https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

A nested time-window framework for evaluating sub-daily precipitation extremes in two convection permitting model datasets

Oliver S Carlo[1,2], Giusy Fedele[1], Paolo Stocchi[3], Sandro Calmanti[4], Mario Raffa[1], Paola Mercogliano[1], Piero Lionello[2]

[1] CMCC Foundation – Euro-Mediterranean Center on Climate Change, Caserta, Italy
[2] DiSTeBA – Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, University of Salento, Lecce, Italy
[3] ISAC – Istituto di Scienze Atmosferiche e Clima, CNR, Bologna, Italy
[4] ENEA – Italian National Agency for New Technologies, Rome, Italy

Correspondence to: Oliver S. Carlo (oliver.carlo@cmcc.it)

Abstract. High-resolution convection-permitting models (CPMs) provide vital datasets for studying localized extreme precipitation and associated multi-hazard risks like landslides and floods. Yet validating CPM datasets against observations remains a challenge. This study evaluates two sub-daily CPM datasets, VHR-REA_IT and LAM-HIND, against observations over a 10-year period across two Italian regions: Emilia-Romagna and Campania. Extreme precipitation is defined using the 95th percentile of wet-event intensities across accumulation durations of 1, 2, 4, 12, and 24 hours.

Conventional verification metrics only partially assess dataset performance and hide differences in the temporal organization of precipitation extremes. Therefore, a new nested time-window framework is presented to better capture the association between short- and long-duration time window extremes (STWEs and LTWEs, respectively). Observations show that STWEs and LTWEs are only partially associated. Depending on region and season, around 7–15% of observed 1-hour STWEs are embedded within 24-hour LTWEs. CPMs reproduce this cross-timescale organisation substantially better than ERA5, yet are overestimated, indicating excess temporal persistence of simulated heavy-rainfall. Consequently, LTWEs are more dependable than STWEs, as CPMs distribute short extreme precipitation bursts into longer durations with persistently higher volumes.

Differences between the CPM datasets are small. LAM-HIND exhibits subtly higher detection and event-overlap scores but a higher false-alarm rate and stronger cross-timescale coupling overestimation. VHR-REA_IT more closely reproduces the observed temporal association between short- and long-duration extremes. These differences may stem from the 'double penalty' effect, choice of an intermediate nesting strategy, and sensitivity in resolving sub-grid turbulent fluxes. However, they are both substantially improved over coarse products like ERA5 for sub-daily extreme precipitation.

1 Introduction

Precipitation is a fundamental driver of the Earth's water cycle, but precipitation phenomena are intermittent and highly variable in space (Cavalleri et al., 2024; Schleiss et al., 2011; Sun et al., 2018). The global rise in extreme weather events includes an increase in extreme precipitation events, causing significant damage to life, infrastructure, and property. For instance, in Italy, the damage suffered from extreme events in 2025 amounted to €11.9 billion, which is projected to rise to €34.2 billion by 2029, with the most frequent events linked to heavy rainfall—causing flooding, landslides, and river overflows (ANSA English Desk, 2025; Formetta et al., 2024; Il Mattino, 2025).

To understand such extreme precipitation events, stakeholders such as governmental agencies, local municipalities, power supply grid operators, policymakers, scientists, and infrastructure stakeholders rely on high-quality climate datasets. These datasets are useful for studying past climate variability and trends, reconstructing specific test cases, and investigating impacts on resources and damage to life and property (Mazzoglio et al., 2025; Viterbo et al., 2024).

Climate reanalysis products are among the most widely used datasets in the geophysical sciences. They combine past observations with models to generate consistent time series of multiple climate variables, providing a comprehensive description of the observed climate as it has evolved during recent decades on 3D grids at sub-daily intervals (Climate reanalysis, 2026). The assimilation of observational data from multiple local sources helps reproduce past weather and climate conditions in the most accurate way possible (Cavalleri et al., 2024; Kalnay et al., 1996; Viterbo et al., 2024).

However, global reanalysis products such as ERA5 or ERA5-Land (over land surfaces only) are produced at relatively coarse spatial resolutions – 31 km and 9 km grid resolutions respectively (ERA5-Land hourly data from 1950 to present; ERA5 hourly data on single levels from 1940 to present; Hersbach et al., 2020; Muñoz-Sabater et al., 2021; Soci et al., 2024). However, they are not optimal for determining localized extreme events. For instance, coarse resolution models tend to spatially smooth out precipitation extremes, and important processes like deep convection or orographically triggered convection that cannot be resolved at large grid scales have to be parametrized instead, which is a major source of model errors and uncertainty (Ban et al., 2021; Kendon et al., 2017; Prein et al., 2015; Stocchi et al., 2022). Therefore, stakeholders depend on high-resolution downscaled products, at less than 4 km resolution, which are often produced by regional research agencies (Cavalleri et al., 2024).

Downscaling techniques include the use of dynamic downscaling, which produces regional climate model (RCM) data. This technique generally uses global climate model (GCM) data or global reanalysis like ERA5 to force the boundary conditions of the RCM, combining core primitive equations (governing fluid motion, mass conservation and heat transfer) with surface characteristics of the region of interest, such as terrain and land-use (What is Statistical and Dynamical downscaling?; Le Roux et al., 2018). These higher resolutions allow for convection-permitting models (CPMs) (< 4 km), supporting the characterization of extreme rainfall events, as they simulate a more direct and accurate description of convective processes rather than relying on parametrized convection used in low-resolution models (Prein et al., 2015).

However, dynamic modelling parameters and equations are a source of uncertainty, which leads to biases in the outputs. Even in the case of data assimilation coupled with a dynamic model, observational data is sparsely located and prone to uncertainties, which also cause bias (Prein and Gobiet, 2017). These biases are systematic deviations compared to actual weather conditions, and are often more pronounced when studying extreme events due to the low probability of their occurrence (Kendon et al., 2012).

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

Studies typically compare the mean annual or seasonal bias between different variables using traditional verification metrics to assess the quality of a model (Capecchi et al., 2023; Cavalleri et al., 2024). However, fewer studies focus on extreme events, particularly the bias between model and observation datasets at sub-daily timescales, largely due to the scarcity of data for these rare but high impact events (Mazzoglio et al., 2025; Trentini et al., 2023). Furthermore, Kendon et al. (2012) have particularly noted that climate models tend to exhibit excessive rainfall persistence, with insufficient representation of short-duration high peak intensity events. Yet high-risk, rainfall-induced landslides primarily occur at or within 6 hours of the peak rainfall intensity, and at or near the geographical location where the rainfall intensity was largest (Roccati et al., 2018). Hardly any studies are dedicated to evaluating the temporal associations across sub-daily time scales, which are vital for multi-hazard risk assessment.

Given the advances in high-performance computing offering a performance 250,000 times higher in a span of 20 years (Prein et al., 2015) and the rise of CPM simulation outputs, this study aims to assess the discrepancies and biases between CPM datasets, ERA5 reanalysis, and observations, for extreme precipitation events across the sub-daily scale. The period of study spans 10 years, from the start of 2014 until the end of 2023, providing a reasonably long period of study and sample size for analysis. Section 2 presents a description of the methodology used, including the data types and the study areas. Section 3 discusses the results from the comparisons between the precipitation variables.

The two chosen CPM datasets are produced for Italy. Therefore, two regions of Italy, Campania, and Emilia-Romagna, were selected for evaluation purposes, because their regional monitoring networks provide sufficiently extensive high-frequency rain-gauge records for sub-daily precipitation validation. But a detailed comparison between the regions of Campania and Emilia-Romagna is beyond the scope of this study for two primary reasons. First, as will be subsequently discussed in Section 2.2, the observational datasets are managed by two different regional agencies, each applying its own unique data processing and quality control standards (Bongiovanni et al., 2025; Desiato et al., 2011). Because our study does not standardize or account for these administrative differences, any direct cross-regional comparison would not be sufficiently thorough. Second, the two regions are geographically separated by the Apennine Mountain range; Emilia-Romagna lies to the north and east (right) side of the Apennines, while Campania lies to the south and west (left) side (see Figure 1). Consequently, they feature distinct local climate zones according to the Köppen Climate Classification (Billi, 2023) with different orographic features, and the study by Mazzoglio et al. (2025) have noted how orography and land-sea contrasts influence CPM biases. Fully accounting for how these factors impact model performance would require a dedicated investigation outside the scope of this paper. The inclusion of these two distinct regions represented by distinct climatic and orographic settings on opposite sides of the Apennines allows the performance of the CPM datasets to be evaluated across contrasting Italian environments, without implying a direct comparison between the two regions.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

2 Data and Methods

2.1 Model data

For this study, three different models were considered: the global ERA5 reanalysis and two high-resolution CPM products. The ERA5 reanalysis is not the primary focus of this study for reasons given in Section 1. However, it is included, first because it drives the initial and boundary conditions for the high-resolution CPMs that are the focus of this study, and second because it highlights the added value of the CPM products over ERA5.

ERA5 reanalysis, released by the European Centre for Medium-Range Weather Forecasts (ECMWF), represents the most recent and thus most reliable baseline for investigating past climate patterns (Hersbach et al., 2020). It provides a detailed hourly historical record of global weather, soil, and ocean conditions from 1950 to the present, at a resolution of 0.25° (≃31 km) (Hersbach et al., 2020; Loprieno et al., 2026).

The two high-resolution CPM data considered are the *Very High-Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy* (VHR-REA_IT) (Raffa et al., 2021b; Reder et al., 2022) and the *LaMMA produced Atmospheric Hindcast Data* (LAM-HIND) (Capecchi et al., 2023).

The downscaled data for VHR-REA_IT was generated by the *Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici* (CMCC Foundation) (https://www.cmcc.it/), one-way nested using the ERA5 reanalysis as forcing conditions, with a horizontal grid spacing of 0.02° (≃2.2 km), over the domain covering the Italian peninsula (Lon = 5° W-20° E; Lat = 5° N-48° N) for the period from January 1981 to December 2024. This data is available to download for free from the CMCC Data Delivery System (https://dds.cmcc.it/). The regional COSMO-CLM model was used with the TERRA-URB module to account for urban parametrizations (Raffa et al., 2021b).

LAM-HIND is an atmospheric hindcast for the period 1979–2019 developed by the *Laboratorio di Monitoraggio e Modellistica Ambientale* (LaMMA) (https://www.lamma.toscana.it/), built using a two-step nesting process – the intermediate *BOlogna Limited Area Model* (BOLAM) is driven by ERA5 reanalysis data, which then further drives *MOdello LOCale in Hybrid coordinates* (MOLOCH) to achieve the high resolution LAM-HIND downscaled data (Capecchi et al., 2023). BOLAM is a primitive equations hydrostatic model with a horizontal grid spacing of 0.07° (≃7.8 km), and convection parameterized using a modification of the Kain–Fritsch convective parameterization. MOLOCH is a non-hydrostatic, fully compressible model with a horizontal grid spacing of 0.0247° (≃2.7 km), in which the microphysical scheme is an upgrade of the parameterization proposed by Drofa and Malguzzi (Capecchi et al., 2023; Drofa and Malguzzi, 2004; Kain, 2004).

Neither of these two data products assimilates local data, and both have been evaluated in previous studies for different variables and performance metrics; thus, the reader is recommended to refer to those studies as well (Capecchi et al., 2023; Cavalleri et al., 2024; Crespi et al., 2024; Giordani et al., 2026; Manco et al., 2025; Mazzoglio et al., 2025; Raffa et al., 2021b).

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

2.2 Observation data

The observational precipitation data for this study was obtained from regional networks operating within the Italian national territory: the high-frequency rain-gauge network of the 'Agenzia Prevenzione Ambiente Energia Emilia-Romagna' (ARPAE) in Emilia-Romagna and the 'Centro Funzionale Multirischi della Protezione Civile Regione Campania' in Campania.

