaanaho, L., Zieher, T., Rutzinger, M., Gallotti, G., Aragão, L., Bagaglini, L., Stefanopoulou, M., Panga, D., Finér, L., Pouta, E., Santo, M.A., Korhonen, N., Pilla, F., Sarkar, A., and Basu, B. (2020). OPEn-air laboRAtories for Nature baseD solUtions to Manage hydro-meteo risks: Critical evaluation of risks and opportunities for OPERANDUM OALs, Report Number D1.2. https://doi.org/10.13140/RG.2.2.29763.43048

Desamsetti, S., Dasari, H.P., Langodan, S., Titi, E.S., Knio, O. and Hoteit, I. (2019) Efficient dynamical downscaling of general circulation models using continuous data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 145(724), 3175–3194.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., and Vogel, G. (2018) A description of the Nonhydrostatic regional COSMO-model. Part II: physical parameterizations. *Deutscher Wetterdienst Report COSMO-Model*, 167.

Dumitrache, R., Velea, L., Barbu, C.D., Ibanescu, I. and Lupascu, A. (2011) Preliminary results of COSMO model forecast for the Romanian territory case studies. *Romanian Reports in Physics*, 63(1), 208–219.

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, 15(1), 51–64.

El-Said, A., Brousseau, P., Ridal, M., & Randriamampianina, R. (2021). A new temporally flow-dependent EDA estimating background errors in the new Copernicus European Regional Re-Analysis (CERRA). Earth and Space Science Open Archive ESSOAr.

Fantini, A. (2019) Climate change impact on flood hazard over Italy. Thesis, University of Trieste.

Feng, T., Hu, Z., Tang, S. and Huang, J. (2021) Improvement of an extreme heavy rainfall simulation using nudging assimilation. *Journal of Meteorological Research*, 35(2), 313–328.

Fosser, G.S.K.P.B., Khodayar, S. and Berg, P. (2015) Benefit of convection permitting climate model simulations in the representation of convective precipitation. *Climate Dynamics*, 44(1–2), 45–60.

Gallus, W.A., Jr. (2002) Impact of verification grid-box size on warm-season QPF skill measures. *Weather and Forecasting*, 17(6), 1296–1302.

Giorgi, F. (1990) Simulation of regional climate using a limited area model nested in a general circulation model. *Journal of Climate*, 3(9), 941–963.

Gleeson, E., Whelan, E. and Hanley, J. (2017) Met Éireann high resolution reanalysis for Ireland. *Advances in Science and Research*, 14, 49–61.

GLOBE Task Team and others (Hastings, David A., Paula K. Dunbar, Gerald M. Elphingstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, John Payne, Nevin A. Bryant, Thomas L. Logan, J.-P. Muller, Gunter Schreier, and John S. MacDonald), eds., 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80305–3328, USA. http://www.ngdc.noaa.gov/mgg/topo/globe.html [Accessed 10th February 2023].

Grazzini, F., Craig, G.C., Keil, C., Antolini, G. and Pavan, V. (2020a) Extreme precipitation events over northern Italy. Part I: a systematic classification with machine-learning techniques. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 69–85.

Grazzini, F., Fragkoulidis, G., Pavan, V. and Antolini, G. (2020b) The 1994 Piedmont flood: an archetype of extreme precipitation events in northern Italy. *Bulletin of Atmospheric Science and Technology*, 1, 283–295.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.N. (2020) The ERA5 global reanalysis.

Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049.

Hochman, A., Mercogliano, P., Alpert, P., Saaroni, H. and Bucchignani, E. (2018) High-resolution projection of climate change and extremity over Israel using COSMO-CLM. *International Journal of Climatology*, 38(14), 5095–5106.

Houze, R.A., Jr. (2014) *Cloud Dynamics*. Cambridge: Academic Press.

Hu, G. and Franzke, C.L. (2020) Evaluation of daily precipitation extremes in reanalysis and gridded observation-based data sets over Germany. *Geophysical Research Letters*, 47(18), e2020GL089624.

