Simulating Hail and Lightning over the Alpine Adriatic Region - A model Intercomparison Study

Barbara Malecic[1], Ruoyi Cui[2], Marie-Estelle Demory[3], Kristian Horvath[4], Damjan Jelic[5], Christoph Schär[2], Maja Telišman Prtenjak[6], Patricio Velasquez[7], and Nikolina Ban[8]

[1]Universtiy of Zagreb, Faculty of Science
[2]ETH Zurich
[3]ETH
[4]Meteorological and Hydrological Service
[5]Andrija Mohorovičić Geophysical Institute, Department of Geophysics, Faculty of Science, University of Zagreb
[6]University of Zagreb, Faculty of Science, Department of Geophysics
[7]Unknown
[8]University of Innsbruck

November 26, 2022

Abstract

Hail is a significant convective weather hazard, often causing considerable crop and property damage across the world. Although extremely damaging, hail still remains a challenging phenomenon to model and forecast, given the limited computational resolution and the gaps in understanding the processes involved in hail formation. Here, eight hailstorms occurring over the Alpine-Adriatic region are analyzed using Weather Research and Forecasting (WRF) and Consortium for Small Scale Modeling (COSMO) simulations, with embedded HAILCAST and Lightning Potential Index (LPI) diagnostics at kilometer-scale grid spacing (~2.2 km). In addition, a model intercomparison study is performed to investigate the ability of the different modeling systems in reproducing such convective extremes, and to further assess the uncertainties associated with simulations of such localized phenomena. The results are verified by hailpad observations over Croatia, radar estimates of hail over Switzerland and lightning measurements from the LINET network. The analysis revealed that both HAILCAST and LPI are able to reproduce the areas and intensities affected by hail and lightning. Moreover, the hail and lightning fields produced by both models are similar, although a slight tendency of WRF to produce smaller hail swaths with larger hailstones and higher LPI compared to COSMO is visible. It is found that these differences can be explained by systematic differences in vertical profiles of microphysical properties and updraft strength between the models. Overall, the promising results indicate that both HAILCAST and LPI could be valuable tools for real-time forecasting and climatological assessment of hail and lightning in current and changing climates.

1 Introduction

Hail is a severe weather hazard that can produce significant crop and property damage across the world (Allen et al., 2020), especially when it occurs over highly populated areas with high-density assets (Kunz et al., 2018). In the literature, a large number of hailstorms causing more than US$1 billion in damage is reported across the world (Schuster et al., 2005; Changnon, 2009; Brown et al., 2015; Kunz et al., 2018; Púcik et al., 2019). As described by Punge and Kunz (2016) and Púcik et al. (2019), several hail hotspots can be found in Europe, including the pre-Alpine and Adriatic areas. Although large hail occurs less often over the highest mountain peaks in the central Alps, severe hailstorms frequently affect Switzerland with up

to 4 large hail days per year (Nisi et al., 2016; Púcik et al., 2019). In this area, the maximum hail diameter can sometimes exceed 10 cm (e.g., see Figure 8 from Púčik et al., 2019). Furthermore, parts of Croatia (Počakal et al., 2018; Jelić et al., 2020) and broader northern Adriatic region (e.g. Manzato (2012)) have similar statistics of hail frequency as southern Germany or southeastern Austria (Punge & Kunz, 2016). Therefore, considering the high economic losses associated with (severe) hailstorms, and high frequencies of hail occurrence, it is very important to have successful hail forecasts, both for short-term numerical weather prediction (NWP) and long-term climate-change adaptation strategies.

One of the largest limitations in understanding processes involved in hail formation is the lack of dense and direct measurements of hail properties on the ground. Hailpads, which are simple meteorological devices consisting of a stand and a measuring plate, represent one of the few methods to detect and measure hailstones directly on the ground. Besides the number of falling hailstones and their diameters, hailpads can also detect the intensity (i.e., kinetic energy) of hail (Smith & Waldvogel, 1989). In Europe, hailpad networks exist in several regions including parts of Spain, France, Greece, northern Italy, eastern Austria and parts of Croatia (Svabik, 1989; Dessens, 1998; Giaiotti et al., 2003; Sioutas et al., 2009; Počakal et al., 2009; Berthet et al., 2011; Počakal, 2011) and have also been used in randomized hail suppression experiments in Switzerland (Federer et al., 1978). Although hailpads are one of the few sources of direct information on hail occurrence, they provide spatially discrete (but unique) information on hail occurrence as they only record hail at the point where they are installed.

Another source of information on hail occurrence is related to weather radars. As the abilities of weather radars to detect different kinds of hydrometeors such as rain, snow and hail progressed over the years, several hail detection algorithms have been developed (e.g. Waldvogel et al., 1979; Witt et al., 1998). At the present, hail detection algorithms are widely used as hail proxies and can provide spatially continuous information on various hail properties, e.g., probability of hail occurrence or maximum expected hailstone size. In Switzerland, two hail detection algorithms are operational in real-time, namely, Probability of Hail (POH, Waldvogel et al., 1979; Foote et al., 2005) that indicates a probability of a hailstorm occurring at a certain location, and Maximum Expected Severe Hail Size (MESHS, Treloar, 1998; Joe et al., 2004) that estimates expected severe hail size at the ground over the Alpine region.

An additional challenge in understanding hail processes is the limited number of high-resolution modeling studies of hailstorms. With increasing computational power, it has become possible to run simulations at convection-permitting scales (horizontal grid spacing < 4 km). Several studies reported the benefits of using models at kilometer scales for more realistic representations of convective processes (Leutwyler et al., 2017), mean diurnal cycles of precipitation (Ban et al., 2014), spatial precipitation patterns and associated extreme values (Prein et al., 2013; Brisson et al., 2016, 2018; Fowler et al., 2021; Pichelli et al., 2021), better representation of convective clouds (Keller et al., 2016; Brisson et al., 2016; Hentgen et al., 2019), local wind systems like sea breeze (Belušić et al., 2018), and complex terrain winds (Horvath et al., 2012). Since models, when run at km scales, can produce a more realistic representation of convective processes, Adams-Selin and Ziegler (2016) integrated a physically improved 1D hail growth scheme – called HAILCAST (Poolman, 1992; Brimelow et al., 2002; Jewell & Brimelow, 2009) – with the km-scale WRF model. When HAILCAST is coupled with WRF, the model simulates the maximum expected hail size at the ground using the profiles of cloud liquid and ice water, vertical velocity, temperature, water vapor and pressure fields from a given model timestep. Several recent studies employed HAILCAST embedded in high-resolution numerical models, such as WRF or COSMO, to study hailstorms occurring over the United States and Europe. The studies found that the models can reproduce the atmospheric conditions and triggering mechanisms responsible for hailstorm formation, resulting in simulating comparable hailstorms to those observed over the complex terrain of the United States (Adams-Selin & Ziegler, 2016; Adams-Selin et al., 2019), Switzerland (Trefalt et al., 2018; Raupach et al., 2021; Cui et al., 2023), Italy (Manzato et al., 2020; Tiesi et al., 2022), and Croatia (Malečić et al., 2022).

Similar to hail, lightning poses a serious threat to human lives (Curran et al., 2000; Holle et al., 2005), wind turbines (Rachidi et al., 2008) and transportation (Kanata et al., 2012; Lee & Collins, 2017; Thornton et al.,

2017). Moreover, lighting is a major cause of wildfires (Latham & Williams, 2001; Abatzoglou & Williams, 2016; Dowdy et al., 2017). Considering the hazards associated with lightning occurrence, the lightning potential index (LPI) was developed as a tool for diagnosing areas prone to lightning discharges (Lynn & Yair, 2008; Yair et al., 2010). With a better representation of convective processes km-scale simulations, LPI offers the possibility to use the parameterizations of lightning that describe the non-inductive process occurring inside a thundercloud (Yair et al., 2010; Brisson et al., 2021). It is defined as a potential for charge formation and separation inside a thundercloud and it relies on the presence of both solid and liquid hydrometeors. Even though LPI is not directly connected to the observed number of lightning flashes, several studies found that LPI could be a valuable tool for implicit lightning forecasting in COSMO (Sokol & Minářová, 2020, Cui et al., 2023) and WRF (Yair et al., 2010; Lagasio et al., 2017a; Malečić et al., 2022) models. Recently, LPI was used in the climatological assessment of lightning over Germany (Brisson et al., 2021) and proved to be a better indicator of lightning occurrence than the commonly used convective available potential energy times precipitation (CAPE x PREC) parameterization (Romps et al., 2014).

Recently, an effort was made to utilize both HAILCAST and LPI to study hailstorms occurring in Croatia using WRF (Malečić et al., 2022) and over the broader Alpine-Adriatic region using COSMO (Cui et al., 2023) models. More specifically, Cui et al., (2023) selected eight days with severe convection over the Alpine-Adriatic region based on the observed impacts and underlying synoptic forcing to study the mechanisms responsible for severe weather effects. Adopting the process-oriented approach, the authors found that both HAILCAST and LPI successfully reproduced observed hail and lightning characteristics over a broad range of synoptic situations.

Motivated by the promising results, this study aims to complement the valuable previous research on hailstorm simulation over the topographically complex Alpine-Adriatic region (Figure 1a). More specifically, we employ these two diagnostic tools in km-scale models (COSMO and WRF) to simulate hailstorms analyzed by Cui et al. (2023), with a geographical focus on Croatia and Switzerland. By analyzing eight hail cases using two models, we aim (i) to conduct a systematic and quantitative evaluation of the model's performance for all cases, and (ii) to identify the robustness of HAILCAST and LPI results produced by two intrinsically different modeling systems. The outcome of this study reveals information about model biases and the origins of disagreements between the two models in simulating severe storms associated with hail and lightning over the complex Alpine-Adriatic region. Moreover, as for Cui et al. (2023), this study benefits from two valuable but intrinsically different datasets of hail observations, namely spatially discrete, but direct, hail measurements from the hailpad network (and two hailpad polygons) in Croatia and spatially continuous, but indirect, radar estimates on hail occurrence from Switzerland. Therefore, the objectives of this paper can be summarized as follows:

To what extent can HAILCAST and LPI credibly diagnose the occurrence of hail and lightning?

How do simulations with two distinct models but the same diagnostic hail and lightning modules differ from each other?

This paper is organized as follows: Section 2 describes the selected hail events and observational data used to evaluate COSMO and WRF models. An overview of the model setups and evaluation approach is indicated in Section 3. The results are presented and discussed in Section 4. The conclusions and remarks are given in Section 5.

2 Selected hailstorms and observational data

Eight days with severe convection occurring over the Alpine-Adriatic region already analyzed in Cui et al. (2023) are selected for the analysis. Out of eight selected days, hailstorms were observed in Croatia and/or Switzerland during seven of these days. During the day without observed hail in the central and eastern Alps, intense precipitation over the Alps leading to severe flooding is reported. Hailstorms are selected based on their intensity, measured by their impact or the kinetic energy recorded with hailpads. Besides that, an attempt was made to select hailstorms occurring over a set of synoptic and mesoscale situations to assess the abilities of both models to reproduce the observed convection during a variety of driving conditions. A

detailed description of synoptic factors contributing to the occurrence of severe weather during the selected days is given by Cui et al. (2023). Moreover, by selecting a day where no hail is observed, the ability of both models to distinguish between convective days with and without hail is assessed. The type of observations available for eight selected events along with a brief description of their impacts are listed in Table 1.

Table 1. Selected convective events over the Alpine-Adriatic region. The type of available observations and a short description of the event's impact is indicated. Radar observations are covering Switzerland while hailpad observations are associated with the hailpad network in Croatia. More information on the impacts of some of these events can be obtained through*http://www.sturmarchiv.ch/index.php/Hagel*.