Quality checks of these stations were adapted using procedures similar to those given in previous studies (Bonanno et al., 2019; Cavalleri et al., 2024). Stations with missing long-term data (longer than 10 days) and rainfall exceeding 500 mm/day were excluded from the study. Occasionally, a weather station would lack data for a few days due to routine maintenance, which is generally required to be carried out every three to four months to clean accumulated dirt and debris from the funnel and buckets, as well as to ensure that the gauge is level (World Meteorological Organization - WMO, 2008). During these periods – considered up to a maximum of ten consecutive days – days with missing values were conservatively masked out and excluded from both the observational and the model datasets, adhering to maintenance guidelines while maintaining the high quality of the long-term datasets. Precipitation data associated with temperatures lower than 2°C was not included to mitigate any under-catch caused by snow. From an initial pool of 187 and 382 unique weather stations in Campania and Emilia-Romagna respectively, the quality checks left 140 and 89 stations respectively for the 10-year study period (Figure 1). Additionally, post the quality check, most of the stations are located within or near the Apennines, which is beneficial for the analysis (Figure 1).

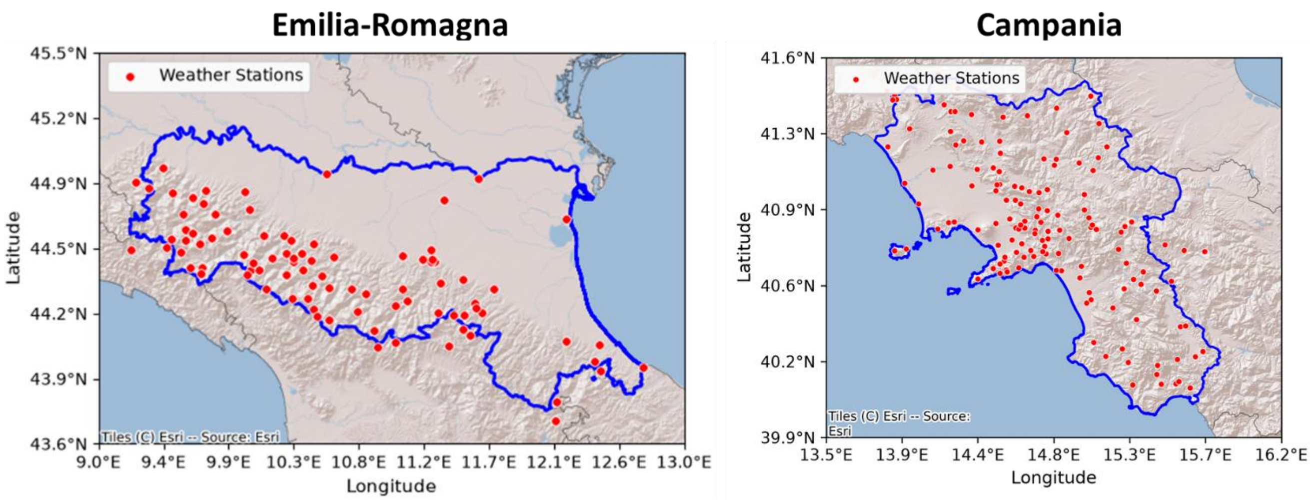


Figure 1 Location of weather stations for Emilia-Romagna (left) and Campania (right) post quality checks. The base map was designed and developed by Esri | Powered by Esri. For more information about this base map, visit https://services.arcgisonline.com/ArcGIS/rest/services/World_Shaded_Relief/MapServer (last access: 29 July 2026).

2.3 Comparison of models against observation

Given that different reanalyses depend on different models and conditions, the goal is to determine how well the models and observations share common characteristics and coincide in time for extreme precipitation. Furthermore, this paper intends to

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

analyse how these discrepancies align at sub-daily temporal resolutions. The hourly timescale was chosen as the smallest common temporal resolution between the reanalysis models and the observational data.

The hourly precipitation data from each of the datasets was extracted based on the geospatial locations of the weather stations. To extract these values, a bilinear interpolation was applied between the cell centres of the 2×2 grids surrounding each observational data point. This approach ensures a smoother transition between precipitation values at neighbouring grid points compared to a nearest-neighbour interpolation, and it is a standard method adopted by several recent studies (Cavalleri et al., 2024; Giordani et al., 2026; Viterbo et al., 2024). No subsequent spatial transformations or scale-decompositions were applied to the data. This treatment aligns with ERA5 guidelines suggesting that gridded reanalysis outputs can be viewed as representing point values at regular intervals (ERA5: What is the spatial reference). However, comparing point-like station data with gridded datasets requires acknowledging the critical distinction between grid spacing and actual model resolution (Cavalleri et al., 2024; Lewis D Grasso, 2000). The length scales at which an atmospheric model truly resolves physical features – also known as its 'effective resolution' – is typically estimated to be at least four times (4Δx) its grid spacing, a factor that is often unknown by users and producers of the model (Cavalleri et al., 2024). In this study, the two high-resolution datasets feature highly comparable horizontal grid spacings: 2.2 km for VHR-REA_IT and 2.7 km for LAM-HIND. Consequently, their respective 'effective resolutions' (not the actual grid spacing) are estimated to differ by roughly 2.2 km (4Δx). While not a direct explanation for the observed variations between the two products, this effective resolution gap represents one possible factor to consider when interpreting how localized, extreme convective events are characterized at these sub-grid scales.

Additionally, most of the comparisons between the datasets draw on the top 95th percentile of the rainfall data over the entire 10-year study period for each station location, considering only wet hours (rainfall values above 0 mm). This allows for a direct comparison of the occurrence probability of extreme weather events between gridded datasets and station observations, assessing their reliability at the sub-daily scale, without the confounding issues of comparing absolute magnitude values.

While the 95th percentile is usually classified as a 'heavy' rather an 'extreme' precipitation event in some climatological studies (Fumière et al., 2020), designating it an 'extreme' event in the context of this study is justified. For instance, sub-daily heavy rainfall can also trigger severe landslides, as they are driven by mesoscale convective systems, squall lines and tropical cyclones (Prein et al., 2015). As noted by Aleotti (2004), the influence of rainfall on landslides differs substantially depending on landslide dimensions, kinematics, material involved, antecedent rainfall, etc., and there can be no critical rainfall threshold applicable to such events. Given this variability, selecting the 95th percentile serves as a robust alternative to a scarcer threshold like the 99th percentile. By including a larger number of such precipitation data points, this choice inherently ensures a sufficient sample size yielding more statistically robust findings, while accounting for the variability in precipitation that also leads to extreme events like landslides.

The sub-daily precipitation data is cumulatively clustered into five time-windows (TWs): 1-hour, 2-hour, 4-hour, 12-hour, and 24-hour. Because of the temporal resolution of the reanalysis outputs, only timescales equal to or longer than 1 hour have been considered. The filtered 'extreme' events vary with the selected TW, as the 95th percentile threshold is applied for each TW

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

and to each weather station. This clustering enables the comparison of extreme events between the different TWs for each dataset and the observation data.

The analysis will first compare the modelled datasets with the observations using performance metrics described in Section 2.4, while further on a novel comparison framework is introduced for comparing extreme events at the sub-daily timescale, by relating extreme events between the sub-daily time windows.

2.4 Performance metrics

Several performance metrics and indices were applied to quantitatively assess the outputs of the models against the observations across the 1-hour to 24-hour time windows. The selected verification metrics follow commonly adopted approaches in the literature for the evaluation of precipitation simulations (Capecchi et al., 2023; Roebber, 2009), and provide complementary information on different aspects of model performance – including correlation, error magnitude, event detection, false alarms and temporal consistency.

Standard verification statistics include the Pearson Correlation Coefficient (*r*) calculated for the whole range of data, alongside root mean squared error *(RMSE)* (in mm of rainfall), normalized root mean squared error *(NRMSE)*, and normalized mean error *(NME)* calculated specifically for the extremes in the data range (Capecchi et al., 2023). Given that the observations consist of sparse data points, a traditional neighbourhood-based spatial fractional skill score (*FSS*) could not be applied (Capecchi et al., 2023). Instead, a temporal fractional skill score *(t-FSS)* has been employed, following the study by Duc, Saito, & Seko (2013). This version utilises a temporal neighbourhood window for each station location to calculate the skill score, rather than a spatial window. This serves to evaluate whether the model performs adequately within a few time windows before and after an observed precipitation event is recorded.

Additionally, performance metrics adapted from Roebber (2009) include the probability of detection *(POD)*, success ratio *(SR)*, and critical success index *(CSI)* calculated for the extremes in the data range. Details on how these verification statistics and performance indices are calculated are provided by Capecchi et al. (2023).

3 Results and Discussions

3.1 Time Profiles and Model Skill

We begin by investigating the monthly aggregated profiles, before moving towards the analysis of the two datasets at the sub-daily timescale emphasised by verification metrics and complementary event overlap statistics. Together, these analyses provide information on both the detection skill and temporal consistency of the two CPM datasets across different aggregation windows.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

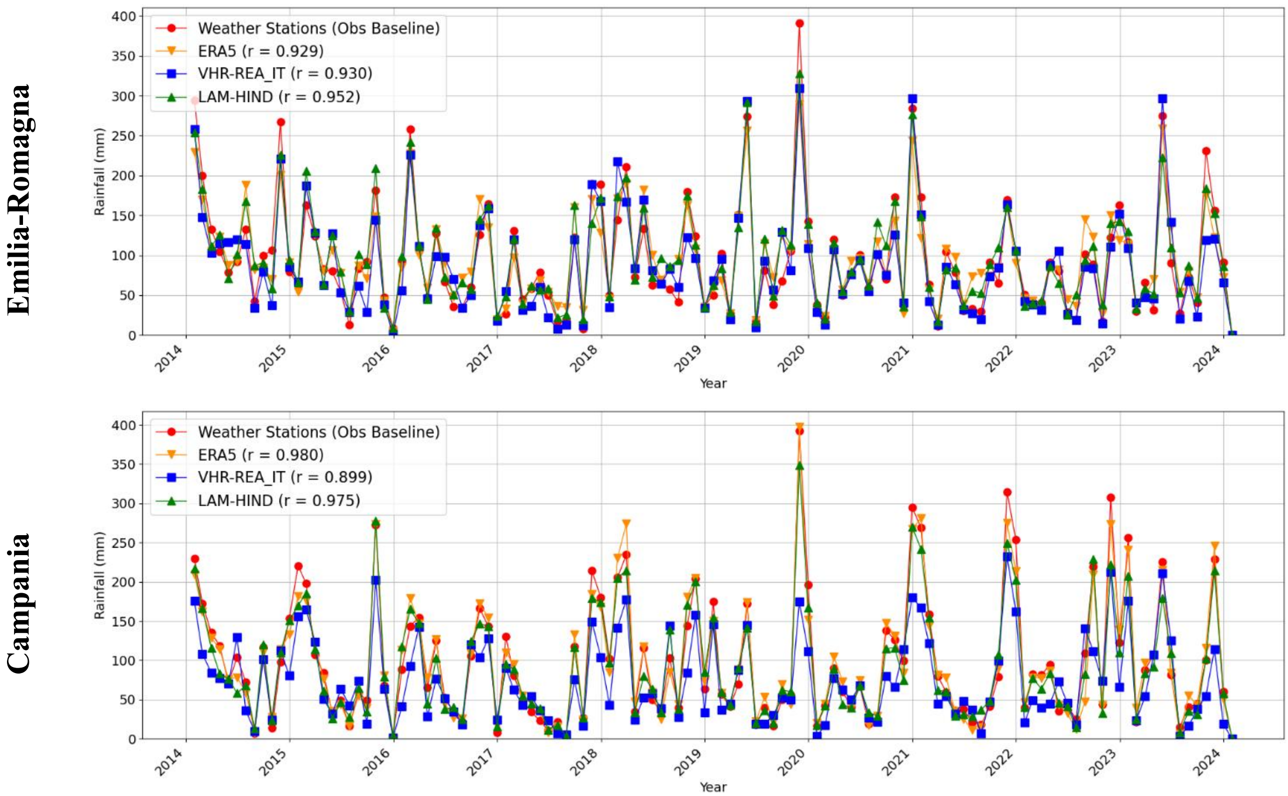


Figure 2 Monthly rainfall timeseries profiles of each model for Emilia-Romagna (top panel) and Campania (bottom panel) across the 10-year study period (2014–2023). The y-axis represents the total monthly rainfall (averaged across all stations), and the x-axis indicates the month/year. Pearson correlation coefficients (r) computed against the weather station observations are provided in the legend for each model.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