Isotta, F.A., Vogel, R. and Frei, C. (2015) Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region. *Meteorologische Zeitschrift*, 24, 15–37.

Italian Civil Protection Department and CIMA Research Foundation. (2014) The Dewetra platform: a multi-perspective architecture for risk management during emergencies. In: *Information Systems for Crisis Response and Management in Mediterranean Countries: First International Conference, ISCRAM-Med 2014*. Toulouse, France, October 15-17, 2014: Springer International Publishing, pp. 165–177.

Jacobsen, I. (1982) A new economic method for the computation of the surface temperature in numerical models. *Beitraege zur Physik der Atmosphaere*, 55, 128–141.

Jansa, A., Genoves, A. and Garcia-Moya, J.A. (2000) Western Mediterranean cyclones and heavy rain. Part 1: numerical experiment concerning the Piedmont flood case. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, 7(4), 323–333.

Jansa, A., Genoves, A., Picornell, M.A., Campins, J., Riosalido, R. and Carretero, O. (2001) Western Mediterranean cyclones and heavy rain. Part 2: statistical approach. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, 8(1), 43–56.

Jermey, P.M. and Renshaw, R.J. (2016) Precipitation representation over a two-year period in regional reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1300–1310.

Khairoutdinov, M. and Randall, D. (2006) High-resolution simulation of shallow-to-deep convection transition over land. *Journal of the Atmospheric Sciences*, 63(12), 3421–3436.

Kirshbaum, D.J., Adler, B., Kalthoff, N., Barthlott, C. and Serafin, S. (2018) Moist orographic convection: physical mechanisms and links to surface-exchange processes. *Atmosphere*, 9(3), 80.

Klasa, C., Arpagaus, M., Walser, A. and Wernli, H. (2018) An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. *Quarterly Journal of the Royal Meteorological Society*, 144(712), 744–764.

Klein Tank, A.M.G. (2010) EURO4M: monitoring weather and climate extremes in Europe. EMS Annual Meeting Abstracts, Vol. 7, EMS2010-202, 10th EMS/8th ECAC. https://meetingorganizer.copernicus.org/EMS2010/EMS2010-202.pdf. [Accessed 10th February 2023].

Køltzow, M., Schyberg, H., Støylen, E., and Yang, X. (2022) Value of the Copernicus Arctic Regional Reanalysis (CARRA) in representing near-surface temperature and wind speed in the north-east European Arctic. *Polar Research*, 41.

Krichak, S.O., Barkan, J., Breitgand, J.S., Gualdi, S. and Feldstein, S.B. (2015) The role of the export of tropical moisture into midlatitudes for extreme precipitation events in the Mediterranean region. *Theoretical and Applied Climatology*, 121(3), 499–515.

Lanciani, A., Mariani, S., Casaioli, M., Accadia, C. and Tartaglione, N. (2008) A multiscale approach for precipitation verification applied to the FORALPS case studies. *Advances in Geosciences*, 16, 3–9.

Lewis, H., Mittermaier, M., Mylne, K., Norman, K., Scaife, A., Neal, R., Pierce, C., Harrison, D., Jewell, S., Kendon, M., Saunders, R., Brunet, G., Golding, B., Kitchen, M., Davies, P. and Pilling, C. (2015) From months to minutes–exploring the value of high-resolution rainfall observation and prediction during the UK winter storms of 2013/2014. *Meteorological Applications*, 22(1), 90–104.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P. and Loumagne, C. (2014) When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, 18(2), 575–594.

Longobardi, A., Buttafuoco, G., Caloiero, T. and Coscarelli, R. (2016) Spatial and temporal distribution of precipitation in a Mediterranean area (southern Italy). *Environmental Earth Sciences*, 75(3), 189.

Lott, F. and Miller, M.J. (1997) A new subgrid-scale orographic drag parametrization: its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, 123(537), 101–127.