Date	Observations of hail	Impact
23 July 2009	Radar	A thunderstorm embedded in a cold front hit Switzerland an
1 June 2013	Radar	Significant damage to vineyards in eastern Switzerland.
18 June 2013	No hail observed in Croatia and Switzerland	Intense precipitation over the Alps led to severe flooding in c
25 June 2017	Radar, Hailpads	Large mesoscale convective system observed in Croatia. 74 ir
8 July 2017	Radar	Damage to crops and vineyards in Switzerland.
24 July 2017	Radar, Hailpads	Large hailstones were observed in Croatia. Maximum record
17 May 2018	Hailpads	Non-gradient pressure field over Croatia. Significant kinetic e
30 May 2018	Radar	Significant damage to cars and buildings in central and easte

To assess the model's ability to reproduce the observed severe weather events, several datasets are used. To validate simulated precipitation over the Alpine-Adriatic region, the Final Run of Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) mission (Huffman et al., 2019) dataset is used. IMERG is a globally gridded precipitation product that estimates surface precipitation rates at 0.1° spatial and 30 min temporal resolution. IMERG incorporates satellite microwave precipitation estimates, microwave-calibrated infrared satellite estimates and rain gauge observations. Combining remote sensing and in situ observations, IMERG provides spatially and temporally continuous and homogeneous precipitation estimates over the whole Alpine-Adriatic region.

Further, to assess the ability of LPI to reproduce the observed lightning activity, lightning data from the Lightning Detection Network (LINET) (Betz et al., 2009) is used. With 190 sensors in 31 countries that are up to 250 km away, LINET successfully detects cloud-to-ground and intracloud lightning flashes and differentiates between positive and negative discharges across Europe. Moreover, the LINET network can detect weaker stroke signals with a current amplitude lower than 1 kA. With continuous improvements of the LINET network, the median values of detected current amplitude values had decreased by half from 2009 (Franc et al., 2016), showing significant improvement in the sensitivity detection toward smaller stroke current amplitudes. For most of the European region (Franc et al., 2016; Jelić et al., 2021), the average minimum detectable signal is 0.7 kA, and the median location accuracy error is ±84 m. Here, we considered total lightning information, i.e., we did not differentiate between types or polarities of lightning flashes as LPI presents the overall potential for lightning activity without preferences to the type or polarity of lightning discharges. The total lightning for the examined cases was taken from the 2D database of lightning flashes at a 3 km x 3 km horizontal and 2 min temporal resolution (developed by Jelić et al., 2021) over the domain shown in Figure 1a.

Next, hail detection products from the Swiss radar network (Germann et al., 2015; Willemse & Furger, 2016) operated by MeteoSwiss are used to assess the HAILCAST results. Namely, operationally computed POH product is used. POH indicates the grid-based probability of hail reaching the ground. It is computed following Waldvogel et al. (1979) and Foote (2005) as a difference in height between the altitude of the center of the highest radar bin at which 45 dBZ echo (i.e., Echo Top of 45 dBZ) is found and the height of the freezing level retrieved from the forecasts of the operational numerical prediction model COSMO. POH has been verified using insurance loss data (Morel, 2014; Nisi et al., 2016) and a good agreement between

hail damage and POH >= 80 % was found. The area in which this product is available is indicated in Figure 1b.

Finally, HAILCAST results are assessed against direct hail measurements from the Croatian hailpad network. It consists of (i) hail suppression stations in the continental region of Croatia, (ii) a specially designed hailpad polygon in northwestern Croatia, and (iii) hailpad stations in the northeastern (NE) Adriatic region (Figure 1c). Overall, 590 hailpads on hail suppression stations and 150 hailpads on the polygon with average spacing between hailpads of ~5.5 km and ~2 km, respectively, have been installed and maintained by the Croatian Meteorological and Hydrological Service (Počakal et al., 2009; Pocakal, 2011). Moreover, during the VITCLIC project (https://www.pmf.unizg.hr/geof/en/research/climatology/vitclic) 65 hailpads were installed in Istria (NE Adriatic) in the vicinity of the vineyards. Notably, the Istrian region is not a part of the hail suppression network; therefore, hail observations from these hailpads are not under the potential influence of hail suppression activities.

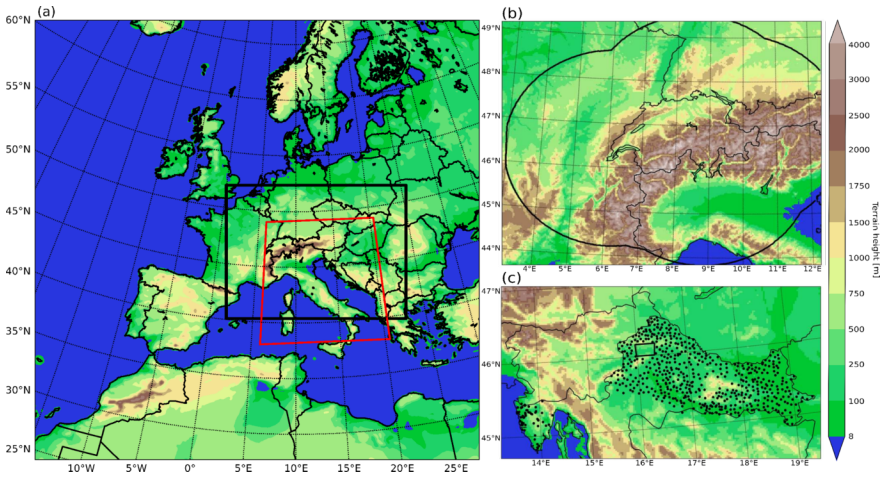

Figure 1 . (a) Terrain height (above sea level) as represented in WRF for the outer 12 km domain. The inner 2.2 km domain is indicated with the black rectangle. The domain where LINET measurements are available is indicated with the red line. (b) Terrain height as represented in WRF for the 2.2 km domain over the Alpine region. The black line indicates the Swiss radar spatial coverage. (c) Terrain height as represented in WRF for the 2.2 km domain over Croatia. The black dots indicate the positions of hailpads.

3 Modeling setup and evaluation approach

3.1 COSMO and WRF setups

Selected hailstorms were simulated using an Advanced Research Weather Research and Forecasting (WRF, version 4.1.5) model (C. Skamarock et al., 2019), and the climate version of the Consortium for Small Scale Modelling (COSMO-crCLIM based on COSMO 5.0) model (Baldauf et al., 2011; Leutwyler et al., 2017; Schär et al., 2020) alongside HAILCAST and LPI. An attempt was made to make a setup of both models as similar as possible. Additionally, one of the hailstorms is simulated using the newest version of COSMO v6.0 model.

The modeling setup consisted of two one-way nested domains with horizontal grid spacing of approximately 12 km (0.11°, 361x361 grid points) and 2.2 km (0.02°, 800x600 grid points for COSMO and 801x601 grid points for WRF) (Figure 1a). Considering the importance of fine grid spacing in the vertical direction (e.g. Fiori et al., 2014) and the sensitivity of sounding derived indices to vertical sampling (e.g. Manzato, 2008), the model setup consists of 65 vertical levels in WRF and 60 vertical levels in COSMO. WRF uses a hybrid sigma-pressure vertical coordinate (Park et al., 2013), while COSMO uses Gal-Chen coordinates. WRF's time step is set to 20 s and 4 s, while COSMO's time step is set to 90 s and 20 s for 12 km and 2.2 km

simulations, respectively. The simulations were initialized and driven at the lateral boundaries using ERA5 reanalysis (Hersbach et al., 2020) at 12 UTC the day before severe convection was observed. It should be noted that unlike for WRF, soil moisture for COSMO was not taken directly from ERA5 reanalysis. Instead, each case simulation was initialized 7 days before the event using the equilibrated monthly mean soil profiles from a 10-year (1999-2008) 12 km COSMO climate simulation (Vergara-Temprado et al., 2020), and let run for 7-days. Then, the model integration started at 12 UTC the day before severe convection was observed using the new soil moisture conditions from the 7 days run. This approach allows for an adjustment of the top soil layers to the conditions of each event.

Grid spacing of approximately 2.2 km allows the model to represent convective processes explicitly (Kain et al., 2006, 2008), and therefore, in the inner domain, no cumulus parameterization is applied. In the outermost domain, convection in WRF is parameterized using the Kain-Fritsch scheme (Kain & Kain, 2004). Other physics options used include the rapid radiative transfer model scheme (RRTM) (Mlawer et al., 1997) for longwave radiation and the Dudhia scheme (Dudhia, 1989) for shortwave radiation. Regarding the planetary boundary layer (PBL) parameterization and microphysics parameterization schemes, MYNN2.5 (Nakanishi & Niino, 2006) and WRF single-moment six-class scheme (WSM6) (Hong & Lim, 2006) are used.

The convection scheme used for the outer COSMO 12 km domain is based on the Tiedtke (1989) with shallow convection parameterized version. Similar to WRF, for the inner domain, explicit convection is applied for shallow and deep convection. The single-moment microphysics scheme (Reinhardt & Seifert, 2006) represents the cloud microphysics with prognostic cloud water, cloud ice, graupel, rain and snow. The radiation scheme is provided by Ritter and Geleyn (1992) with a**?** -two-stream approach.

For the simulation of hail and lightning, HAILCAST (Adams-Selin & Ziegler, 2016) and LPI (Lynn & Yair, 2008; Yair et al., 2010) are used in both models. HAILCAST is a time-dependent hail growth model that provides the forecast of the maximum hailstone diameter at the ground. In our setup, HAILCAST is activated every 5 min on the inner 2.2 km convection-permitting domain if the updraft in a particular grid point exceeds 10 ms^{-1} for more than 15 min. Similarly, we adopt the same formulation of LPI in both WRF and COSMO models. In those formulations the following conditions must be met so that the LPI for a particular grid point is nonzero: (i) vertical velocity in a particular grid point must be greater than 0.5 ms^{-1}, (ii) vertical velocity in adjacent grid points (within a five-grid radius) must be greater than 2 ms^{-1}, and finally (iii) a particular grid point and its adjacent grid points must be in an unstable environment. An unstable environment is defined by analysis of a parameter similar to mixed-layer CAPE obtained by the integration over a 500 hPa layer starting at 50 hPa above ground. More details on these requirements can be found in Brisson et al. (2021). In this study, LPI is computed every 15 min and 15 min fields are stored for both models. More information on HAILCAST and LPI can be found in the Supplement.

3.2 Evaluation approach

When evaluating the results of diagnostic tools such as HAILCAST and LPI against observations, it should be considered that their performance relies on the skill of the convection-permitting model to represent the convection properly. For this reason, the results are evaluated in three sequential phases. First, the model's skill to represent the observed precipitation is evaluated. Second, simulated LPI is assessed against LINET lightning data using the minimum coverage neighborhood verification method (Ebert, 2008). Third, HAILCAST results are evaluated against radar estimates on hail occurrence from Switzerland and direct hail measurements from the Croatian hailpad network using a minimum coverage neighborhood verification method (Ebert, 2008) and upscaled neighborhood verification method proposed by Malečić et al. (2022), respectively.

The first step considered an evaluation of the skill with which models produced the observed convection. Specifically, precipitation simulated in both models is assessed against precipitation estimated by IMERG. Considering that simulated precipitation could be shifted in time compared to the observations, daily aggregated fields of both simulated and observed precipitation are evaluated. Evaluation is done by determining

standard deviations, correlation coefficients and root mean square errors between observed and modeled fields. The results are summarized using Taylor diagrams (Taylor, 2001).

Second, LPI is assessed against lightning observations from the LINET network. Here, we had to consider that it is difficult for a high-resolution simulation to precisely match the observation in space and/or time. However, even simulations that do not precisely match observations in space, time, or even intensity, can still be useful (Ebert, 2008). For this reason, in our verification approach, we are using a minimum coverage neighborhood method. Thus, a useful forecast is defined as the one where lightning is simulated anywhere in the neighborhood of the point where it is observed. Based on this method, a contingency table is built and a symmetric extremal dependence index (SEDI) (Ferro & Stephenson, 2011) is computed. Moreover, here we are varying verification window sizes as well as thresholds for the number of lightning flashes to examine the scale-intensity combination at which high-resolution simulation is useful.

Third, HAILCAST results are assessed against radar products POH and MESHS from Switzerland and hail measurements from the Croatian hailpad network. To evaluate HAILCAST results against radar products, a minimum coverage verification method with varying verification windows sizes is utilized and categorical skill score such as probability of detection (POD), false alarm ratio (FAR) and extremal dependence index (EDI) (Ferro & Stephenson, 2011) are determined. Next, HAILCAST results are assessed against hailpad observation from Croatian hailpad network. Here, to overcome challenges associated with the limited spatial information from hailpad networks and to limit the effect of double penalty that occurs when verifying slightly offset high-resolution forecasts of extremely rare events (Ebert, 2008), an upscaled neighborhood verification method is used (Malečić et al., 2022). This verification methodology is composed of the elements of methods such as point to point, upscaling and a minimum coverage verification method (as described by Ebert (2008)) and is further described by Malečić et al. (2022). Based on this method, a contingency table is built and categorical skill scores are determined.