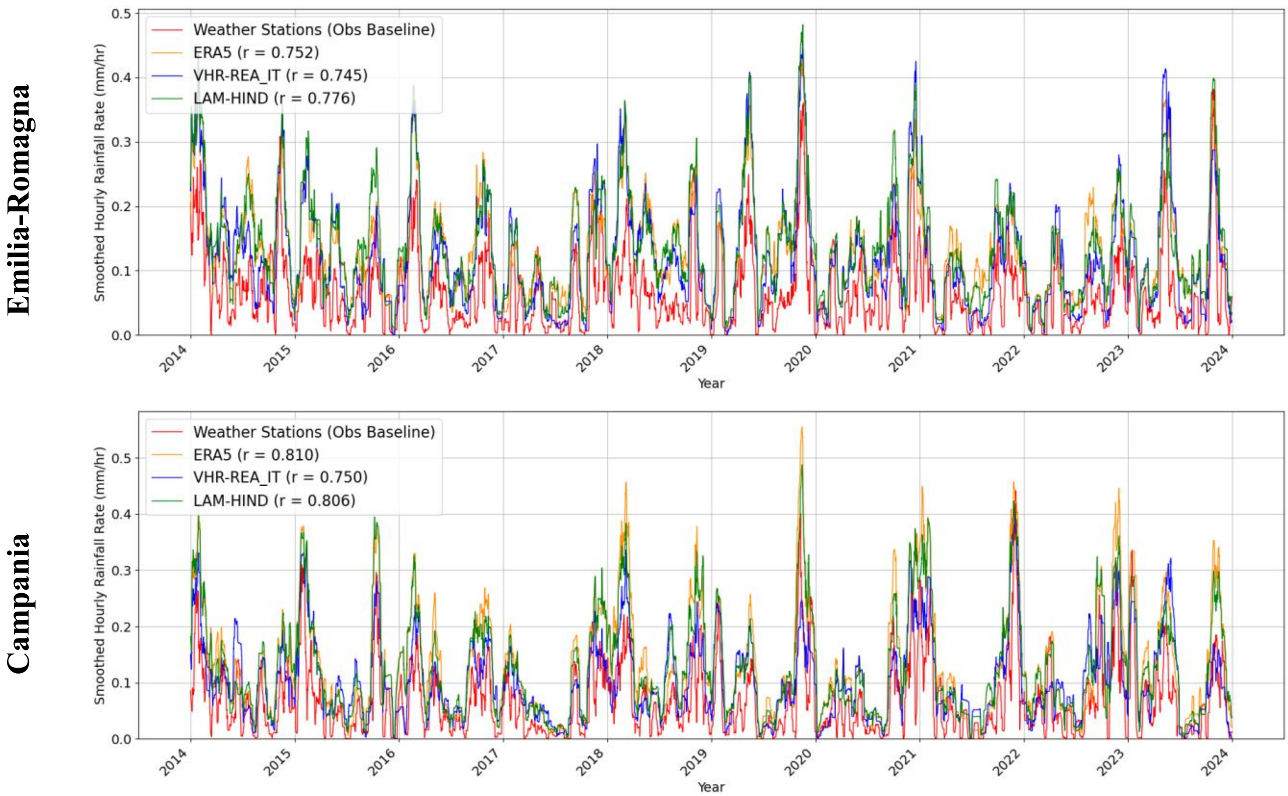


Figure 3 Station-averaged low-pass hourly precipitation profiles for Emilia-Romagna (top panel) and Campania (bottom panel) across the 10-year study period (2014–2023). The y-axis represents the smoothed hourly rainfall rate (mm/hr) computed using a 30-day (720-hour) centred rolling window, and the x-axis indicates the timeline (month/year). Pearson correlation coefficients (r) computed against the weather station observations are provided in the legend for each model.

Figure 2 conveys that at the monthly timescale, both CPM datasets demonstrate comparable cumulative variations with the differences between the two being relatively small, while ERA5 also reproduces the observed temporal variability remarkably well. This indicates that the added value of CPM datasets is not primarily expressed at monthly timescales, but rather emerges at finer temporal resolutions, providing a strong motivation for the analyses that follow.

The lower Pearson correlation coefficients observed for the low pass filtered profiles in Figure 3 reveal that the models temporally cluster rainfall differently, even if the cumulative values over the month converge as shown in Figure 2. Specifically, the hourly low-pass profiles in Figure 3 tend to be systematically elevated over the study period compared to the observed values. This indicates that while an observed extreme short-duration convective event tends to get diluted over a rolling average window, the models tend to distribute these extreme events into persistent, high-volume blocks of rainfall. The models thus perform better for longer cumulative precipitation periods compared to shorter ones, which provides useful context for further interpretations discussed further in Section 3.2.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

	Time Window	r	NME	RMSE (mm)	NRMSE	POD	CSI	SR	t-FSS
	Ideal Score	***1***	***0***	***0***	***0***	***1***	***1***	***1***	***1***
ERA5	*1-hour*	*0.407*	*-0.783*	*8.745*	*1.003*	*0.463*	*0.106*	*0.121*	*0.313*
	2-hour	*0.476*	*-0.723*	*12.475*	*0.907*	*0.512*	*0.135*	*0.155*	*0.382*
	4-hour	*0.563*	*-0.638*	*17.011*	*0.808*	*0.578*	*0.175*	*0.201*	*0.444*
	12-hour	*0.687*	*-0.490*	*25.257*	*0.659*	*0.688*	*0.255*	*0.288*	*0.520*
	24-hour	*0.740*	*-0.417*	*31.854*	*0.588*	*0.686*	*0.309*	*0.359*	*0.582*
VHR-REA_IT	1-hour	**0.267**	-0.754	9.039	1.041	0.201	0.093	**0.145**	**0.385**
	2-hour	**0.343**	-0.686	12.964	0.950	0.251	0.123	**0.192**	**0.448**
	4-hour	0.434	-0.599	17.813	0.859	0.311	0.162	**0.249**	**0.493**
	12-hour	0.585	-0.447	26.344	0.710	0.427	0.244	**0.360**	**0.544**
	24-hour	0.654	-0.393	32.819	0.635	0.471	0.285	**0.416**	**0.571**
LAM-HIND	1-hour	0.258	**-0.751**	8.954	**1.033**	**0.289**	**0.094**	0.121	0.381
	2-hour	0.340	**-0.682**	12.725	**0.934**	**0.348**	**0.124**	0.160	0.447
	4-hour	**0.442**	**-0.596**	17.191	**0.831**	**0.416**	**0.164**	0.210	0.490
	12-hour	**0.607**	**-0.443**	24.599	**0.668**	**0.532**	**0.248**	0.314	0.536
	24-hour	**0.677**	**-0.377**	30.334	**0.593**	**0.560**	**0.288**	0.367	0.560

Table 1 Model skill of extreme precipitation events for both models, evaluated for Emilia-Romagna. Bold cells indicate the better result between the two CPM datasets, while ERA5 is separately marked in italics.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

	Time Window	**r**	**NME**	**RMSE (mm)**	**NRMSE**	**POD**	**CSI**	**SR**	**t-FSS**
	Ideal Score	***1***	***0***	***0***	***0***	***1***	***1***	***1***	***1***
ERA5	*1-hour*	*0.379*	*-0.830*	*10.924*	*1.048*	*0.402*	*0.086*	*0.098*	*0.267*
	2-hour	*0.457*	*-0.763*	*14.749*	*0.957*	*0.455*	*0.117*	*0.136*	*0.342*
	4-hour	*0.556*	*-0.672*	*19.032*	*0.855*	*0.532*	*0.162*	*0.188*	*0.404*
	12-hour	*0.692*	*-0.524*	*26.346*	*0.696*	*0.626*	*0.235*	*0.271*	*0.481*
	24-hour	*0.755*	*-0.428*	*31.165*	*0.599*	*0.657*	*0.295*	*0.347*	*0.562*
VHR-REA_IT	1-hour	**0.205**	-0.839	11.389	1.092	0.124	0.059	**0.102**	**0.319**
	2-hour	**0.278**	-0.775	15.624	1.014	0.165	0.085	**0.147**	**0.368**
	4-hour	0.379	-0.686	20.637	0.928	0.226	0.123	**0.211**	**0.398**
	12-hour	0.527	-0.595	29.941	0.794	0.296	0.169	**0.280**	**0.422**
	24-hour	0.596	-0.532	37.455	0.724	0.346	0.207	**0.335**	**0.468**
LAM-HIND	1-hour	0.230	**-0.801**	11.181	**1.073**	**0.219**	**0.072**	0.097	0.339
	2-hour	0.316	**-0.723**	15.232	**0.991**	**0.286**	**0.104**	0.140	0.399
	4-hour	**0.427**	**-0.630**	19.827	**0.895**	**0.364**	**0.146**	0.195	0.442
	12-hour	**0.584**	**-0.504**	27.854	**0.744**	**0.460**	**0.212**	0.280	0.492
	24-hour	**0.657**	**-0.450**	33.802	**0.659**	**0.495**	**0.248**	0.330	0.544

Table 2 Model skill of extreme precipitation events for both models, evaluated for Campania. Bold cells indicate the better result between the two CPM datasets, while ERA5 is separately marked in italics.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

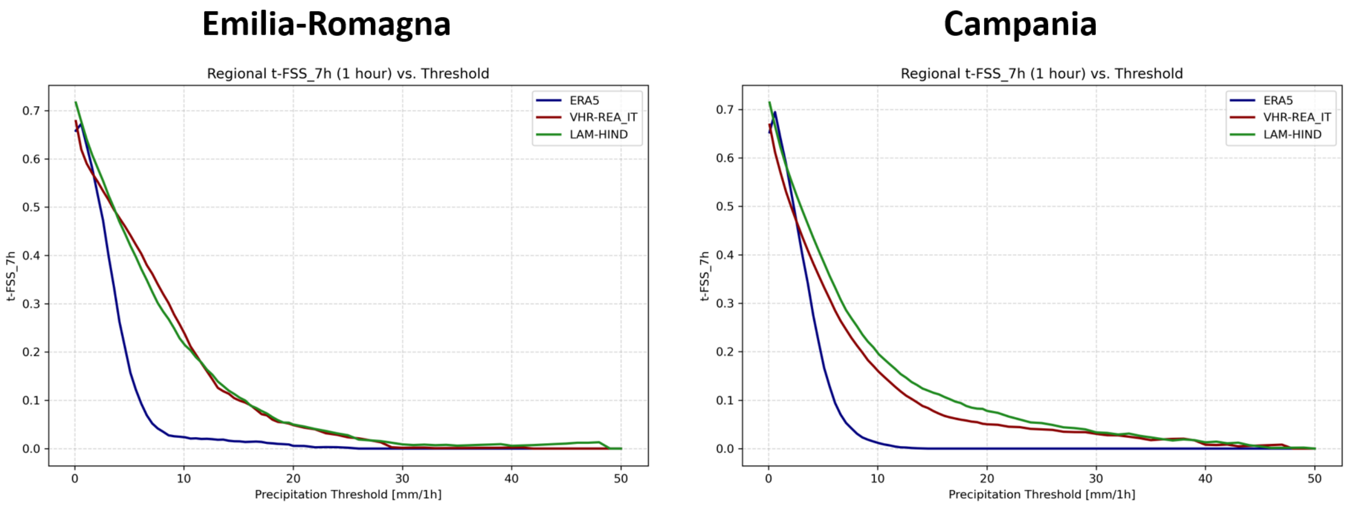


Figure 4 Temporal Fractional Skill Score (t-FSS) as a function of the precipitation threshold for the 1-hour time window. The neighbourhood length is ±3 hours, therefore using a 7-hour rolling window.