Marsigli, C., Boccanera, F., Montani, A. and Paccagnella, T. (2005) The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. *Nonlinear Processes in Geophysics*, 12(4), 527–536.

Marsigli, C., Ebert, E., Ashrit, R., Casati, B., Chen, J., Coelho, C.A., Dorninger, M., Gilleland, E., Haiden, T., Landman, S., and Mittermaier, M. (2021) Observations for high-impact weather and their use in verification. *Natural Hazards and Earth System Sciences*, 21(4), 1297–1312.

Marsigli, C., Montani, A. and Paccagnella, T. (2008) A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorological Applications: A Journal of Forecasting, Practical Applications, Training Techniques and Modelling*, 15(1), 125–143.

Mayaux, P., Eva, H., Gallego, J., Strahler, A.H., Herold, M., Agrawal, S., Naumov, S., de Miranda, E.E., di Bella, C.M., Ordoyne, C., Kopin, Y. and Roy, P.S. (2006) Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), 1728–1739.

Mironov, D.V. (2008) *Parameterization of Lakes in Numerical Weather Prediction: Description of a Lake Model*. Offenbach, Germany: DWD, p. 41.

Murphy, A.H. (1993) What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8(2), 281–293.

Napoli, A., Crespi, A., Ragone, F., Maugeri, M. and Pasquero, C. (2019) Variability of orographic enhancement of precipitation in the Alpine region. *Scientific Reports*, 9(1), 13352.

Niedda, M., Amponsah, W., Marchi, L., Zoccatelli, D., Marra, F., Crema, S., Pirastru, M., Marrosu, R., and Borga, M. (2014) Il ciclone Cleopatra del 18 Novembre 2013 in Sardegna: analisi e modellazione dell'evento di piena. *Quaderni di Idronomia Montana*, 32(1), 47–58.

Nieto, R., Gimeno, L., de La Torre, L., Ribera, P., Gallego, D., García-Herrera, R., García, J.A., Nuñez, M., Redaño, A. and Lorente, J. (2005) Climatological features of cutoff low systems in the northern hemisphere. *Journal of Climate*, 18(16), 3085–3103.

Pal, S., Chang, H.I., Castro, C.L. and Dominguez, F. (2019) Credibility of convection-permitting modeling to improve seasonal precipitation forecasting in the southwestern United States. *Frontiers in Earth Science*, 7, 11.

Pavan, V., Antolini, G., Barbiero, R., Berni, N., Brunier, F., Cacciamani, C., Cagnati, A., Cazzuli, O., Cicogna, A., De Luigi, C. and Di Carlo, E. (2019) High resolution climate precipitation analysis for north-Central Italy, 1961–2015. *Climate Dynamics*, 52(5–6), 3435–3453.

Pichelli, E., Coppola, E., Sobolowski, S., Ban, N., Giorgi, F., Stocchi, P., Alias, A., Belušić, D., Berthou, S., Caillaud, C., Cardoso, R.M., Chan, S., Christensen, O.B., Dobler, A., de Vries, H., Goergen, K., Kendon, E.J., Keuler, K., Lenderink, G., Lorenz, T., Mishra, A.N., Panitz, H.J., Schär, C., Soares, P.M.M., Truhetz, H. and Vergara-Temprado, J. (2021) The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation. *Climate Dynamics*, 56(11), 3581–3602.

Porcù, F., Aragão, L., 2019, OPEn-air laboRAtories for nature baseD solUtions to manage hydro-meteo risks: data record on extreme events by OAL and by hazard. Deliverable 4.2, Technical Report.

Prein, A.F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N.K., Keuler, K. and Georgievski, G. (2013) Added value of convection permitting seasonal simulations. *Climate Dynamics*, 41(9–10), 2655–2677.

Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Lipzig, N.P.M. and Leung, R. (2015) A review on regional convection-permitting climate modeling: demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53(2), 323–361.