4 Results and discussion

4.1 Precipitation

The first step in the analysis of results consists of the comparison between simulated and observed precipitation. The comparison refers to the period from 00 UTC to 24 UTC on the day when severe convection was observed. For the evaluation, the IMERG dataset is selected, since it covers the whole domain of interest (domain 2 from Figure 1). When comparing the daily accumulated fields of both observed and simulated precipitation, a generally good agreement between observed and simulated fields can be found (Figure 2). Therefore, both models reproduce the observed precipitation patterns fairly well, even though there are slight local variations. Moreover, the areas with more intense precipitation correspond well between simulated and observed fields. However, a tendency of both models to produce more peaked and more scattered precipitation objects compared to IMERG data can be found in all analyzed cases except 8 July 2017 and 18 June 2013. This tendency could partially be attributed to the horizontal resolution differences between simulated and observed fields (simulated fields are obtained at 2.2 km horizontal grid spacing, while observed fields are obtained at 11 km horizontal grid spacing). On the other hand, for 18 June 2013 and 8 June 2017, both models produced mostly smaller and less peaked objects than observed. In addition, comparing the fields produced by COSMO-crCLIM and WRF, it is noted that WRF tends to produce slightly less peaked precipitation objects than COSMO-crCLIM.

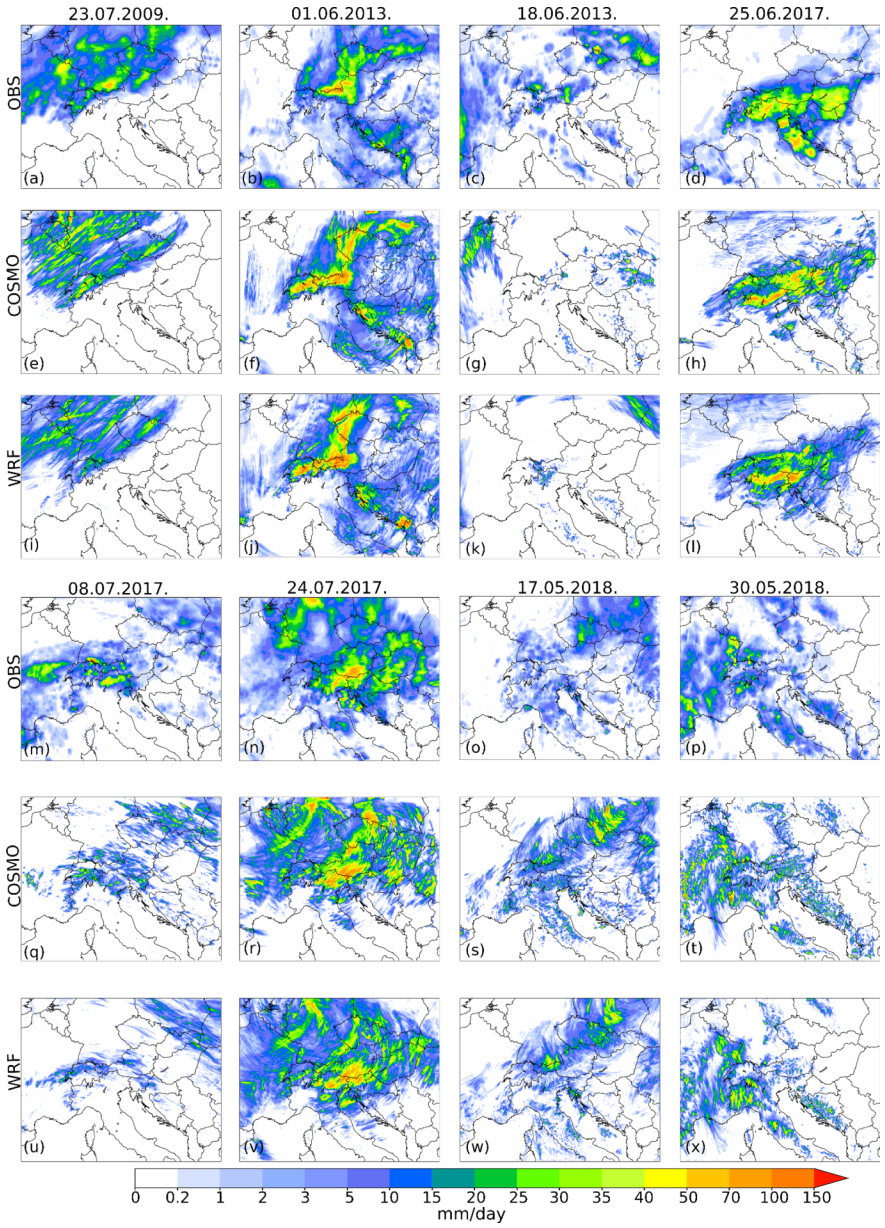

Figure 2 . Accumulated precipitation for the period between 00 and 24 UTC on the day severe convection was observed for the eight case studies. Fields observed by (a-d; m-p) IMERG and simulated by (e-h; q-t) COSMO-crCLIM and (i-l; u-x) WRF are presented.

To further expand and complement this analysis, the hourly accumulated precipitation averaged over the whole inner domain (Figure 1a) is compared between the observations and the models (Figure 3). Considering the discrepancy between the model's and IMERG horizontal resolution, both observed and simulated fields are interpolated to the outer 12 km grid with bounds corresponding to the inner 2.2 km grid presented in Figure 1a. The comparison reveals that both models captured the temporal evolution of precipitation fairly well in all cases, except 25 June 2017. However, it should be noted that some underestimations/overestimations or time shifts might be present, depending on the case analyzed. For 25 June 2017, both models failed to represent the two local maximums of precipitation observed in the early morning and evening hours. In-

terestingly, during most cases, both models simulate comparable or slightly larger amounts of precipitation compared to the observations. An underestimation of precipitation is present only for 18 June 2013 and 8 July 2017 cases as already noted in the analysis of precipitation spatial patterns. Therefore, for the events analyzed, WRF produces, on average, less precipitation compared to COSMO-crCLIM.

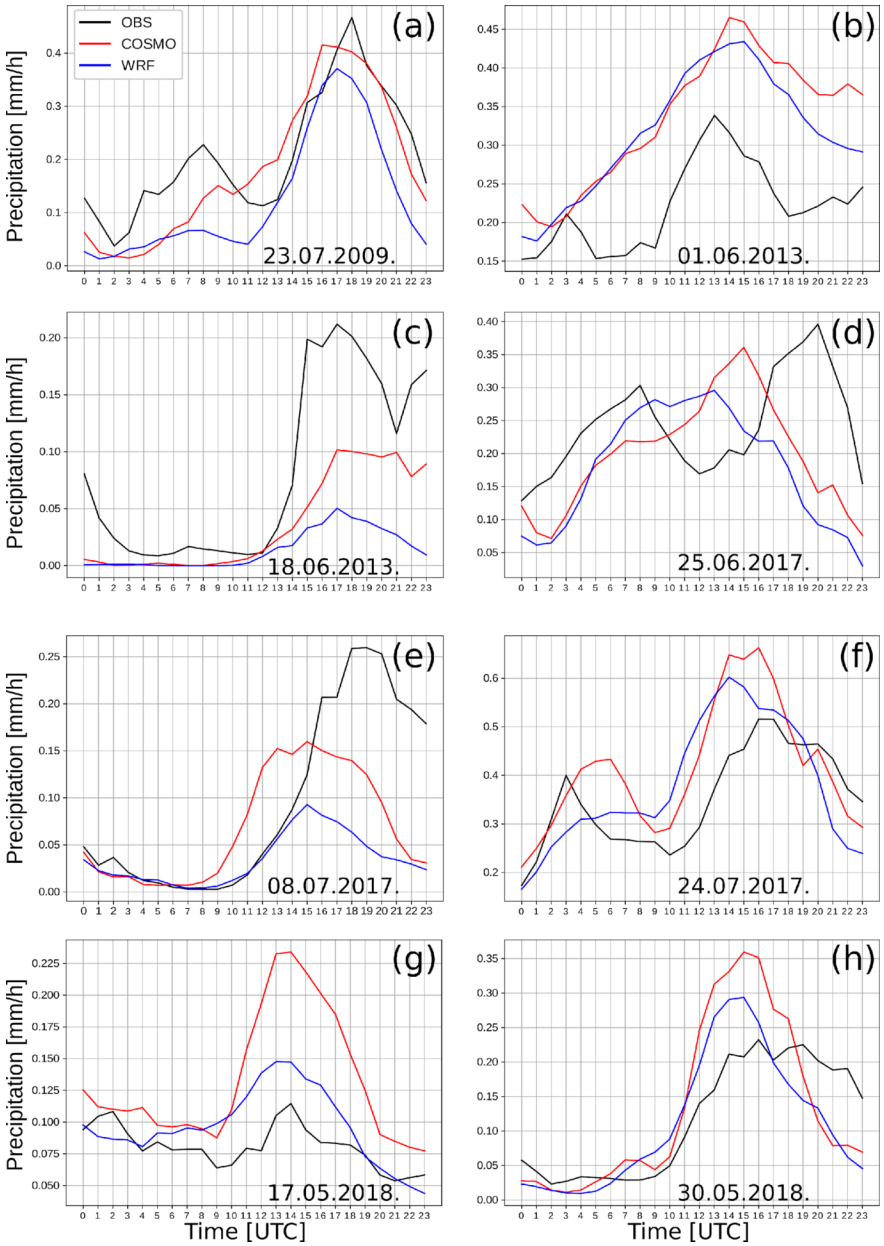

Figure 3 . Hourly accumulated precipitation obtained from IMERG (black) observations and simulated by COSMO-crCLIM (red) and WRF (blue) for all eight cases (a)-(h). The observed and simulated hourly precipitation amounts are interpolated to a 12 km grid and then averaged over the 2.2 km domain.

The precipitation results are further compared quantitatively using Taylor diagrams (Taylor, 2001b). To account for possible time shifts between observed and simulated convection, and different horizontal resolutions between IMERG and the models, Taylor diagrams are obtained by interpolating the daily accumulated

observed and simulated fields to a common 12 km grid. As shown in Figure 4 both models perform similarly, although larger differences in standardized deviations are found for 18 June 2013 and 8 July 2017 case. Both models show similar correlation coefficients between simulated and observed fields. Looking at the median performance for all cases together, it can be concluded that both models perform similarly in terms of simulating the observed precipitation with standardized deviations of 1.14 and 1 mm/day, correlation coefficients of 0.48 and 0.46 and root mean square errors of 1.14 and 1.04 mm/day for COSMO-crCLIM and WRF, respectively.

Overall, given the presented analysis it can be concluded that both models successfully represent the observed precipitation, and WRF tends to simulate less precipitation than COSMO-crCLIM.

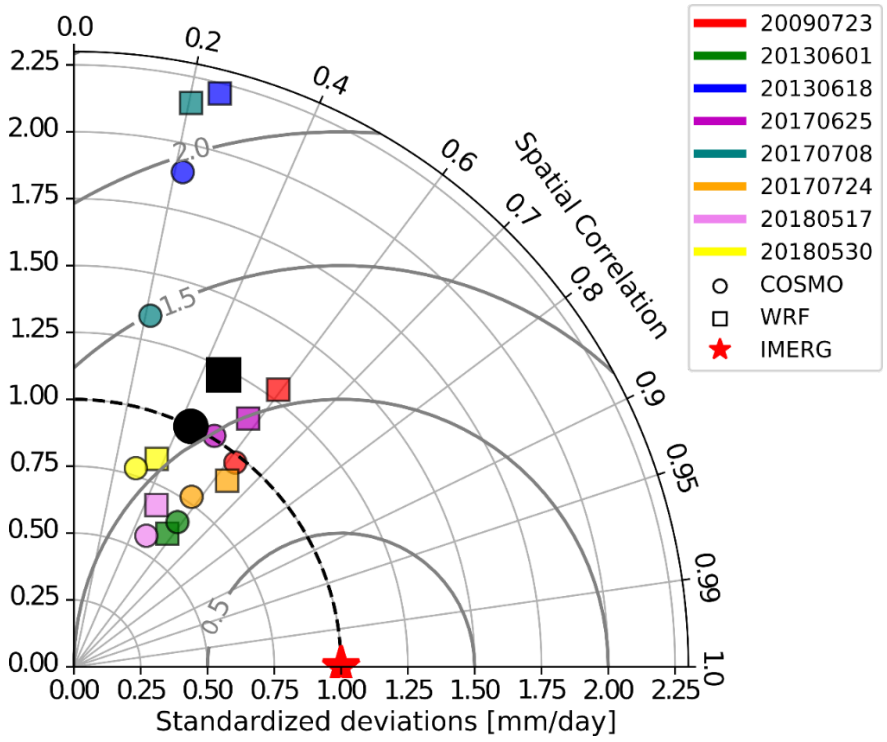

Figure 4 . Taylor diagram showing the performance of COSMO-crCLIM (circle) and WRF (square) when simulating daily accumulated precipitation observed by IMERG (red star). The performance for each case is indicated by colored markers, while the corresponding median values are indicated by black markers.