From the performance of the CPM datasets shows by the skill scores in Table 1 and Table 2, LAM-HIND scores slightly better across most performance metrics, while VHR-REA_IT maintains a higher Success Ratio (SR). This suggests that LAM-HIND's skill is accompanied by a higher frequency of false alarms, with a tendency to over-forecast 95th-percentile extreme events. One potential hypothesis could be the model configuration; LAM-HIND dataset may have been triggered by more aggressive convection, yielding a higher POD, but also generating false storms that lower the SR (Capecchi et al., 2023). Another factor may be the grid size – the larger grid cells in LAM-HIND (2.7 km) might ensure a better hit with point-to-point metrics at weather station locations, but they also tend to cover nearby areas where no rain was present, leading to a lower SR in a spatial validation context.

However, the scores from Table 1 and Table 2 provide only a narrow picture, as noted from the scores for the ERA5 dataset; by demonstrating skill scores on par with the CPM datasets (and in some cases much better), it is more indicative of the fact that these metrics do not clearly justify the value of CPM modelling, and other methods would be more suited for this purpose. For instance, from Table 1 and Table 2, for Emilia-Romagna the VHR-REA_IT model reveals t-FSS scores on par with LAM-HIND, while for Campania, LAM-HIND t-FSS scores only slightly better. On closer investigation of the hourly t-FSS plotted against precipitation thresholds shown in Figure 4, the scores even-out at high precipitation thresholds (>30 mm/1h) for both regions. Therefore, both CPM datasets behave similarly across moderate-to-high precipitation thresholds, despite the difference in their grid size and model configuration. Most notably, however, both CPM datasets remain clearly separate from ERA5. This suggests the principal added value arising from convection-permitting modelling itself compared to the low resolution ERA5 dataset, whereas the differences between the two CPM datasets are comparatively small and become negligible at higher precipitation thresholds.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

It could be argued that the negligible advantage of LAM-HIND over VHR-REA_IT for most of the skill scores shown in Table 1 and Table 2 may relate to the dynamics of regional model nesting hierarchies. As explored in previous studies, a large resolution jump in a single-nesting configuration – as used in the VHR-REA_IT model, which features a 1:14 jump from the 31 km ERA5 grid spacing forcing a 2.2 km grid model – can result in an underestimation of small-scale transient-eddy amplitudes, also known as the spatial spin-up issue (Lucas-Picher et al., 2021; Matte et al., 2016). This occurs when the domain is too small to allow for the full development of small scales for a given resolution jump between the driving lateral boundary conditions (LBCs) and the RCM grid mesh. Thus, LAM-HIND's double-nesting hierarchy (ERA5 -> BOLAM -> LAM-HIND) could minimize the spatial spin-up constraint, allowing small-scale convective eddies to fully recover their physical amplitudes over the study domain. In fact, for VHR-REA_IT, the study by Raffa *et al.* (Raffa et al., 2021a) noted that while the direct one-step nesting showed better performance compared to a two-step nesting for total precipitation amount, the two-step nesting strategy showed better results for extreme precipitation above the 95th percentile of wet days. While it is not necessarily the case that a direct one-step nesting strategy deteriorates the overall quality of the model (Marsigli et al., 2014; Raffa et al., 2021a), this may shed light on the fact that an intermediate nesting stage used in producing the LAM-HIND dataset, may play a role in the representation of precipitation extremes.

The intermediate nesting strategy comes with the drawback that it weakens the strict boundary control exerted by the driving reanalysis model, as it relies on boundary conditions generated by the intermediate-level model codes. In this case, BOLAM is run using a primitive equation hydrostatic model at an intermediate grid spacing of 7.8 km (Capecchi et al., 2023), which operates at the entry point of the convective 'grey zone' (Jeevanjee, 2017). In this region, hydrostatic solvers tend to overestimate vertical velocities (Jeevanjee, 2017), which can amplify vertical parcel transport and generate more active moisture convergence. When these higher-energy fields are subsequently passed on as lateral boundary conditions to the finer non-hydrostatic MOLOCH mode producing the LAM-HIND dataset, they may increase the probability of extreme convective precipitation, thus explaining the slightly higher skill scores of the LAM-HIND model. However, this is only a complementary hypothesis, and further studies are needed to assess the contribution of this phenomenon in isolation, in context with extreme precipitation events.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

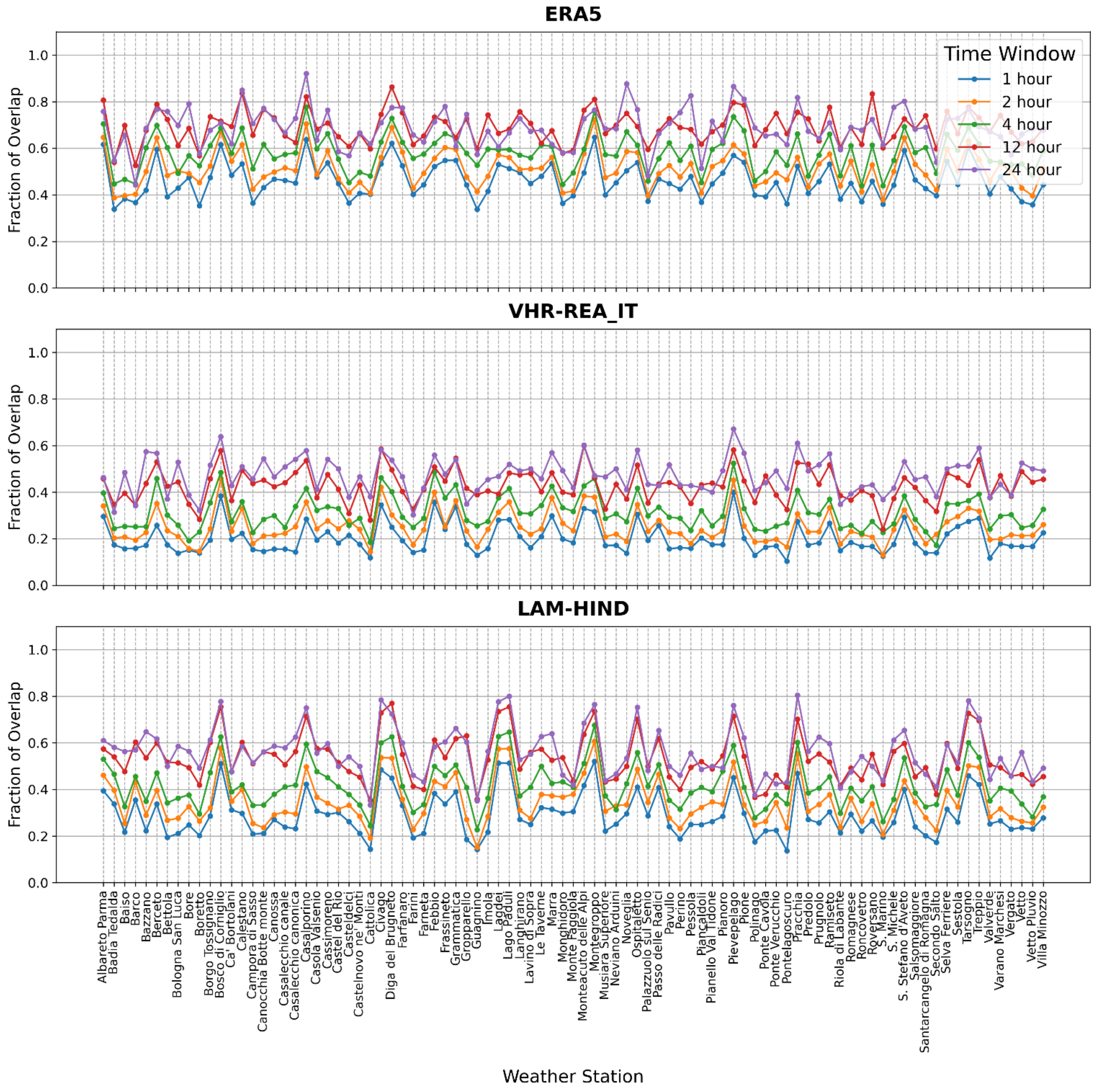


Figure 5 Model to observation overlap ratio for the 95th percentile extreme rainfall events for Emilia-Romagna.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

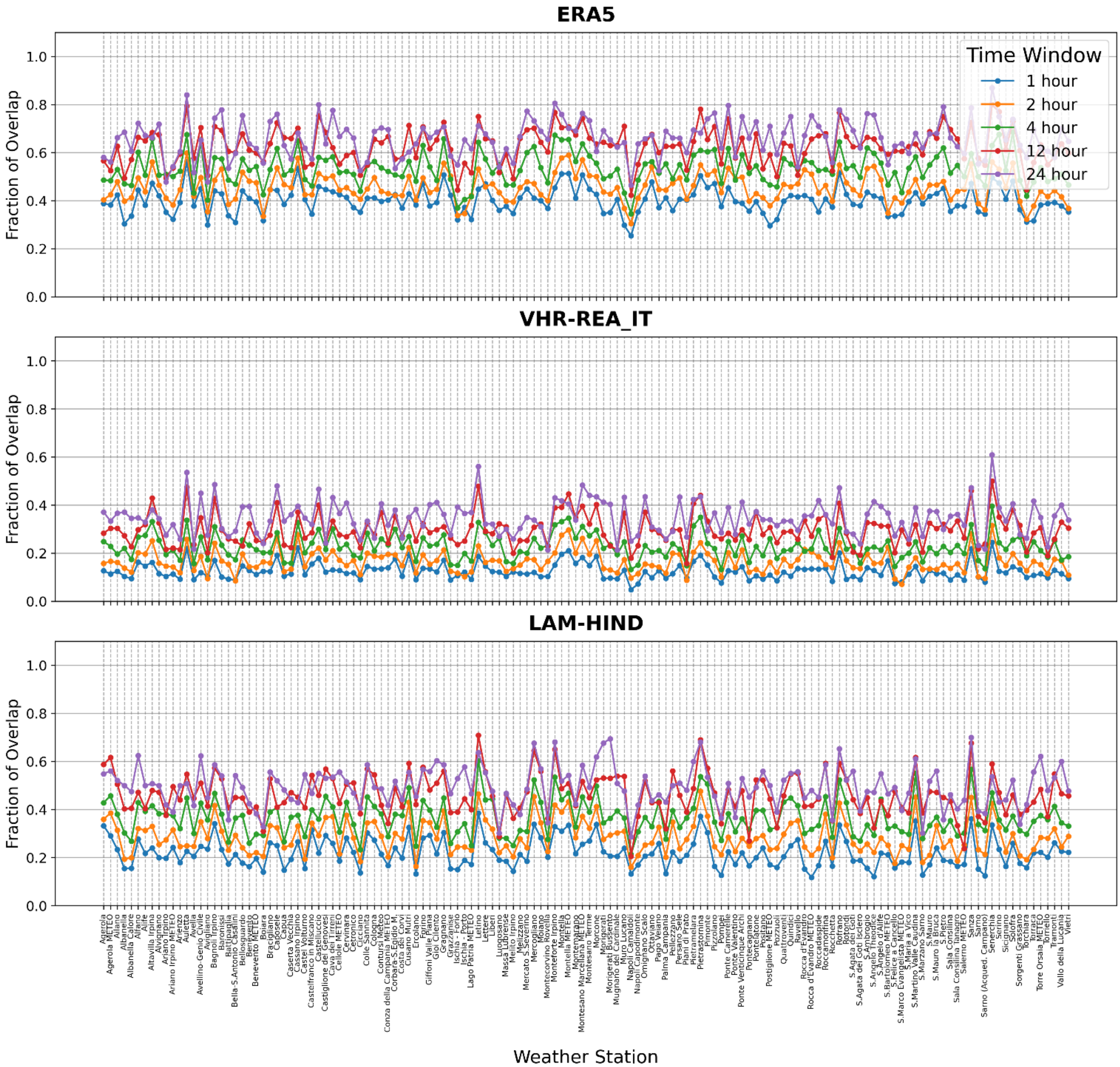


Figure 6 Model to observation overlap ratio for the 95th percentile extreme rainfall events for Campania.