Qiu, Y., Ichiba, A., Paz, I.D.S.R., Chen, F., Versini, P.A., Schertzer, D. and Tchiguirinskaia, I. (2019) Evaluation of low impact development and nature-based solutions for stormwater management: a fully distributed modelling approach. *Hydrology and Earth System Sciences Discussions*, 1–27. https://doi.org/10.5194/hess-2019-347

Raffa, M., Reder, A., Marras, G.F., Mancini, M., Scipione, G., Santini, M. and Mercogliano, P. (2021) VHR-REA_IT dataset: very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM. *Data*, 6(8), 88.

Raschendorfer, M. (2001) The new turbulence parameterization of LM. *COSMO Newsletter*, 1, 89–97.

Rebora, N., Molini, L., Casella, E., Comellas, A., Fiori, E., Pignone, F., Siccardi, F., Silvestro, F., Tanelli, S. and Parodi, A. (2013) Extreme rainfall in the Mediterranean: what can we learn from observations? *Journal of Hydrometeorology*, 14(3), 906–922.

Reder, A., Raffa, M., Padulano, R., Rianna, G. and Mercogliano, P. (2022) Characterizing extreme values of precipitation at very high resolution: an experiment over twenty European cities. *Weather and Climate Extremes*, 35, 100407.

Riishojgaard, L.P. (2020) Impacts of COVID-19 restrictions on observations and monitoring. *Bulletin-World Meteorological Organization (WMO)*, 69(2), 16–19.

Ritter, B. and Geleyn, J.F. (1992) A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly Weather Review*, 120(2), 303–325.

Rivin, G.S., Rozinkina, I.A., Vil'fand, R.M., Alferov, D.Y., Astakhova, E.D., Blinov, D.V., Bundel, A.Y., Kazakova, E.V., Kirsanov, A.A., Nikitin, M.A. and Perov, V.L. (2015) The COSMO-Ru system of nonhydrostatic mesoscale short-range weather forecasting of the Hydrometcenter of Russia: the second stage of implementation and development. *Russian Meteorology and Hydrology*, 40(6), 400–410.

Rivoire, P., Martius, O. and Naveau, P. (2021) A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. *Earth and Space Science*, 8(4), e2020EA001633.

Roberts, N.M. and Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136(1), 78–97.

Roebber, P.J. (2009) Visualizing multiple measures of forecast quality. *Weather and Forecasting*, 24(2), 601–608.

Ruangpan, L., Vojinovic, Z., Sabatino, S.D., Leo, L.S., Capobianco, V., Oen, A.M., McClain, M.E., and Lopez-Gunn, E. (2020) Nature-based solutions for hydro-meteorological risk reduction: a state-of-the-art review of the research area. *Natural Hazards and Earth System Sciences*, 20(1), 243–270.

Sahani, J., Kumar, P., Debele, S., Spyrou, C., Loupis, M., Aragão, L., Porcù, F., Shah, M.A.R. and di Sabatino, S. (2019) Hydro-meteorological risk assessment methods and management by nature-based solutions. *Science of the Total Environment*, 696, 133936.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried, K., Panosetti, D., Rüdisühli, S., Schlemmer, L., Schulthess, T.C., Sprenger, M., Ubbiali, S., and Wernli, H. (2020) Kilometer-scale climate models: prospects and challenges. *Bulletin of the American Meteorological Society*, 101(5), E567–E587.

Schättler, U., Doms, G. and Schraff, C. (2018) *A Description of the Nonhydrostatic Regional COSMO-Model - Part VII: User's Guide*. Offenbach: Consortium for small scale modeling, Deutscher Wetterdienst, p. 195.

Schraff, C. and Hess, R. (2013) A Description of the Nonhydrostatic Regional COSMO-model - Part III: Data Assimilation. In: *User's Guide, Consortium For Small Scale Modeling*. Offenbach: Deutscher Wetterdienst.