4.2 Lightning potential index results

The second step of the evaluation considers the assessment of LPI [J/kg] against the observed number of lightning flashes from the LINET network. LPI indicates the potential for lightning activity, and as such, it is not directly connected to the observed number of lightning flashes. To make a direct comparison between LPI and the observed number of lightning flashes, a conversion of LPI to the number of lightning flashes following Brisson et al. (2021) and Malečić et al. (2022) is done. Conversion assumes a linear relationship between the LPI and the observed number of lightning flashes as well as the existence of the threshold value of LPI for which a lightning flash is produced, such that:

$$\mathrm{LPI}_{\mathrm{adj}} = \begin{cases} 0, & \& LPI \leq t \\ k \bullet LPI + l, & \& LPI > t \end{cases} \quad (1)$$

where LPI_{adj} [$km^{-2}h^{-1}$] denotes the adjusted LPI, i.e., LPI converted to the number of lightning flashes, parameter t denotes the minimum value of LPI for which a lightning flash is produced, $k$and l represent the parameters of a straight line. The parameters$t$, k and l are iterated across [0, 20], [0, 10], [-20, 20] intervals, respectively. For every combination of parameters t , k , and l , hourly means of LPI_{adj} are calculated. Then, a distribution function of both simulated and observed hourly means of lightning flashes during all cases is determined. Further, a root mean square error (RMSE) between the two discussed distributions is found. The optimal combination of parameters is the one that minimizes the RMSE. Here, a conversion is done by using $t = 0.045$, $k = 3.3$ and $l = 0.1$ for COSMO-crCLIM and $t = 0.65$, $k = 0.65$ and $l =$ -0.2 for WRF for all cases. The discrepancy in optimal parameter values between WRF and COSMO highlights the discrepancies between LPI produced by COSMO-crCLIM and WRF. Namely, higher values of t and lower values of k associated with WRF indicate that WRF produces higher LPI, contrary to the results obtained for precipitation where WRF produced slightly lower precipitation amounts compared to COSMO. Since LPI is highly dependent on the updraft intensity and hydrometeor fields (as indicated by Equation S2.1), this discrepancy in t , $k$and l values between the two models could be due to the differences in simulated updrafts or differences in microphysical fields produced by COSMO-crCLIM and WRF. These discrepancies will be further discussed in Section 4.4.

Finally, to reduce uncertainties in simulating the temporal characteristics of observed convection (as seen in Figure 3), the daily sums of both LPI_{adj} produced by COSMO-crCLIM and WRF and the observed number of lightning flashes for each case analyzed are presented in Figure 5. Overall, it seems that the general spatial pattern of the observed lightning activity is well reproduced by both models although the simulated fields appear to be more scattered than the observed. This could be partially attributed to the fact that LPI is calculated every 15 min, while the LINET network detects lightning flashes continuously. Moreover, considering all cases, it is noted that the conversion of LPI to lightning flashes is better fitted towards less intense lightning activity. This is explained by the fact that the fit is performed on all grid points: as there are more grid points with low flash counts then intense lightning activity, the fit is intrinsically better for lower flash counts. The discrepancy in fit between lower and higher flash counts is more pronounced during the cases with more intense and widespread lightning activity, i.e., 25 June 2017 and 24 July 2017. Nonetheless, in general, the spatial distribution of lightning, i.e., the distribution of the areas with more and less intense lightning activity, corresponds well between simulated and observed fields, although local discrepancies could be present, depending on the case and model analyzed. Looking at the differences between fields produced by COSMO-crCLIM and WRF, a tendency of COSMO-crCLIM to produce more scattered and less peaked fields can be found.

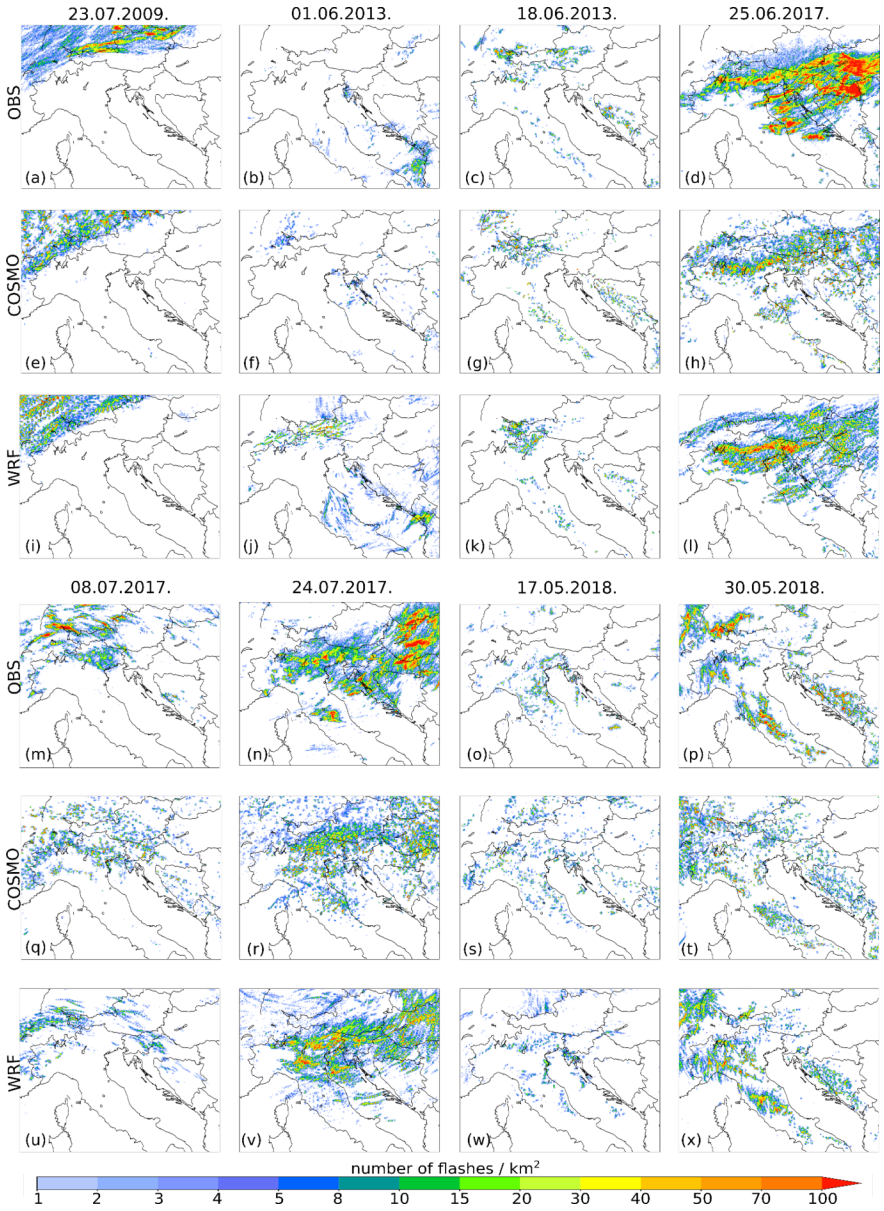

Figure 5 . Simulated and measured lightning flash accumulation in the time window from 00 to 24 UTC on the day with severe convection. Columns denote cases, while rows denote measurements from the (a-d; m-p) LINET network, and fields produced by (e-h; q-t) COSMO-crCLIM and (i-l; u-x) WRF models.

Furthermore, to quantitatively evaluate the capabilities of COSMO-crCLIM and WRF to simulate the observed lightning activity, a minimum coverage method is utilized. To get more robust results, the evaluation is done by aggregating all analyzed cases together and analyzing the daily sums of observed and simulated lightning activity (fields presented in Figure 5). Using the minimum coverage method combined with various radiuses of verification windows and various thresholds for the number of lightning flashes, contingency tables are constructed and SEDI index is calculated (Figure 6). Both models show similar performance which is better for the lower thresholds of lightning flashes. Moreover, we get good performance (SEDI > 0.6) even for more intense thresholds if we consider larger verification window sizes. WRF tends to have higher SEDI values than COSMO-crCLIM towards higher and more localized lightning flashes (bottom right side

of diagrams), which confirms the previous findings that COSMO-crCLIM tends to produce more scattered lightning activity.

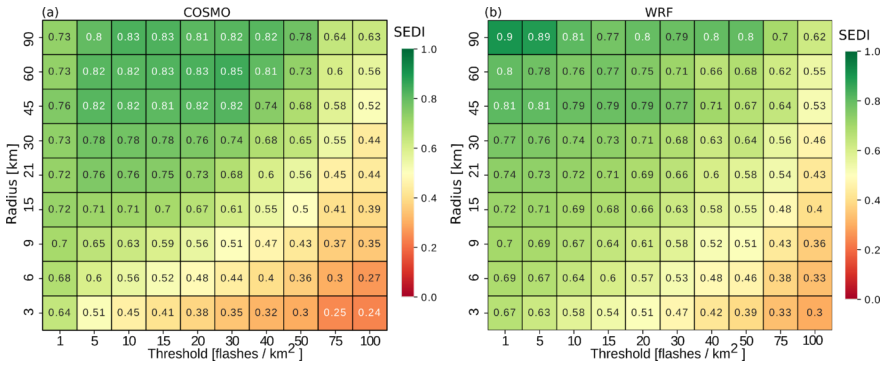

Figure 6 . Performance of (a) COSMO-crCLIM and (b) WRF in simulating the observed lightning flashes. Performance depending on the threshold for the number of lightning flashes and verification window sizes (radius) is indicated in terms of SEDI skill score (shading). The higher/lower SEDI score means better/worse performance of the model, as reflected by the green/red colors.

4.2 HAILCAST results

HAILCAST results are assessed against remote-sensing and ground observations for a period from 00 UTC to 24 UTC on the day severe convection was observed. Moreover, to allow for possible temporal shifts between simulated and observed convection, we aggregate simulated and observed datasets over a 24 h period. First, we perform a qualitative comparison between hail swaths produced by the two models. Figure 7 suggests that both models produced generally similar hail swaths over the same area of interest, although some local discrepancies between simulated hail swaths exist. Despite the overall similarity of the results, a tendency of COSMO-crCLIM to produce more hail in all analyzed cases is apparent. Interestingly, both models correctly reproduced heavy precipitation without hail over the Alps for 1 June 2013, which suggests that both models are able to distinguish intense precipitation events form hail events. However, it should be noted that both models still produced hail over only a few grid points over the Alps.