The comparative results from Figure 5 and Figure 6 indicate that LAM-HIND has a slightly higher aggregated overlap ratio of extreme events with the observations compared to VHR-REA_IT. However, this finding is only a generic observation of the overall trend and is not uniform across all geographical locations. For instance, in the multi-model comparison shown in Figure 5 for Emilia-Romagna, weather stations such as Monghidoro, Neviano Arduini, Pavullo, or Ponte Verucchio show an

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

increased fraction of overlap for the 24-hour TW under VHR-REA_IT rather than LAM-HIND. This indicates that local forcing or specific storm tracks may be better captured by different high-resolution model architectures depending on the sub-regional context.

As indicated earlier, a possible hypothesis for the increased overlap ratios for certain datasets may be due to its larger grid size, given that fine-grid datasets would face the 'double-penalty' effect (Cavalleri et al., 2024; Jermey and Renshaw, 2016). For instance, although ERA5 indicates the highest overlap ratio in Figure 5 and Figure 6 (as well as some of the highest skill scores in Table 1 and Table 2), this is an imprecise representation of precipitation at a specific point, and the results must be interpreted with caution. The ERA5 reanalysis relies on convection parameterization and distributes precipitation over relatively coarse ~31 km grid spaces, masking local-scale intensity peaks and homogenizing the rainfall distribution. Thus, the high overlap values in ERA5 are an artifact of its coarse horizontal resolution. By aggregating rainfall over broad boundaries, ERA5 gives the illusion of being precise but realistically avoids the penalties typically associated with high-resolution downscaled models, which are strictly penalized for minor spatial displacements or intensity mismatches when evaluated against point-to-point metrics (Ebert, 2008; Skok and Roberts, 2016; Taszarek et al., 2021).

Similarly, even though the difference in horizontal resolution between VHR-REA_IT (2.2 km) and LAM-HIND (2.7 km) is marginal, the difference in 'effective resolution' is higher (see Section 2.3). Therefore, the slightly higher overlap ratio observed for LAM-HIND may be attributed to the paradox observed with the ERA5 data above. Models with a slightly lower effective resolution or smoother spatial fields often exhibit higher point-to-point overlap ratios by reducing sensitivity to fine-scale spatial displacement errors, and this may therefore not be indicative of superior model performance, as partially evidenced from the t-FSS scores shown in Figure 4. Consequently, as noted by Skok & Roberts (2016) the choice of the reanalysis domain size for rainfall verification should consider typical spatial errors of the model and differences in the model code.

An important observation emerging from Figure 5 and Figure 6 is the systematic increase in overlap ratios, also complemented with the increasing skills scores, with increasing time-aggregation windows. The comparisons between models and observations show a limited consensus on the timing of extreme events, particularly if short (hourly) timescales are considered. Smaller TWs typically have a lower overlap fraction, while larger TWs show higher fraction ratios. Only in rare cases does a smaller TW have a higher fraction than that of a larger TW. For instance, in Figure 5 for the VHR-REA_IT dataset, stations such as Ponte Cavola, Riola di Labante, and Varano Marchesi show a higher fraction overlap for the 12-hour TW compared to the 24-hour TW, while stations such as Casteldelci and Secondo Salto show a higher overlap for the 2-hour TW compared to the 4-hour TW. The typical decreasing consensus with decreasing timescales may be an effect of the 'double penalty' effect (described earlier) when applied temporally, since a larger TW aggregating rainfall over larger time boundaries could give the illusion of being more precise. Notably, however, the skill of a CPM dataset is temporally scale-dependent and should be carefully considered for evaluation.

Some stations simultaneously observe higher or lower ratio across all models. In Figure 5 for Emilia-Romagna, stations like Casalporino, Civago, Febbio, Lago Paduli, Ospitaletto, and Pievepelago show a high fraction of overlap, while stations like Cattolica, Farini, Guagnino, Musiara Superiore, and Pontelagoscuro show a low fraction of overlap simultaneously for all

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

models. Given that both CPMs run on different architectures while being forced by ERA5, the persistence of ERA5 boundary conditions provide a reasonable justification for this pattern. Large-scale atmospheric energy and moisture that are injected by ERA5 at the boundaries would dominate the energy cascade driven by the numerical architecture of the model, thus overriding variations in how CPM datasets resolve sub-grid turbulent fluxes (A. Leonard, 1975; Prein et al., 2015). This highlights that the models have an acute sensitivity to the accuracy of the forcing conditions and is a limiting factor in areas demonstrating a low fraction of overlap.

3.2 Intercomparison and relationality of sub-daily precipitation windows within half-year periods (April-September and October-March)

The traditional verification metrics used in the previous section reveal clear limitations when evaluating CPM datasets to coarse resolution dataset like ERA5. In particular, standard metrics struggle to capture the temporal organization of precipitation extremes and its inherent scale dependence. To address this, this section introduces an analytical framework that complements that of traditional verification metrics. Specifically, a nested time-window analysis is carried out to understand how the downscaled datasets relate extreme precipitation events between time windows at the sub-daily scale. These time windows are clustered as 1-hour, 2-hour, 4-hour, 12-hour, and 24-hours.

The analysis is shown separately for two periods: the first period covers April to September, and the second period covers October to March These intervals correspond to the summer and winter halves of the year, consistent with the Köppen-Geiger climate classification for Mediterranean regimes (Mediterranean climate, 2026). This seasonal division is physically robust from a hydrogeological perspective: the April–September (summer) period is characterized by lower rainfall volumes that primarily generate surface runoff due to dry antecedent soil conditions. Conversely, the October – March (winter) period exhibits higher precipitation amounts that vary considerably by duration, intensity, and cumulative totals (Aleotti, 2004), causing river flooding and leaving the soil profile saturated, unstable, and prone to landslides (Kotsias et al., 2021, 2023; Lombardo and Bitting, 2024; Roccati et al., 2018; The CLIMSAVE Project, 2013).

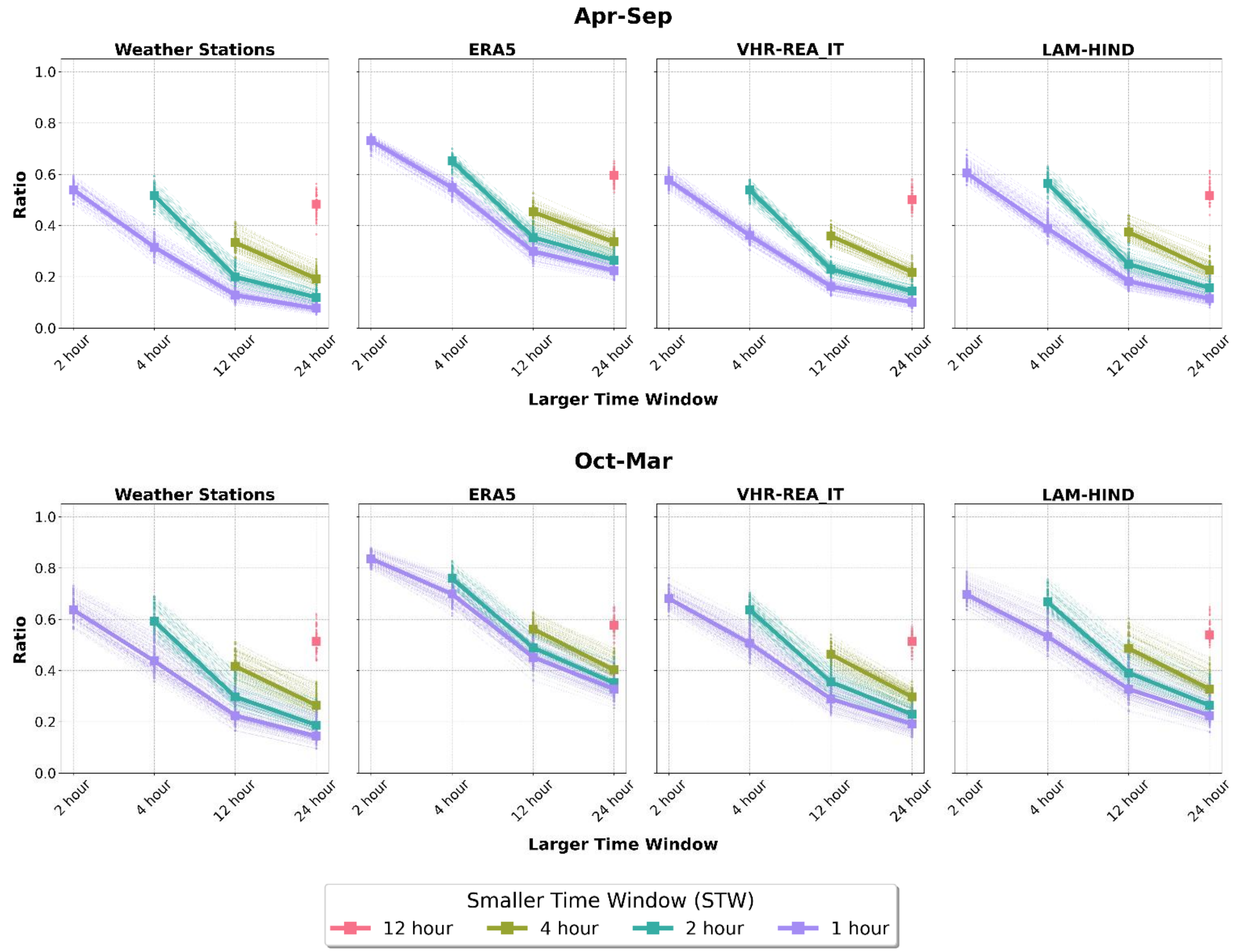


Figure 7 Graphs showing the fraction of extreme to total events (y-axis) in a STW (see legend) that occur within a LTW extreme event (x-axis) for Emilia-Romagna.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