Shah, M.A.R., Renaud, F.G., Anderson, C.C., Wild, A., Domeneghetti, A., Polderman, A., Votsis, A., Pulvirenti, B., Basu, B., Thomson, C., Panga, D., Pouta, E., Toth, E., Pilla, F., Sahani, J., Ommer, J., el Zohbi, J., Munro, K., Stefanopoulou, M., Loupis, M., Pangas, N., Kumar, P., Debele, S., Preuschmann, S. and Zixuan, W. (2020) A review of hydro-meteorological hazard, vulnerability, and risk assessment frameworks and indicators in the context of nature-based solutions. *International Journal of Disaster Risk Reduction*, 50, 101728.

Simon, T., Wang, D., Hense, A., Simmer, C. and Ohlwein, C. (2013) Generation and transfer of internal variability in a regional climate model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 22485.

Sommeria, G. and Deardorff, J.W. (1977) Subgrid-scale condensation in models of nonprecipitating clouds. *Journal of Atmospheric Sciences*, 34(2), 344–355.

Spekkers, M., Rözer, V., Thieken, A., ten Veldhuis, M.C. and Kreibich, H. (2017) A comparative survey of the impacts of extreme rainfall in two international case studies. *Natural Hazards and Earth System Sciences*, 17(8), 1337–1355.

Starosta, K. and Wyszogrodzki, A. (2016) Assessment of model generated wind energy potential in Poland. *COSMO News Letter*, 16, 16–24.

Su, C.H., Eizenberg, N., Steinle, P., Jakob, D., Fox-Hughes, P., White, C.J., Rennie, S., Franklin, C., Dharssi, I. and Zhu, H. (2019) BARRA v1. 0: the Bureau of Meteorology atmospheric high-resolution regional reanalysis for Australia. *Geoscientific Model Development*, 12(5), 2049–2068.

Teixeira, J., Stevens, B., Bretherton, C.S., Cederwall, R., Doyle, J.D., Golaz, J.C., Holstslag, A.A.M., Klein, S.A., Lundquist, J.K., Randall, D.A., Siebesma, A.P., and Soares, P.M.M. (2008) Parameterization of the atmospheric boundary layer: a view from just above the inversion. *Bulletin of the American Meteorological Society*, 89(4), 453–458.

Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117(8), 1779–1800.

Unden, P., Renshaw, R., Bazile, E., Brunet, M., Kaiser-weiss, A. and Klein Tank, A. (2016) UERRA-uncertainties in ensembles of regional reanalyses. In: *EGU General Assembly Conference Abstracts*, Vienna, Austria: EGU, p. EPSC2016-15450.

Wahl, S., Bollmeyer, C., Crewell, S., Figura, C., Friederichs, P., Hense, A., Keller, J.D. and Ohlwein, C. (2017) A novel convective-scale regional reanalysis COSMO-REA2: improving the representation of precipitation. *Meteorologische Zeitschrift*, 26(4), 345–361.

Weisman, M.L., Skamarock, W.C. and Klemp, J.B. (1997) The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, 125(4), 527–548.

Weusthoff, T., Ament, F., Arpagaus, M. and Rotach, M.W. (2010) Assessing the benefits of convection-permitting models by neighborhood verification: Examples from MAP D-PHASE. *Monthly Weather Review*, 138(9), 3418–3433.

Weygandt, S.S., Loughe, A.F., Benjamin, S.G. and Mahoney, J.L. (2004) Scale sensitivities in model precipitation skill scores during IHOP. *22nd Conf. On Severe Local Storms*. Hyannis, MA: American Meteorological Society: Vol. 16.

Wilks, D.S. (2019) *Statistical Methods in the Atmospheric Sciences*, 4th edition. Amsterdam, The Netherlands: Elsevier Science.

Zhang, Q., Pan, Y., Wang, S., Xu, J. and Tang, J. (2017) High-resolution regional reanalysis in China: evaluation of 1 year period experiments. *Journal of Geophysical Research: Atmospheres*, 122(20), 10801–10819.