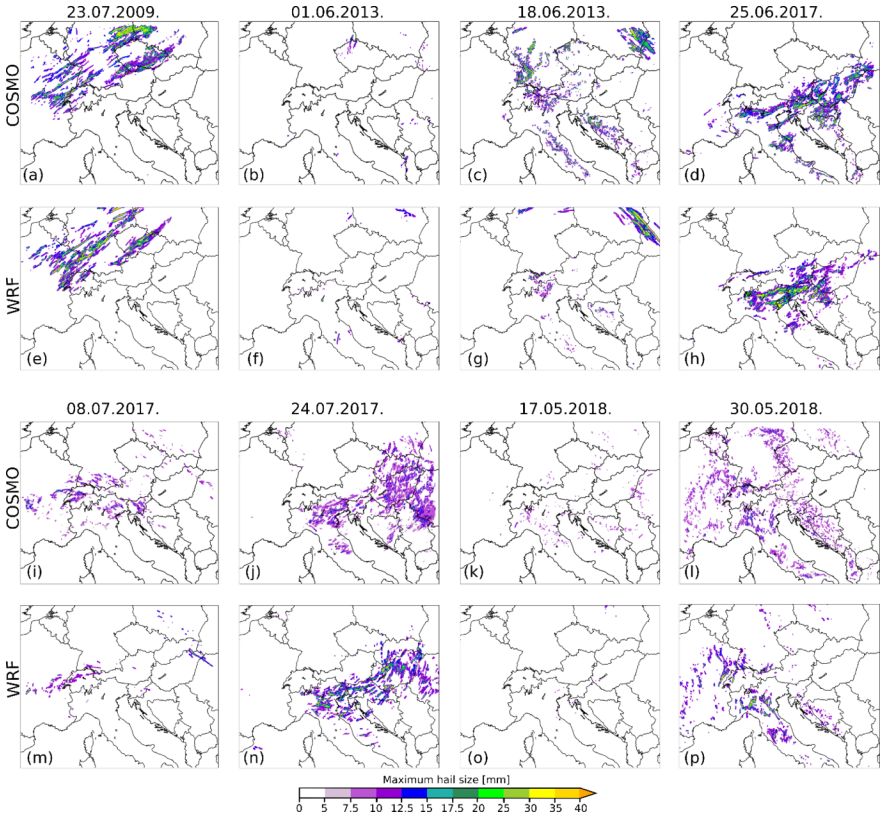

Figure 7 . Maximum hailstone diameters in the time window from 00 UTC to 24 UTC on the day with severe convection simulated by COSMO-crCLIM (a)-(d); (i)-(l), and WRF (e)-(h); (m)-(p).

Further, we compare simulated fields against remote sensing observations of hail in Switzerland. Figure 8 shows the simulated and observed hail swaths over the Alpine region. It is clear that both models can produce hail swaths comparable to those observed, both in the context of the area affected by hail and the shapes of the observed hail swaths. Notably there is not an exact match between simulated and observed fields, as, some deviations are present. Interestingly, in most of the cases analyzed, WRF produces smaller hail swaths than COSMO-crCLIM. On the other hand, WRF simulates more grid points having maximum hailstone diameters greater than 35 mm.

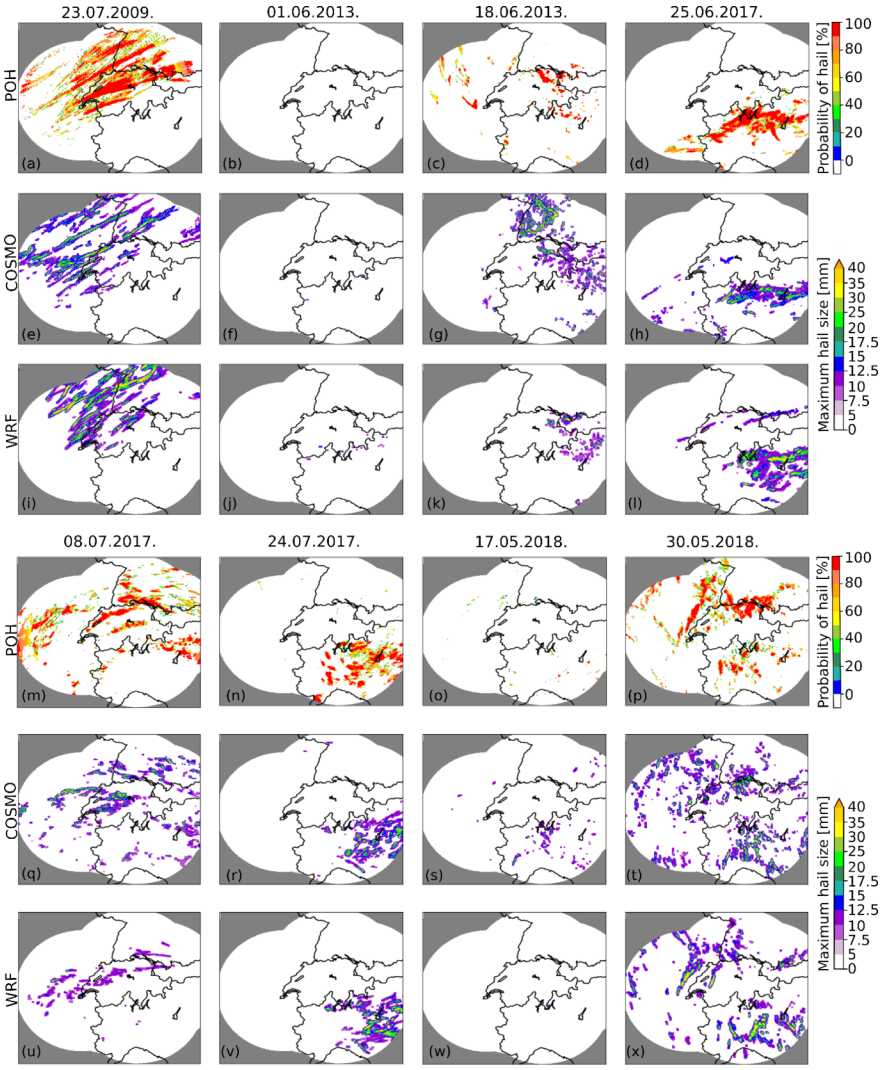

Figure 8 . Daily maximums of the hailstone diameter simulated by (e)-(h); (q)-(t) COSMO-crCLIM and (i)-(l); (u)-(x) WRF and daily maximums of (a)-(d); (m)-(p) POH radar product.

Next, we assess simulated fields against another source of hail observations – hailpad measurements from the Croatian hailpad network. Out of eight cases with severe convection over the Alpine-Adriatic region, hailpads in Croatia recorded hail on only three of those days (25 June 2017, 24 July 2017, 17 May 2018). For these days, simulated hail swaths with indicated impacted hailpads are presented in Figure 9. There is a generally good agreement between observed and simulated hail produced by both models. Most of impacted hailpads lie in the area of simulated hail. However, both models exhibit a certain number of false alarms, i.e. hail is not observed, but the model simulated hail. Notably, some of these false alarms could be attributed to the limited spatial information on hail occurrence provided by the hailpad network. Unlike radars, the hailpad network provides information on hail occurrence only at the exact position where the hailpad is installed. In theory, hail could easily occur anywhere between the two hailpads and be left unrecorded. Nonetheless, the greatest number of false alarms is present for the 24 July 2017 case with WRF producing more false alarms than COSMO-crCLIM. Surprisingly, both models successfully reproduced even a highly localized hailstorm occurring on 17 May 2018 with a flat surface pressure distribution over the north-eastern Adriatic (Cui et al., 2023), although it should be noted that, unlike WRF, COSMO-crCLIM produces a few spurious false signals

in the continental part of Croatia. Moreover, two of the analyzed cases, namely 25 June 2017 and 24 July 2017, were previously analyzed in Malečić et al. (2022). Even with different modeling settings compared to Malečić et al. (2022), i.e., different domains, horizontal resolutions, input data or HAILCAST activation time, WRF-HAILCAST produced similar hail swaths in both studies.

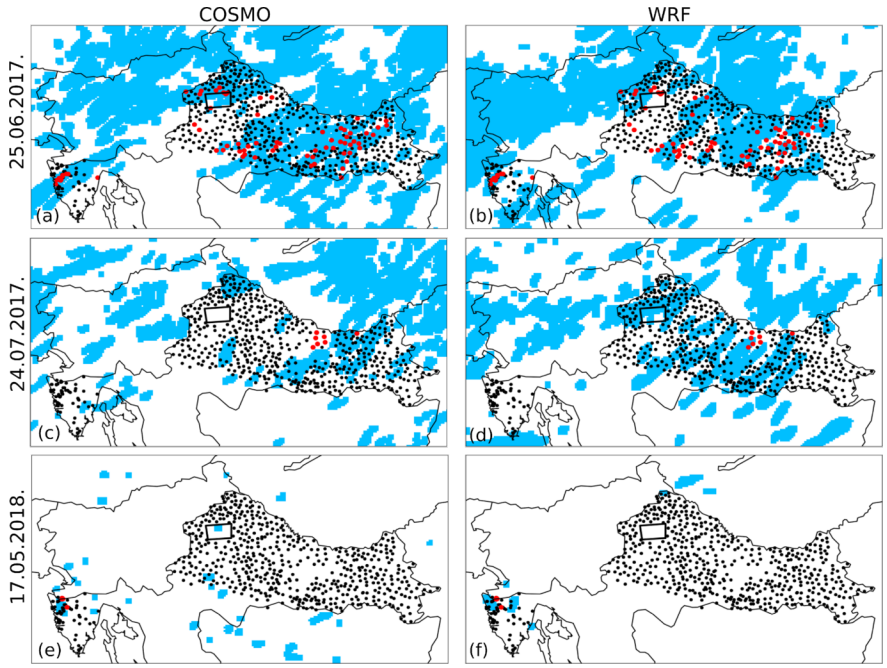

Figure 9 . Simulated and recorded hail during the three cases with hail in Croatia. The shaded blue area represents simulated hail swaths (maximum hailstone diameter larger of equal to 5 mm) from 00 to 24 UTC on the day hail was observed. The position of hailpads is indicated with black dots. Impacted hailpads are marked with red circles. The position of a densely populated hailpad polygon is marked with a black rectangle and the stations within the polygon are colored only if the hail was observed at that specific station. Fields produced by (a, c, e) COSMO-crCLIM and (b, d, f) WRF are presented.

To quantitatively describe the results, simulated fields are evaluated against observations. To get more robust results, the evaluation is done by aggregating all analyzed cases together and analyzing the daily maximums of observed and simulated fields. More specifically, the fields presented in Figure 8 are assessed against POH $>$ 80 % signals, as the region corresponding to POH $>$ 80 % is highly probable to have hail on the ground (Nisi et al., 2016). For the analysis, POH product is interpolated to the model's grid and observed and simulated fields are evaluated using a minimum coverage verification method with varying verification window sizes, as already described in Section 3.3. Obtained performance diagrams presented in Figure 10a-b reveal that COSMO-crCLIM performs better in terms of POD and EDI skill scores for all considered verification window sizes. On the other hand, WRF performs better in terms of FAR for all verification window sizes except the one corresponding to 30 grid points. These findings could be attributed to the fact that COSMO-crCLIM produces more hail compared to WRF which leads to higher POD and FAR values. Summarizing the insights obtained in Figure 8 and Figure 9a-b, it seems that COSMO-crCLIM produces hail swaths more similar to those observed over the Alpine region.

Next, simulated fields are evaluated against measurements from the hailpad network in Croatia using an upscaled verification method (as described in Section 3.3 and Malečić et al. (2022)). The obtained performance diagrams (Figure 10c-d) show similar performance between the models. High POD values for larger verification window sizes indicate that models simulated hail where it was observed. However, unlike the

results connected with radars (Figure 10a-b), FAR values associated with hailpad network are much higher. That could be connected to the potential tendency of the model to overestimate the area affected by hail, if not also to the lack of spatially continuous information on hail occurrence in Croatia. Notably, there is a great contribution to the FAR values from the case on 24 July 2017 where both models produced a lot of false alarms. Interestingly, the same case, 24 July 2017, was also weakly represented in Malečić et al. (2022) with a lot of false alarms indicating a low predictability of the atmospheric conditions leading to the initiation and evolution of the observed convection.