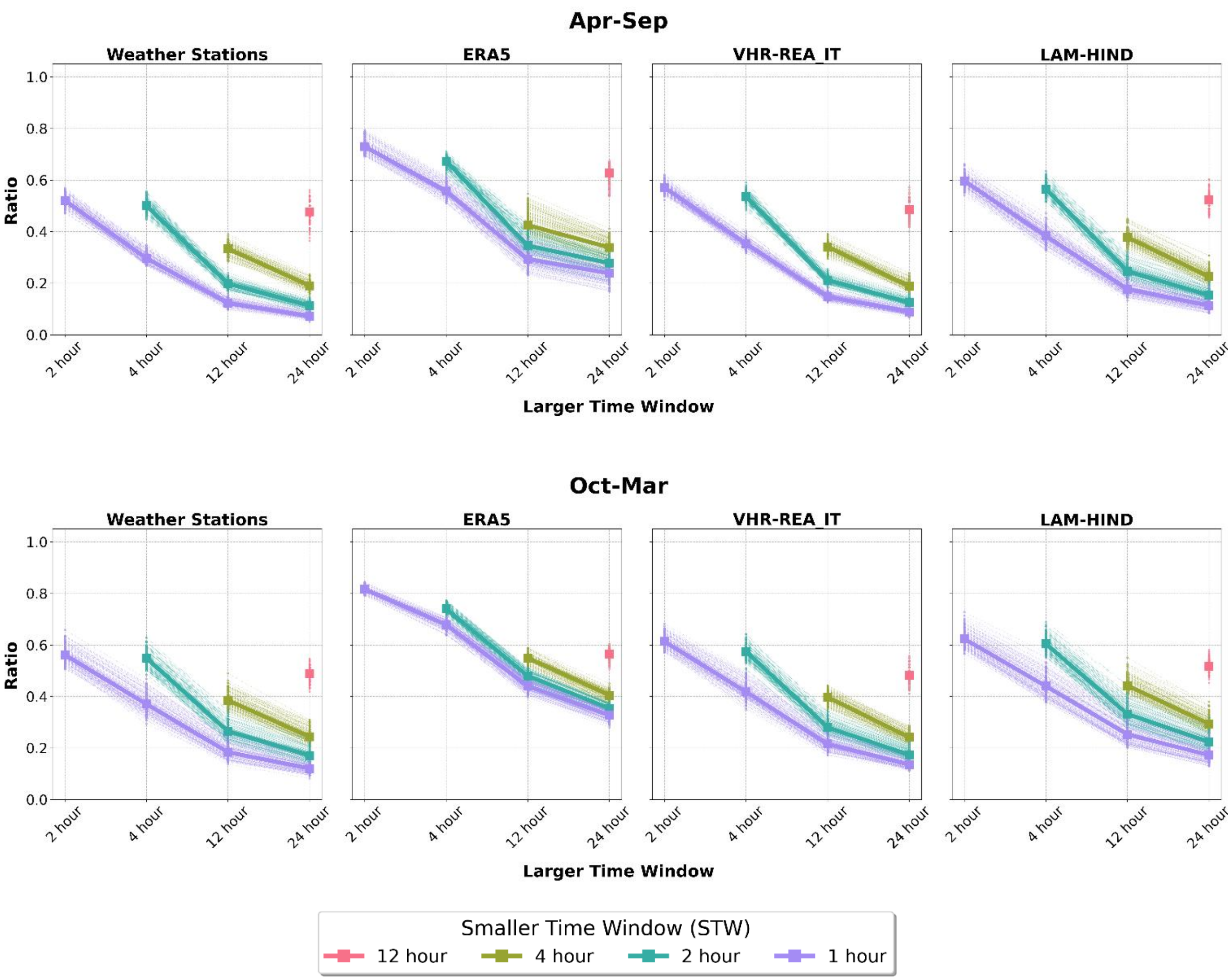


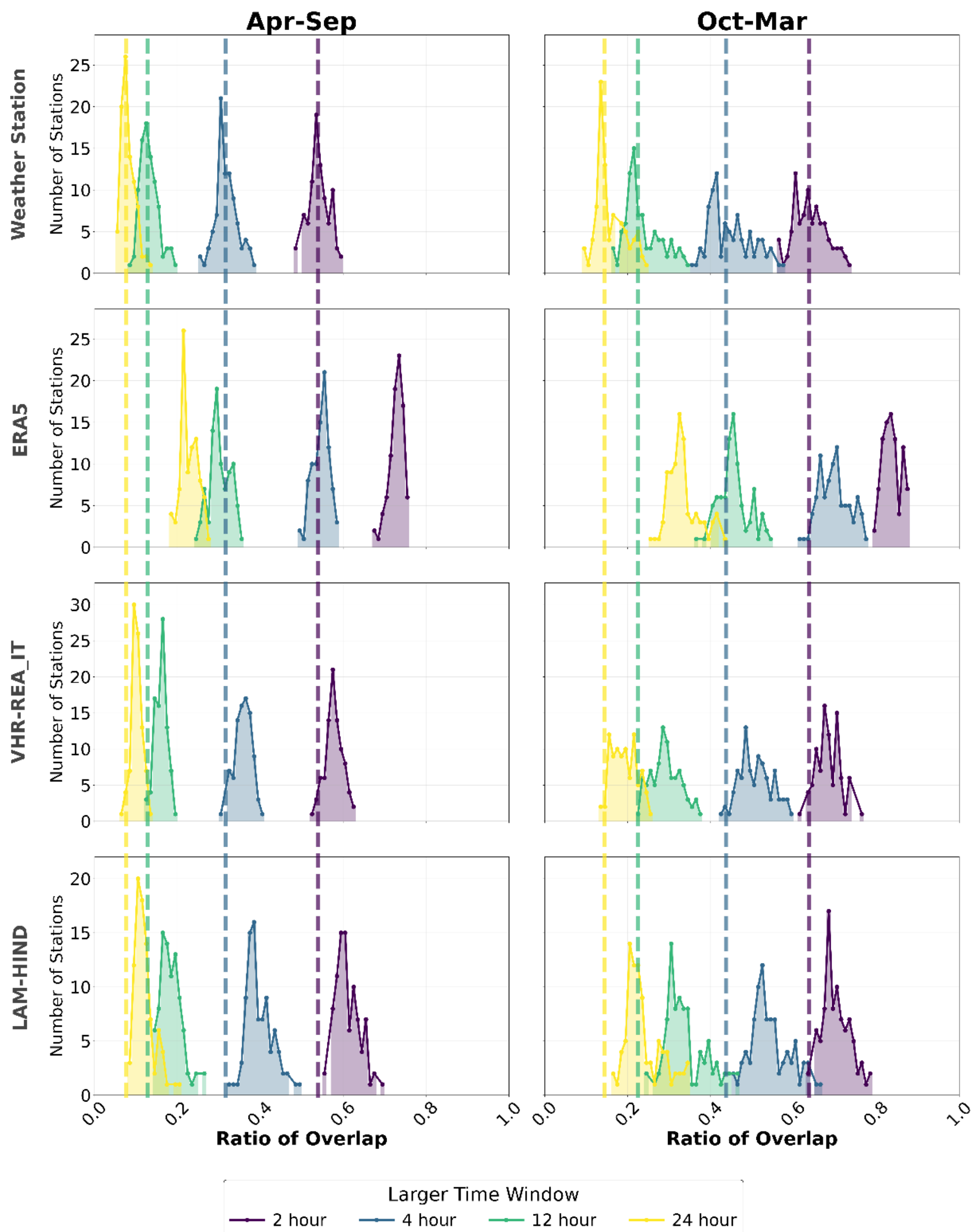
Figure 8 Graphs showing the fraction of extreme to total events (y-axis) in a STW (see legend) that occur within a LTW extreme event (x-axis) for Campania.

For the analysis of the inter-scale temporal links, Figure 7 and Figure 8 illustrate the fraction of extreme events recorded for various Short-Time Windows (STWs) nested within each corresponding Large-Time Window (LTW). For each dataset, precipitation values exceeding the 95th percentile are defined as ‘extremes’ as previously discussed. The overlap fraction is calculated by identifying the number of extreme STW records (e.g., 4-hour) occurring within the duration of an extreme LTW (e.g., 24-hour). Consequently, Figure 7 and Figure 8 display the LTWs along the x-axis, while the y-axis represents the fraction of nested extreme events. The specific STWs are distinguished by line colour, as indicated in the legend. The thin dotted lines indicate individual stations, while the thick lines indicate the spatial average across all stations.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

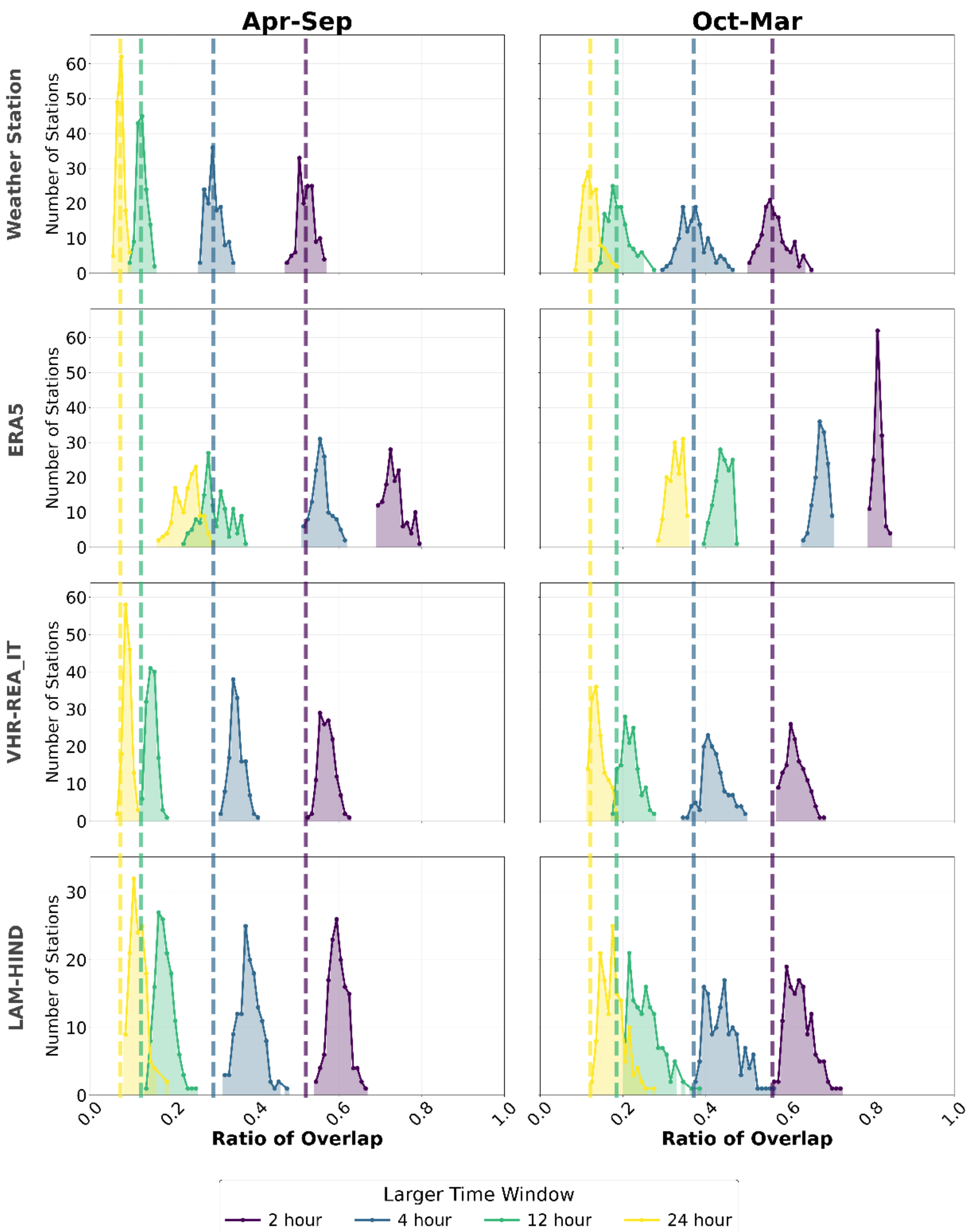
The CPM datasets reproduce the temporal organization of sub-daily extremes much more realistically than ERA5, although they tend to slightly overestimate the temporal linkages between short and long duration events relative to the observations. In contrast, ERA5 fractions are substantially higher than the observed baseline.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026



365 **Figure 9 Histogram for Emilia-Romagna showing the 1-hour extreme STW contributing to each subsequent LTW (2-hour, 4-hour, 12-hour, 24-hour). The ratio is denoted on the x-axis, while the number of stations is denoted on the y-axis. Thick dotted line indicates the median ratio of the observed data.**

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026



370 **Figure 10 Histogram for Campania showing the 1-hour extreme STW contributing to each subsequent LTW (2-hour, 4-hour, 12-hour, 24-hour). The ratio is denoted on the x-axis, while the number of stations is denoted on the y-axis. Thick dotted line indicates the median ratio of the observed data.**

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

Figure 9 and Figure 10 explore a more detailed view of the data presented in Figure 7 and Figure 8, focusing specifically on the inter-scale temporal coupling of the 1-hour STW extreme (STWE) fractions nested within each LTW extreme (LTWE). The thick dashed line denotes the observational baseline median, serving as a reference to quantify the systematic offset of the simulated distributions.