How to cite this article: Giordani, A., Cerenzia, I.M.L., Paccagnella, T. & Di Sabatino, S. (2023) SPHERA, a new convection-permitting regional reanalysis over Italy: Improving the description of heavy rainfall. *Quarterly Journal of the Royal Meteorological Society*, 1–28. Available from: https://doi.org/10.1002/qj.4428

APPENDIX A - SKILL SCORES

This section summarizes the statistical methods used to assess the performance of the reanalysis systems. For a more detailed description, the reader is referred to Wilks (2019). These methods calculate verification indices for categorical forecasts, that is, forecasts distinguishing only between events and non-events (yes/no forecasts) related to the occurrence of exceedance of certain precipitation thresholds. For each threshold, a 2×2 contingency table (see Table A1) is built from the joint distributions of forecasts (i.e. reanalysis, in our case) and observations, determining the frequency of forecast/observation pairs falling within each of the four possible categories (a, b, c and d in Table A1).

From the results of the contingency tables, it is possible to calculate a variety of scalar attributes assessing different aspects of the simulations under examination. The following reports the indices adopted in the present work using the terminology defined in Table A1:

- POD, or hit rate: describes the fraction of observed "yes" that has been correctly forecasted, ranges between 0 (i.e. worst possible forecast: not even one observed event has been forecasted) and 1 (i.e. perfect forecast: all the observed events have been correctly simulated). It is defined as

$$\mathrm{POD} = \frac{a}{a+c}$$

- FAR: describes the fraction of the simulated "yes" events that actually did not occur, ranges between 0 (i.e. perfect forecast: the simulations did produce not even one false detection) and 1 (i.e. worst possible forecast: all the simulations produced false event detections). It can be expressed as the SR through the relation $\mathrm{SR} = 1 - \mathrm{FAR}$, which is alternatively used (e.g. in the abscissa of the performance diagram). It is defined as

$$\mathrm{FAR} = \frac{b}{a+b}$$

- TS or Critical Success Index (CSI): measures the fraction of observed and/or forecast events that were correctly predicted, ranges between 0 (i.e. worst possible forecast: not even one forecast was correctly simulated) and 1 (i.e. perfect forecast: all the forecasts were correctly predicted without missing any observed event or producing any false alarm). TS can be expressed as a nonlinear combination of POD and FAR, and this property is exploited in the performance diagram where POD is plotted against SR. It is defined as

$$\mathrm{TS} = \frac{a}{a+b+c} = \left(\frac{1}{POD} + \frac{1}{1-FAR} - 1\right)^{-1}$$

- Frequency Bias (BIAS): evaluates the deviation of the simulated from the observed frequency of event occurrences, and it is defined as their ratio. With the

TABLE A1 2 × 2 contingency table representing the relationship between forecast/observation pairs for the dichotomous non-probabilistic verification related to the occurrence of an event

		Observed		
		Yes	**No**	
Forecast	Yes	a	b	a + b
		Hits	False alarms	Forecast yes
	No	c	d	c + d
		Misses	Correct negatives	Forecast no
		a + c	b + d	a + b + c + d
		Observed yes	Observed no	Total

frequency bias, it is possible to evaluate the tendency of the forecasting system to underestimate (BIAS < 1) or overestimate (BIAS > 1) the number of observed events. It ranges from 0 to infinity, and the score for the perfect forecast is 1 (i.e. equal frequency of occurrence between forecasts and observations). It is defined as

$$\text{BIAS} = \frac{a+b}{a+c}$$

- A way to summarize all four indices is with the performance diagram (e.g. Figure 6), which exploits the geometrical relationships between the scalar attributes, permitting a more comprehensive evaluation of the performance. A good forecast would lie in the top-right corner of the diagram, where POD, SR, TS and BIAS approach unity. At the same time, deviations in a particular direction would indicate relative differences in POD and SR and, thus, in TS and BIAS. The variability in the scores from sampling is estimated with a bootstrap resampling technique of 1,000 new data samples and shown in the diagram as "cross-hairs". The reader is referred to Roebber (2009) for a detailed description of the performance diagram.