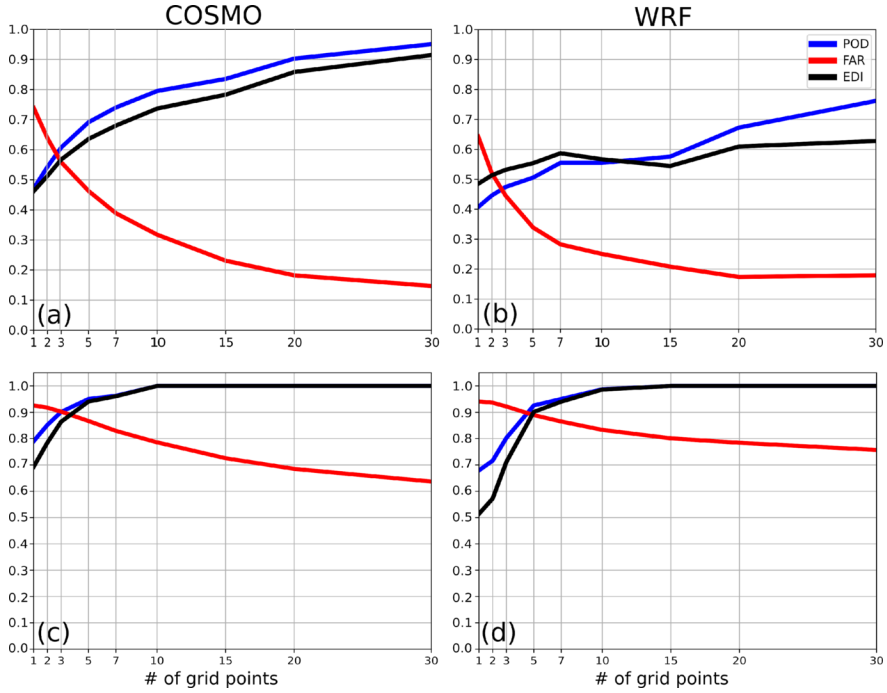

Figure 10 . Performance of (a) COSMO-crCLIM and (b) WRF to simulate hail swaths as observed by the probability of hail (POH) radar product, and performance of (c) COSMO-crCLIM and (d) WRF in simulating hail as observed by the Croatian hailpad network in terms of probability of detection (POD, blue), false-alarm rate (FAR, red) and extremal dependence index (EDI, black) skill scores. Performance depending on the verification window size is presented.

Hailpad networks, besides delivering information on hail occurrence, also provide information on hailstone sizes on the ground. Based on this information, a comparison of simulated and observed maximum hailstone diameters is performed. To account for possible spatial shifts between observed and simulated fields, a neighborhood inside a radius of 5 grid points (roughly corresponding to 12 km) of each impacted hailpad is scanned. The maximum simulated hailstone diameter inside this area is compared to the observed maximum hailstone diameter (Figure 11a). Both models underestimate the occurrence of smaller hailstones (diameters of 5-10 mm and 10-20 mm), and overestimate the occurrence of larger hailstones (diameters larger than 20 mm). However, when analyzing such results, one should proceed with caution, as it has been known from previous studies that hailpads are unlikely to record the largest hailstones given that they cover only 0.25 m^2 (e.g. Smith & Waldvogel, 1989). Indeed, to obtain a more realistic comparison between simulated and observed maximum hailstone diameter, it would be better to use the information on hailstone size observed by an observer – data which was not available for this study. Nonetheless, some tendencies could be extrapolated from the present comparison. Namely, COSMO-crCLIM mostly simulates hailstones in the 20-30 mm category, while WRF mostly simulates hailstones in the 30-50 mm category. Interestingly, WRF was able to reproduce an observed hailstone larger than 50 mm. Those differences between the two models

are further confirmed if we compare the distribution of maximum hailstone sizes over the whole domain and all cases analyzed (Figure 11b). Here, it is clear that COSMO-crCLIM produces more hailstones in the 5-10 mm category than WRF, while WRF tends to produce more larger hailstones. These findings are consistent with Cui et al. (2023) who showed that COSMO tends to produce too many small hailstones and not enough larger ones compared to MESHS radar product.

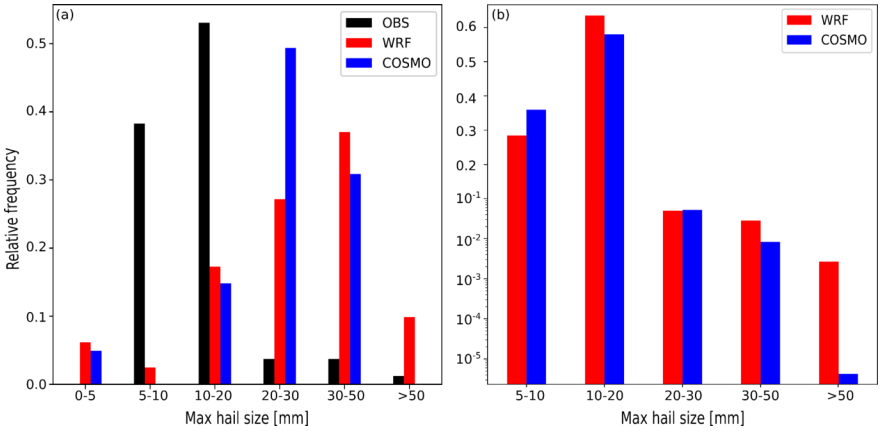

Figure 11 . (a) Relative frequency of maximum recorded hailstone size from hailpads (black) and simulated maximum hailstone size by COSMO-crCLIM (blue) and WRF (red), and (b) relative frequency of simulated maximum hailstone sizes over the whole domain for COSMO-crCLIM (blue) and WRF (red). Histograms are normalized by dividing the count of hailstone sizes in each category with the total observed number of hailstones. To better depict differences between models and runs, the$y$ -axis in (b) is partly linear and partly logarithmic.

To summarize the results, HAILCAST, whether integrated in WRF or COSMO, is able to reproduce observed hail swaths over the Alpine-Adriatic region as already reported by previous studies (Trefalt et al., 2018; Manzato et al., 2020; Raupach et al., 2021; Malečić et al., 2022; Tiesi et al., 2022; Cui et al., 2023). Further by comparing the simulated and observed maximum hailstone diameters, we find that both models tend to overestimate the maximum hailstone size compared to the Croatian hailpad network. Given the limitations of hailpad networks to record the largest hailstones, a detailed analysis with maximum hailstone sizes observed by an observer needs to be conducted to get more justifiable conclusions. Although the results are similar between the two models, some differences can be found. Overall, COSMO-crCLIM tends to produce larger areas covered by hail but smaller hailstones compared to WRF. Based on the analyzed cases, WRF tends to produce smaller areas covered by hail and wider spectra of hailstone sizes, while COSMO-crCLIM shows better performance in terms of simulating hail where it was observed over the whole domain, and the simulated hailstone sizes are somewhat closer to the observed ones in Croatia.

4.4 Differences between models and model internal variability

The analysis reveals that WRF tends to produce less precipitation, smaller hail swaths but higher values of LPI and more large hailstones compared to COSMO-crCLIM. Here, we study the potential origins of these differences, and consider the role of model internal variability in our results. For this reason, we form an ensemble of simulations with different initialization times for one of the cases with widespread hail and lightning across the Alpine-Adriatic region, namely the 30 May 2018. Both models were initialized at 06, 12 and 18 UTC the day before hail was observed (29 May 2018). Additionally, we utilize a simulation of the newest version of COSMO, namely COSMO 6.0, to further increase the ensemble size increase the robustness of our conclusions. The simulation using COSMO 6.0 is initialized at 12 UTC the day before hail was observed.

First, we analyze simulated daily precipitations fields between ensemble members (Figure 12). It is clear that

all ensemble members produce precipitation patterns similar to the observed. Moreover, there is a greater difference in the fields produced by COSMO and WRF than between the members of the same model. This finding suggests that the differences in results are rather linked to systematic differences between models than to the model internal variability. However, it should be noted that differences between model members exist. Interestingly, both versions of COSMO produced comparable fields, although local differences are present that may be relevant for forecasting applications. For instance, the observed heavy precipitation along the French-German border is largely or completely missed in two of the COSMO ensemble members, but present in the member initialized at 06 UTC. This highlights the need for using ensemble techniques in numerical weather prediction (NWP) applications (Klasa et al., 2018).

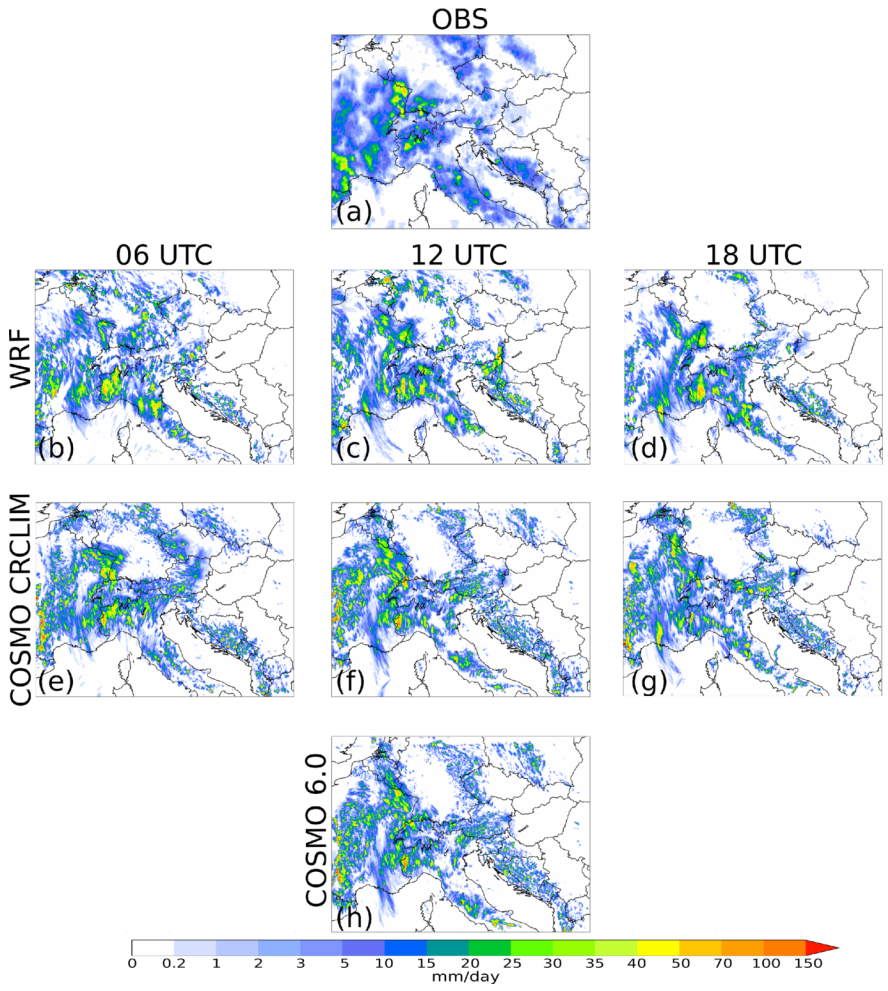

Figure 12 . Accumulated precipitation for the period between 00 and 24 UTC on 30 May 2018. From top to bottom rows are (a) IMERG observation and simulated fields using (b-d) WRF, (e-g) COSMO-crCLIM, (h) COSMO 6.0. The columns from left to right represent the simulations initialized at 06, 12 and 18 UTC on the day before the event, respectively.

Similarly, daily maximums of LPI produced by each of the ensemble members are compared against daily sums of the number of lightning flashes from the LINET network. Here we use the raw LPI instead of the adjusted LPI because we want to avoid filtering the signals to have a direct comparison. It is clear that both models reproduced the area affected by lightning fairly well, although WRF produces higher values of LPI. This finding is consistent with the above results for which we applied a higher threshold for WRF to convert

LPI to the observed number of lightning flashes. Although there are differences in LPI between the model members, there are larger differences in LPI produced by the two different models. This suggests that the conclusions that WRF produces higher LPI than COSMO could be due to differences in simulating convection rather than to the model internal variability. Interestingly, COSMO 6.0 produced LPI in agreement with COSMO-crCLIM, although with slightly higher values.