The observations show that extremes at different accumulation times are only partially nested. A short-duration extreme is often not embedded within an extreme accumulated over 12 or 24 hours. The conditional fraction is well below one, and generally decreases as the separation between the short and long accumulation windows increases. Figure 7 to Figure 10 therefore show that STWEs and LTWEs are related, but they are not simply different representations of the same events. This also has an important physical and practical implication: many short-duration extremes are concentrated and relatively isolated rainfall bursts, whereas long-duration extremes more often correspond to sustained or repeated precipitation.

This evaluation framework also adds a second important result. The CPMs reproduce STWEs and LTWEs reasonably well when the different accumulation windows are assessed separately, and they represent their conditional association much more realistically than ERA5. However, they still overestimate this association relative to the observations: STWEs are too frequently embedded within the LTWE; another way to put it is that LTWEs are systematically more dependent on short-term convective bursts than what is physically observed. This indicates excessive temporal continuity in simulated heavy-rainfall episodes, a bias that is not apparent from conventional verification metrics applied separately to each duration.

A possible explanation of this temporal persistence bias may be commonly observed in the CPMs, where the simulated decay of convective cells is slower than natural systems, leading to an over-representation of extreme sub-daily events in the models (Kendon et al., 2012). Alternatively, this could highlight the representativeness problem, as local-scale spatial variability is not fully captured by grid-scale variability, misrepresenting the exact location, size, or intensity of a thunderstorm (Volosciuk et al., 2017; Zwiers et al., 2013). The spatial variance is inflated, leading to an overestimation of spatial extremes (Volosciuk et al., 2017).

It is also observed from Figure 9 and Figure 10 that the interlinked temporal fractions for VHR-REA_IT are more closely aligned with the observations (the median line) than those for LAM-HIND. For instance, the VHR-REA_IT 1-hour STW shows a closer fraction match with the 12-hour and 24-hour LTWs than LAM-HIND. Consequently, while LAM-HIND demonstrates slightly better skill in capturing absolute spatial point-to-point overlap with observations (as discussed for Figure 5 and Figure 6), it displays lower fidelity in reproducing the internal temporal dependencies between nested sub-daily extreme windows (as seen in Figure 9 and Figure 10).

The structural differences between the CPMs mentioned earlier, such as effective resolution, model physics, nesting strategies, etc, may all be considered as plausible contributing mechanisms for why the LAM-HIND dataset is slightly better than VHR-REA_IT at capturing the magnitude and frequency of extremes. However, its overestimation of the inter-temporal link ratios implies that it also generates a greater number of false alarms. Therefore, studies isolating these different contributions would be helpful.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

The results suggest that while high-resolution models characterize the magnitudes of extremes well, they still suffer from an overestimation of the temporal continuity of extreme events at the sub-daily scale, resulting from the model's tendency to distribute these events over persistent, high-volume blocks of rainfall (Figure 3). ERA5, on the other hand, shows the weakest performance regarding temporal continuity, significantly exaggerating the influence of STWEs on LTWEs. While traditional model skill scores are effective for assessing the general performance of all datasets, they do not provide a complete representation of how sub-daily time windows relate to one another for high-resolution CPM products, which is a critical factor in determining dataset fidelity. The STW-to-LTW nested framework presented in this section offers an additional diagnostic tool for validating high-resolution precipitation datasets.

4 Conclusions

High-resolution modelled data are an essential tool for evaluating precipitation extremes, which serve as critical indicators for multi-hazard risk assessments, including floods and landslides. These datasets are particularly valuable for mitigating risks to life and property in regions characterized by a low density of in-situ rain gauges (Formetta et al., 2024). This study evaluated the quality of two high-resolution downscaled datasets over Italy, for the sub-daily precipitation extremes defined by the 95th percentile of wet events. The two datasets are the VHR-REA_IT (*Very High-Resolution REAnalysis for Italy*), developed by CMCC and resolved to a 2.2 km spatial grid using the COSMO-CLM Regional Climate Model (RCM), and LAM-HIND (*LaMMA produced Atmospheric Hindcast Data*), developed by LaMMA and resolved to a 2.7 km grid. The evaluation was performed against a 10-year observational baseline (January 1, 2014, to December 31, 2023) obtained from the high-frequency rain gauge networks of the '*Agenzia Prevenzione Ambiente Energia Emilia-Romagna*' (ARPAE) and the '*Centro Funzionale Multirischi della Protezione Civile Regione Campania*'. The results reveal a clear and substantial added value in utilizing these convection permitting datasets over standard coarse global products like ERA5 at the sub-daily scale, which typically smooth out localized, high-intensity atmospheric features.

The analysis finds that coarse global reanalysis like ERA5, despite good skill scores using traditional verification metrics, struggle to capture the rapid, sub-daily transitions of heavy rain from one time window to the next. To address these complexities, the introduction of the nested time-window framework provides a strong diagnostic overview of the dataset performance, successfully capturing the persistence biases between STWEs and LTWEs. It shows that only a limited fraction of observed STWEs is embedded within LTWEs. CPMs reproduce this cross-timescale organization substantially better than ERA5, but still overestimate the association, indicating excessive temporal persistence of simulated heavy-rainfall episodes. This behaviour is largely hidden by conventional verification metrics. Furthermore, the CPM datasets show a limited consensus regarding the exact spatial and temporal placement of sub-daily extreme events, more particularly for shorter than for longer accumulation windows. Therefore, the skill of the CPMs is temporally scale-dependent and yields more accuracy when longer time windows are considered, due to the tendency of the model to distribute short extreme convective events into persistent, high-volume chunks of rainfall spread out over longer durations. By offering a deeper insight into the simulated convective

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

life cycles, the novelty of this framework extends beyond the comparisons between the two CPM datasets and could provide a valuable complementary diagnostic for future validation studies of high-resolution datasets.

Between the two CPM datasets, the study found that the LAM-HIND dataset (2.7 km) yields a higher fraction of event overlap, as well as marginally better skill metrics in detecting the occurrence and magnitude of extremes as evidenced by higher Probability of Detection (POD) and Critical Success Index (CSI) scores. Conversely, this elevated detection capability introduces a higher false alarm rate and a lower Success Ratio (SR) compared to the more conservative VHR-REA_IT configuration (2.2 km). So, while VHR-REA_IT may miss more extreme events, its slightly higher SR and closer alignment with observed inter-scale temporal fractions represent a more physically consistent rainfall life cycle. However, for evaluating specific hazard and risk factors of precipitation, the choice of the CPM dataset should consider potential inherent structural differences. Micro-scale spatial displacement caused by minor differences in the resolution may still be penalized by the 'double penalty' effect. Meanwhile a multi-stage dynamical downscaling technique using an intermediate hydrostatic nest may potentially optimize the final non-hydrostatic CPM output in the case of extreme precipitation. Furthermore, CPM datasets appear to show an acute sensitivity to the boundary conditions employed and may be an indication of unresolved energy cascades in the CPM architecture. Consequently, both CPM datasets can be utilized synergistically to address distinct research questions, as the differences are subtle depending on the verification metrics considered and are mainly contrasted by the performance enhancements achieved when moving away from coarse global model products

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

CRediT authorship contribution statement

Oliver S. Carlo: Conceptualization, Writing – original draft, Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data Curation. Giusy Fedele: Writing – review & editing, Validation, Methodology, Resources. Paolo Stocchi: Writing – review & editing, Methodology, Resources. Sandro Calmanti: Writing – review & editing, Resources, Project administration. Mario Raffa: Writing – review & editing, Resources. Paola Mercogliano: Writing – review & editing, Supervision, Resources. Piero Lionello: Conceptualization, Writing – review & editing, Supervision, Resources, Methodology.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data used in this study is publicly available online and can be accessed as follows:
The VHR-REA_IT dataset is accessible here: https://dds.cmcc.it/#/dataset/era5-downscaled-over-italy/hourly
The LAM-HIND dataset is accessible here: https://meteohub.mistralportal.it/app/datasets
Rain gauge data from the 'Centro Funzionale Multirischi della Protezione Civile Regione Campania' is accessible here: https://centrofunzionale.regione.campania.it/dati
Rain gauge data from the 'Agenzia Prevenzione Ambiente Energia Emilia-Romagna' is accessible here: https://www.arpae.it/it/temi-ambientali/meteo/dati-e-osservazioni/rete-di-monitoraggio-idrometeorologica

Use of Generative Artificial Intelligence

During the preparation of this study, the authors use ChatGPT (OpenAI) and Gemini (Google) in order to enhance readability. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the manuscript.

Acknowledgements

The authors acknowledge the financial support received under the PNRR innovation grant project RETE (Resilience of the Electric Transmission Grid to Extreme Events), and financial support from ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU (Project code CN_00000033, CUP C83C22000560007)

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

References

A. Leonard: Energy Cascade in Large-Eddy Simulations of Turbulent Fluid Flows, Adv. Geophys., 18, 237–248, 1975.

Aleotti, P.: A warning system for rainfall-induced shallow failures, Eng. Geol., 73, 247–265, https://doi.org/10.1016/j.enggeo.2004.01.007, 2004.

ANSA English Desk: Italy saw more extreme weather events in 2025, SOS due to record temperatures, , 30th December, 2025.

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., and Adinolfi, M.: The first multi - model ensemble of regional climate simulations at kilometer - scale resolution , part I : evaluation of precipitation, Clim. Dyn., 57, 275–302, https://doi.org/10.1007/s00382-021-05708-w, 2021.

Billi, P.: Mean Annual and Mean Monthly Temperature Variation in Italy Across the Last Century, Atti della Soc. Toscana Sci. Nat. Memorie, Ser. A, 130, 105–115, https://doi.org/10.2424/ASTSN.M.A.2023.08, 2023.

Bonanno, R., Lacavalla, M., and Sperati, S.: A new high-resolution Meteorological Reanalysis Italian Dataset: MERIDA, Q. J. R. Meteorol. Soc., 145, 1756–1779, https://doi.org/10.1002/qj.3530, 2019.

Bongiovanni, G., Matiu, M., Crespi, A., Napoli, A., Majone, B., and Zardi, D.: EEAR-Clim: a high-density observational dataset of daily precipitation and air temperature for the Extended European Alpine Region, Earth Syst. Sci. Data, 17, 1367–1391, https://doi.org/10.5194/essd-17-1367-2025, 2025.

Mediterranean climate: https://www.britannica.com/science/Mediterranean-climate, last access: 18 May 2026.

Capecchi, V., Pasi, F., Gozzini, B., and Brandini, C.: A convection-permitting and limited-area model hindcast driven by ERA5 data: precipitation performances in Italy, Clim. Dyn., 61, 1411–1437, https://doi.org/10.1007/s00382-022-06633-2, 2023.