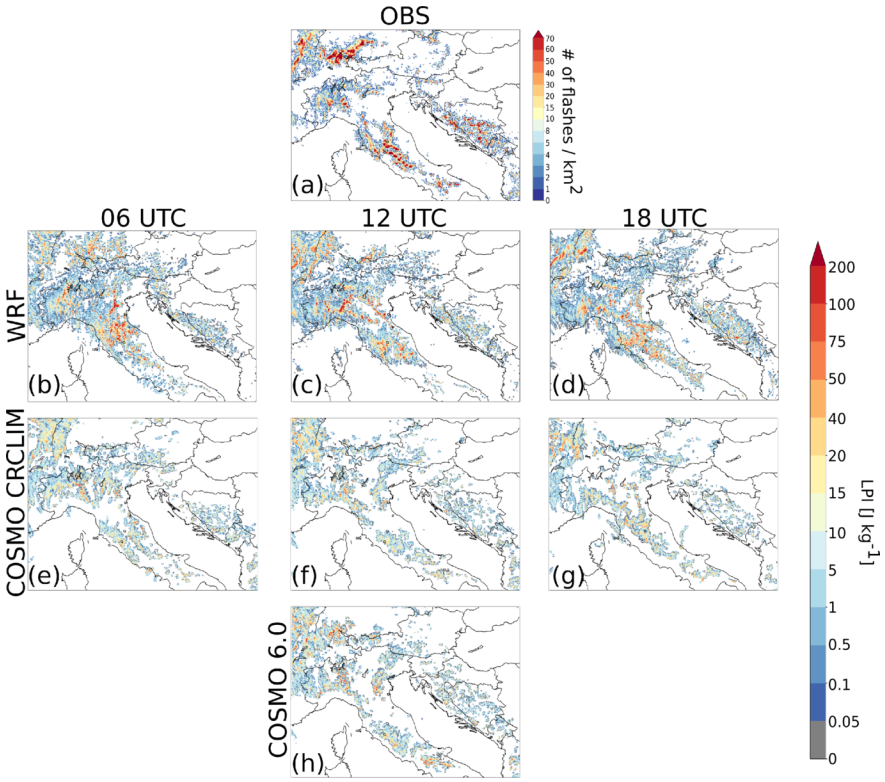

Figure 13 . (a) Daily sum of the observed lightning flashes by the LINET network for the period between 00 and 24 UTC on 30 May 2018. Daily maximum of LPI produced by (b-d) WRF, (e-g) COSMO-crCLIM, and (h) COSMO 6.0. The columns represent simulations initialized at with 06, 12 and 18 UTC on the day before the event.

By comparing daily maximums of the hailstone diameters produced by ensemble members for both spatial (Figure 14) and cumulative distribution (Figure 15), similar conclusions are found. Figure 14 shows that the simulated fields are overall similar, although WRF produces less hail compared to COSMO regardless of the initialization time. Hail produced by the same model but different initialization times (06, 12 and 18 UTC) and different model versions (COSMO-crCLIM and COSMO 6.0) is more similar than hail produced by different modeling systems (WRF vs COSMO). Additionally, when comparing simulated maximum hailstone diameters (Figure 15), we notice that, for hailstones smaller than 30 mm, the differences between COSMO and WRF are within each model internal variability. However, for hailstones larger than 30 mm, not only the differences between models becomes larger than model internal variability, but also it is clear that, out of all ensemble members, only WRF produced hailstones larger than 50 mm. This gives confidence to the overall conclusion that WRF produces more of the larger hailstones than COSMO (whatever its model version).

It should be noted that these findings are valid for one case only, and since the magnitude of the internal variability depends on the synoptic situation, model configuration, region and season (e.g. Lavin-Gullon et al., 2021), more cases should be analyzed to get more robust conclusions. Nonetheless, our results are encouraging and indicate that hailstorms occurring over the Alpine-Adriatic region can be simulated well

using km-scale models, while the variability among different models and initialization times indicates that one could benefit from employing multi-model ensembles when simulating these events in an NWP context.

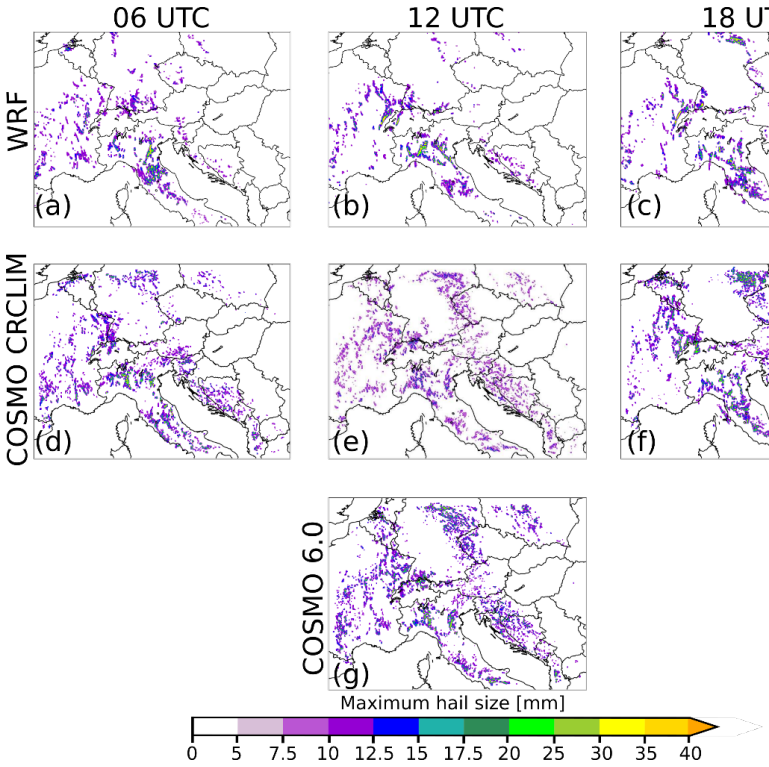

Figure 14 . Daily maximum of hailstone size for the period between 00 and 24 UTC on 30 May 2018 simulated by (a-c) WRF, (d-f) COSMO-crCLIM, and (g) COSMO 6.0. The columns represent the simulations initialized at 06, 12 and 18 UTC on the day before the event.

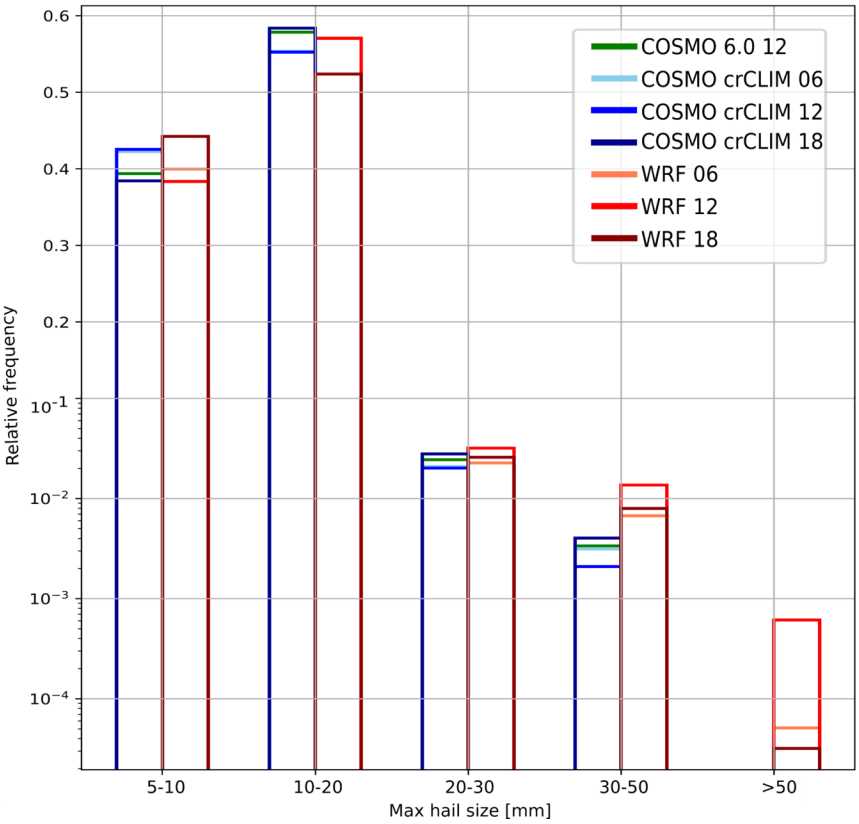

Figure 15 . Relative frequency of the simulated maximum hailstone diameters over the whole domain for COSMO-crCLIM initialized at 06, 12 and 18 UTC (shades of blue), WRF initialized at 06, 12 and 18 UTC (shades of red) and COSMO 6.0 initialized at 12 UTC (green). The histograms are normalized by dividing the count of hailstone sizes in each category with the total number of grid points where hail occurs. To better depict differences between models and members, the y-axis is partly linear and partly logarithmic.

Further, considering the importance of the updrafts, as well as solid and liquid hydrometeors in the LPI and HAILCAST formulations, vertical profiles of these variables are compared among the models. Figure 16 presents vertical profiles averaged over time and grid points that have LPI greater than 0 during the 30 May 2018 case. The models produced different distributions of solid and liquid hydrometeors inside thunderclouds. Namely, WRF produces higher cloud water mixing ratios compared to both versions of COSMO and higher rain water mixing ratios compared to COSMO-crCLIM, but lower rain water mixing ratios compared to COSMO 6.0. Similarly, both versions of COSMO produce higher ice and graupel water mixing ratios, but lower snow water mixing ratios. Since in the LPI formulation the ratio between solid and liquid hydrometeors inside a thundercloud is more important than their exact values, total liquid water and ice fractional liquid ratio terms from the LPI formulation (Q_L and $Q_I$terms from Equation S2.2; S2.3) are computed alongside a dimensionless parameter ε representing the scaling factor for the updraft in the LPI formulation (Equation S2.1). ε obtains maximum values when total liquid water and ice fractional mixing ratios are equal (Equation S2.2). Surprisingly, analysis of ε shows no apparent discrepancies between the models even though there are some differences in Q_L and Q_I parameters. Interestingly, there is a difference in the simulated updrafts, i.e., both COSMO versions simulate, on average, weaker and higher updraft cores compared to WRF. WRF simulates stronger updrafts with the updraft core exactly at the position of the maximum value of ε coinciding with the region with nearly equal amounts of solid and liquid hydrometeors. Since the presence of both solid and liquid hydrometeors is important for lightning and hail growth processes,

it is not surprising that WRF simulates higher LPI and more large hailstones compared to COSMO. On the other hand, the updraft core in both versions of COSMO is in the region with much more solid than liquid hydrometeors which is not as favorable for lightning or hail growth processes. This could be the reason why COSMO simulates lower LPI values and higher amounts of smaller hailstones. Interestingly, COSMO 6.0 simulates stronger updraft cores compared to COSMO-crCLIM which could explain why COSMO 6.0 produces slightly higher LPI values than COSMO-crCLIM (Figure 13) and shows a tendency to produce slightly larger hailstones compared to COSMO-crCLIM (Figure 15a).

Several studies reported a sensitivity of hail and lightning related variables such as updrafts and graupel mixing ratios on the choice of microphysics (Lagasio et al., 2017b; Trefalt et al., 2018; Manzato et al., 2020; Sokol & Minářová, 2020; Raupach et al., 2021), a combination of microphysics and planetary boundary layer parameterization scheme (Malečić et al., 2022), and large-scale forcing and initialization time (Manzato et al., 2020). Thus, different models with different configurations can produce large variability for different cases which encourages the use of a multi-model and/or multi-physics ensemble to analyze or forecast such events and to investigate the response of such events to further warming of the atmosphere.

Likewise, differences in updrafts strength and updraft structure strongly depend upon the dynamical core of the models. Such differences can objectively be assessed using kinetic energy spectra (W. C. Skamarock, 2004). While both models considered here have similar dynamical cores using the split-explicit approach, there are significant differences in terms of advection schemes. The role of model formulation for heavy summer convection over Europe has recently been investigated in an intercomparison of the COSMO and the ECMWF-IFS models (Zeman et al., 2021). Results revealed a strong sensitivity with respect to the dynamical core (split-explicit versus spectral) but also with respect to time-step size as well as (explicit or implicit) numerical diffusion.