Cavalleri, F., Lussana, C., Viterbo, F., Brunetti, M., Bonanno, R., Manara, V., Lacavalla, M., Sperati, S., Raffa, M., Capecchi, V., Cesari, D., Giordani, A., Cerenzia, I. M. L., and Maugeri, M.: Multi-scale assessment of high-resolution reanalysis precipitation fields over Italy, Atmos. Res., 312, 107734, https://doi.org/10.1016/j.atmosres.2024.107734, 2024.

ERA5-Land hourly data from 1950 to present:

ERA5: What is the spatial reference: https://confluence.ecmwf.int/spaces/CKB/pages/65237704/ERA5+What+is+the+spatial+reference.

What is Statistical and Dynamical downscaling? https://climate.copernicus.eu/sites/default/files/2021-01/infosheet8.pdf.

ERA5 hourly data on single levels from 1940 to present:

Climate reanalysis: https://climate.copernicus.eu/climate-reanalysis, last access: 20 May 2026.

Crespi, A., Napoli, A., Galassi, G., Lazzeri, M., Parodi, A., Zardi, D., and Pittore, M.: Leveraging observations and model reanalyses to support regional climate change adaptation activities: An integrated assessment for the Marche Region (Central Italy), Clim. Serv., 36, https://doi.org/10.1016/j.cliser.2024.100512, 2024.

Desiato, F., Fioravanti, G., Fraschetti, P., Perconti, W., and Toreti, A.: Climate indicators for Italy: calculation and dissemination, Adv. Sci. Res., 6, 147–150, https://doi.org/10.5194/asr-6-147-2011, 2011.

Drofa, O. V and Malguzzi, P.: Parameterization of Microphysical Processes in a Non Hydrostatic Prediction Model, in: 14th International Conference on Clouds and Precipitation (ICCP), 2004.

Duc, L., Saito, K., and Seko, H.: Spatial-temporal fractions verification for high-resolution ensemble forecasts, Tellus, Ser. A Dyn. Meteorol. Oceanogr., 65, https://doi.org/10.3402/tellusa.v65i0.18171, 2013.

Ebert, E. E.: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework, in: Meteorological Applications, 51–64, https://doi.org/10.1002/met.25, 2008.

Formetta, G., Dallan, E., Borga, M., and Marra, F.: Sub-daily precipitation returns levels in ungauged locations: Added value of combining observations with convection permitting simulations, Adv. Water Resour., 194, https://doi.org/10.1016/j.advwatres.2024.104851, 2024.

Fumière, Q., Déqué, M., Nuissier, O., Somot, S., Alias, A., Caillaud, C., Laurantin, O., and Seity, Y.: Extreme rainfall in Mediterranean France during the fall: added value of the CNRM-AROME Convection-Permitting Regional Climate Model, Clim. Dyn., 55, 77–91, https://doi.org/10.1007/s00382-019-04898-8, 2020.

Giordani, A., Ruggieri, P., and Di Sabatino, S.: Added value of a multi-model ensemble of convection-permitting rainfall reanalyses over Italy, Atmos. Res., 328, https://doi.org/10.1016/j.atmosres.2025.108402, 2026.

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., De Chiara,

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

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., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Jeevanjee, N.: Vertical Velocity in the Gray Zone, J. Adv. Model. Earth Syst., 9, 2304–2316, https://doi.org/10.1002/2017MS001059, 2017.

Jermey, P. M. and Renshaw, R. J.: Precipitation representation over a two-year period in regional reanalysis, Q. J. R. Meteorol. Soc., 142, 1300–1310, https://doi.org/10.1002/qj.2733, 2016.

Kain, J. S.: The Kain-Fritsch Convective Parameterization: An Update, J. Appl. Meteorol. Climatol., 43, 170, https://doi.org/https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, 1996.

Kendon, E. J., Roberts, N. M., Senior, C. A., and Roberts, M. J.: Realism of rainfall in a very high-resolution regional climate model, J. Clim., 25, 5791–5806, https://doi.org/10.1175/JCLI-D-11-00562.1, 2012.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G., and Wilkinson, J. M.: Do convection-permitting regional climate models improve projections of future precipitation change?, Bull. Am. Meteorol. Soc., 98, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1, 2017.

Kotsias, G., Lolis, C. J., Hatzianastassiou, N., Lionello, P., and Bartzokas, A.: An objective definition of seasons for the Mediterranean region, Int. J. Climatol., 41, E1889–E1905, https://doi.org/10.1002/joc.6819, 2021.

Kotsias, G., Lolis, C. J., Hatzianastassiou, N., Bakas, N., Lionello, P., and Bartzokas, A.: Objective climatology and classification of the Mediterranean cyclones based on the ERA5 data set and the use of the results for the definition of seasons, Theor. Appl. Climatol., 152, 581–597, https://doi.org/10.1007/s00704-023-04374-8, 2023.

Lewis D Grasso: The Differentiation between Grid Spacing and Resolution and Their Application to Numerical Modelling, Bull. Am. Meteorol. Soc., 2000.

Lombardo, K. and Bitting, M.: A Climatology of Convective Precipitation over Europe, Mon. Weather Rev., 152, 1555–1585, https://doi.org/10.1175/MWR-D-23, 2024.

Loprieno, L. L., Raffa, M., Campanale, A., Adinolfi, M., and Mercogliano, P.: Comparative study of the COSMO and ICON models performances over Italy: Key insights from the impactful year 2023, Atmos. Res., 327, https://doi.org/10.1016/j.atmosres.2025.108339, 2026.

Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., Kotlarski, S., and Caillaud, C.: Convection-permitting modeling with regional climate models: Latest developments and next steps, Wiley Interdiscip. Rev. Clim. Chang., 12, e731, https://doi.org/https://doi.org/10.1002/wcc.731, 2021.

Manco, I., Riviera, W., Zanetti, A., Briscolini, M., Mercogliano, P., and Navarra, A.: A new conditional generative adversarial neural network approach for statistical downscaling of the ERA5 reanalysis over the Italian Peninsula, Environ. Model. Softw., 188, https://doi.org/10.1016/j.envsoft.2025.106427, 2025.

Marsigli, C., Montani, A., and Paccagnella, T.: Provision of boundary conditions for a convection-permitting ensemble: Comparison of two different approaches, Nonlinear Process. Geophys., 21, 393–403, https://doi.org/10.5194/npg-21-393-2014, 2014.

Matte, D., Laprise, R., and Thériault, J. M.: Comparison between high-resolution climate simulations using single- and double-nesting approaches within the Big-Brother experimental protocol, Clim. Dyn., 47, 3613–3626, https://doi.org/10.1007/s00382-016-3031-9, 2016.

Il Mattino: Italy's Extreme Weather Challenge, , 7th November, 2025.

Mazzoglio, P., Lompi, M., Marra, F., Dallan, E., Deidda, R., Claps, P., Manfreda, S., Noto, L. V., Viglione, A., Raffa, M., Mercogliano, P., Marani, M., Caporali, E., and Borga, M.: Orographic and land-sea contrast effects in convection-permitting simulations of extreme sub-daily precipitation, Weather Clim. Extrem., 49, https://doi.org/10.1016/j.wace.2025.100798, 2025.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J. N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

https://doi.org/10.5194/egusphere-2026-4515
Preprint. Discussion started: 30 July 2026

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, Int. J. Climatol., 37, 305–327, https://doi.org/10.1002/joc.4706, 2017.

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. van, and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev. Geophys., 53, https://doi.org/https://doi.org/10.1002/2014RG000475, 2015.

Raffa, M., Reder, A., Adinolfi, M., and Mercogliano, P.: A Comparison between One-Step and Two-Step Nesting Strategy in the Dynamical Downscaling of Regional Climate Model COSMO-CLM at 2 . 2 km Driven by ERA5 Reanalysis, Atmosphere (Basel)., 12, 260, 2021a.

Raffa, M., Reder, A., Marras, G. F., Mancini, M., Scipione, G., Santini, M., and Mercogliano, P.: VHR-REA_IT dataset: Very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM, Data, 6, https://doi.org/10.3390/data6080088, 2021b.

Reder, A., Raffa, M., Padulano, R., Rianna, G., and Mercogliano, P.: Characterizing extreme values of precipitation at very high resolution: An experiment over twenty European cities, Weather Clim. Extrem., 35, 100407, https://doi.org/10.1016/j.wace.2022.100407, 2022.

Roccati, A., Faccini, F., Luino, F., Turconi, L., and Guzzetti, F.: Rainfall events with shallow landslides in the Entella catchment, Liguria, northern Italy, Nat. Hazards Earth Syst. Sci., 18, 2367–2386, https://doi.org/10.5194/nhess-18-2367-2018, 2018.

Roebber, P. J.: Visualizing multiple measures of forecast quality, Weather Forecast., 24, 601–608, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Le Roux, R., Katurji, M., Zawar-Reza, P., Quénol, H., and Sturman, A.: Comparison of statistical and dynamical downscaling results from the WRF model, Environ. Model. Softw., 100, 67–73, https://doi.org/10.1016/j.envsoft.2017.11.002, 2018.

Schleiss, M., Jaffrain, J., and Berne, A.: Statistical analysis of rainfall intermittency at small spatial and temporal scales, Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL049000, 2011.

Skok, G. and Roberts, N.: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts, Q. J. R. Meteorol. Soc., 142, 2599–2610, https://doi.org/10.1002/qj.2849, 2016.

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J. N.: The ERA5 global reanalysis from 1940 to 2022, Q. J. R. Meteorol. Soc., 150, 4014–4048, https://doi.org/10.1002/qj.4803, 2024.

Stocchi, P., Pichelli, E., Alavez, J. A. T., Coppola, E., Giuliani, G., and Giorgi, F.: Non-Hydrostatic Regcm4 (Regcm4-NH): Evaluation of Precipitation Statistics at the Convection-Permitting Scale over Different Domains, Atmosphere (Basel)., 13, https://doi.org/10.3390/atmos13060861, 2022.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L.: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, Rev. Geophys., 56, 79–107, https://doi.org/10.1002/2017RG000574, 2018.

Taszarek, M., Pilguj, N., Allen, J. T., Gensini, V., Brooks, H. E., and Szuster, P.: Comparison of Convective Parameters Derived from ERA5 and MERRA-2 with Rawinsonde Data over Europe and North America, J. Clim., 34, 3211–3237, https://doi.org/10.1175/JCLI-D-20-0484.1, 2021.

The CLIMSAVE Project: Climate Change Integrated Assessment Methodology for Cross-Sectoral Adaptation and Vulnerability in Europ, 2013.

Trentini, L., Dal Gesso, S., Venturini, M., Guerrini, F., Calmanti, S., and Petitta, M.: A Novel Bias Correction Method for Extreme Events, Climate, 11, https://doi.org/10.3390/cli11010003, 2023.

Viterbo, F., Sperati, S., Vitali, B., D'Amico, F., Cavalleri, F., Bonanno, R., and Lacavalla, M.: MERIDA HRES: A new high-resolution reanalysis dataset for Italy, Meteorol. Appl., 31, https://doi.org/10.1002/met.70011, 2024.

Volosciuk, C., Maraun, D., Vrac, M., and Widmann, M.: A combined statistical bias correction and stochastic downscaling method for precipitation, Hydrol. Earth Syst. Sci., 21, 1693–1719, https://doi.org/10.5194/hess-21-1693-2017, 2017.

World Meteorological Organization - WMO: Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8), 2008.

Zwiers, F. W., Alexander, L. V., Hegerl, G. C., Knutson, T. R., Kossin, J. P., Naveau, P., Nicholls, N., Schär, C., Seneviratne, S. I., and Zhang, X.: Climate Extremes: Challenges in Estimating and Understanding Recent Changes in the Frequency and Intensity of Extreme Climate and Weather Events, in: Climate Science for Serving Society, Springer Netherlands, 339–389, https://doi.org/10.1007/978-94-007-6692-1_13, 2013.