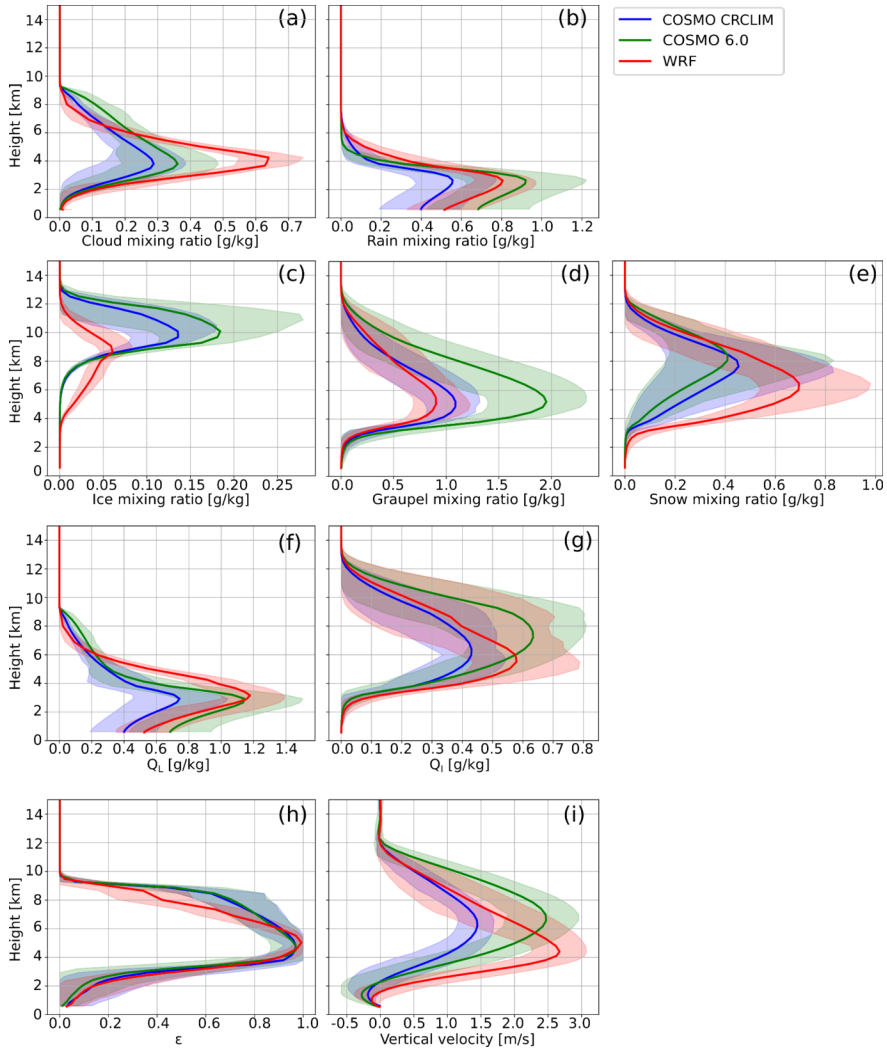

Figure 16 . Vertical profiles of (a) cloud, and (b) rain mixing ratio; (c) ice, (d) graupel, (e) snow, (f) total liquid water, and (g) ice fractional mixing ratio; (h) updraft scaling parameter; and (i) vertical velocity as simulated by COSMO-crCLIM (blue), COSMO 6.0 (green) and WRF (red) models. The lines indicate the mean values across all grid points with LPI > 0, while the shading indicates the range between 5^{th} and 95^{th} percentile.

5 Conclusions

Hail and lightning, which are damaging and relatively frequent phenomena over the Alpine-Adriatic region, still remain difficult to model. Thus, this study employed two km-scale models, namely COSMO and WRF, with hail (HAILCAST) and lightning (LPI) diagnostic tools to simulate eight severe convective events occurring over the Alpine-Adriatic region. This study has been encouraged by and complement the two studies performed on WRF (Malečić et al., 2022) and COSMO (Cui et al., 2023) separately. The main aim was to analyze the robustness of HAILCAST and LPI results produced by the two different modeling systems, to explore their differences and to systematically and quantitatively evaluate the performance of each model. The main conclusions from this analysis can be summarized as follows:

- Both models reproduced the observed precipitation patterns and amounts, but WRF tended to produce slightly lower precipitation amounts than COSMO. Moreover, temporal evolution of precipitation is

captured well in most cases.

- Both models showed good performance in reproducing the observed lightning activity despite WRF's tendency to simulate higher LPI values.
- Simulated hail swaths produced by the two models are overall similar, but a tendency of COSMO to produce more hail than WRF was found. Both models showed a good performance in reproducing hail observed by radar estimates over Switzerland and in-situ measurements over Croatia, although COSMO performed slightly better than WRF. Moreover, both models, on average, overestimated observed maximum hailstone diameters. Interestingly, a tendency of WRF to produce larger hailstones compared to COSMO was revealed, which is attributed to differences in model structures.

Furthermore, the potential origins of differences between the models and their internal variability are studied by employing an ensemble of simulations for one case with varying initialization times (06, 12, 18 UTC) using both COSMO and WRF models. Additionally, a simulation initialized at 12 UTC performed with the newest version of COSMO, namely COSMO 6.0, is added to the ensemble. The results for precipitation, lightning and hail are overall similar, but differences depending on the initialization time and modeling system exist. Moreover, the tendencies of WRF to produce less precipitation, smaller hail swaths but higher values of LPI and more large hailstones are present regardless of the initialization time. A careful analysis of the most important variables for hail and lightning formation, namely updrafts and hydrometeor mixing ratios, revealed that the analyzed modeling systems produce, on average, different distributions of updrafts, solid and liquid hydrometeors inside thunderclouds. The different distributions of updrafts, solid and liquid hydrometeors inside thunderclouds could be linked to differences in model structures.

In conclusion, we show that atmospheric conditions leading to hailstorm formation and evolution are well simulated using state-of-the-art km-scale modeling systems. Moreover, diagnostic tools such as HAILCAST and LPI have a great potential for real-time forecasting and climatological assessment of hail and lightning in current and future climate. However, the variability of the results depending on the modeling system used encourages the use of a multi-model and/or multi-physics ensemble when modeling such events. Additionally, in an operational setting, data assimilation could further improve the predictability of such extreme events. Despite the promising results, it should be noted that this study is based on a small number of cases. To get statistically more robust conclusions a larger number of hailstorms needs to be analyzed. Moreover, this study would highly benefit from employing other data sources of hail observations covering the whole Alpine-Adriatic region. Nonetheless, given all limitations, this study represents the first attempt to systematically analyze and evaluate the performance of two intrinsically different km-scale modeling systems to reproduce the main characteristics of multiple hailstorms occurring over the Alpine-Adriatic region.

Acknowledgments

This research is enabled by the SWALDRIC (IZHRZO-180587) project, which is financed within the Croatian-Swiss Research Program of the Croatian Science Foundation and the Swiss National Science Foundation with funds obtained from the Swiss-Croatian Cooperation Programme. Lightning data were obtained from the Lightning Detection Network in Europe (LINET) (https://www.nowcast.de/en/solutions/linet-data; 5 June 2022). Authors acknowledge MeteoSwiss for providing radar observations (POH), Croatian Meteorological and Hydrological Service for providing hailpad observations and NASA for providing IMERG data. Hersbach et al., (2020) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. This research was supported by the WRF model (freely available at www.wrf-model.org/index.php). The WRF simulations were run using Isabella cluster (https://www.srce.unizg.hr/en/). RC, PV, NB, MED and CS acknowledge the Partnership for advanced computing in Europe (PRACE) for awarding access to Piz Daint at ETH Zürich at the Swiss National Supercomputing Centre (CSCS, Switzerland). The authors also acknowledge the Federal Office for Meteorology and Climatology (MeteoSwiss), CSCS, the Center for Climate Systems Modeling (C2SM) and ETH Zürich for their contributions to the development and maintenance of the GPU-accelerated version of COSMO.

Open Research

Data for this research were obtained from three sources. Lightning data were obtained from the Lightning Detection Network in Europe (LINET)*https://www.nowcast.de/en/solutions/linet-data*(Betz et al., 2009). ERA 5 reanalysis fields used as initial and boundary conditions can be obtained through the following link (*https://cds.climate.copernicus.eu/#!/home*) while the IMERG dataset can be obtained through (*https://gpm.nasa.gov/data/imerg*). Hail measurements from the hailpad network in Croatia are available through inquiries of the Croatian Meteorological and Hydrological Service (*usluge@cirus.dhz.hr*) while radar products from Switzerland are available through inquiries at MeteoSwiss (*Contact form*).

Author contributions

BM, RC and PV did run the WRF, COSMO-crCLIM and COSMO 6.0 simulations, respectively. BM drafted the paper, with all co-authors providing input.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences of the United States of America* ,*113* (42), 11770–11775. https://doi.org/10.1073/PNAS.1607171113/-/DCSUPPLEMENTAL

Adams-Selin, R. D., Clark, A. J., Melick, C. J., Dembek, S. R., Jirak, I. L., & Ziegler, C. L. (2019). Evolution of WRF-HAILCAST during the 2014-16 NOAA/Hazardous Weather Testbed Spring Forecasting Experiments. *Weather and Forecasting* , *34* (1), 61–79. https://doi.org/10.1175/WAF-D-18-0024.1

Adams-Selin, R. D., & Ziegler, C. L. (2016). Forecasting hail using a one-dimensional hail growth model within WRF. *Monthly Weather Review* , *144* (12), 4919–4939. https://doi.org/10.1175/MWR-D-16-0027.1

Allen, J. T., Giammanco, I. M., Kumjian, M. R., Jurgen Punge, H., Zhang, Q., Groenemeijer, P., Kunz, M., & Ortega, K. (2020). Understanding Hail in the Earth System. In *Reviews of Geophysics* (Vol. 58, Issue 1). Blackwell Publishing Ltd. https://doi.org/10.1029/2019RG000665

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

Ban, N., Schmidli, J., & Schär, C. (2014). Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research* , *119* (13), 7889–7907. https://doi.org/10.1002/2014JD021478

Belušić, A., Prtenjak, M. T., Güttler, I., Ban, N., Leutwyler, D., & Schär, C. (2018). Near-surface wind variability over the broader Adriatic region: insights from an ensemble of regional climate models. *Climate Dynamics* , *50* (11–12), 4455–4480. https://doi.org/10.1007/S00382-017-3885-5

Berthet, C., Dessens, J., & Sanchez, J. L. (2011). Regional and yearly variations of hail frequency and intensity in France. *Atmospheric Research* , *100* (4), 391–400. https://doi.org/10.1016/J.ATMOSRES.2010.10.008

Betz, H. D., Schmidt, K., Laroche, P., Blanchet, P., Oettinger, W. P., Defer, E., Dziewit, Z., & Konarski, J. (2009). LINET-An international lightning detection network in Europe. *Atmospheric Research* ,*91* (2–4), 564–573. https://doi.org/10.1016/j.atmosres.2008.06.012

Brimelow, J. C., Reuter, G. W., & Poolman, E. R. (2002). Modeling maximum hail size in Alberta thunderstorms. *Weather and Forecasting* , *17* (5), 1048–1062. https://doi.org/10.1175/1520-0434(2002)017<1048:MMHSIA>2.0.CO;2

Brisson, E., Blahak, U., Lucas-Picher, P., Purr, C., & Ahrens, B. (2021). Contrasting lightning projection using the lightning potential index adapted in a convection-permitting regional climate model. *Climate Dynamics 2021* , *1* , 1–15. https://doi.org/10.1007/S00382-021-05791-Z

Brisson, E., Brendel, C., Herzog, S., Ahrens, B., Brisson, E., Brendel, C., Herzog, S., & Ahrens, B. (2018). Lagrangian evaluation of convective shower characteristics in a convection-permitting model;Lagrangian evaluation of convective shower characteristics in a convection-permitting model. *MetZe* , *27* (1), 59–66. https://doi.org/10.1127/METZ/2017/0817

Brisson, E., Van Weverberg, K., Demuzere, M., Devis, A., Saeed, S., Stengel, M., & van Lipzig, N. P. M. (2016). How well can a convection-permitting climate model reproduce decadal statistics of precipitation, temperature and cloud characteristics? *Climate Dynamics* , *47* (9–10), 3043–3061. https://doi.org/10.1007/S00382-016-3012-Z

Brown, T. M., Pogorzelski, W. H., & Giammanco, I. M. (2015). Evaluating hail damage using property insurance claims data. *Weather, Climate, and Society* , *7* (3), 197–210. https://doi.org/10.1175/WCAS-D-15-0011.1

Changnon, S. A. (2009). Increasing major hail losses in the U.S.*Climatic Change* , *96* (1), 161–166. https://doi.org/10.1007/s10584-009-9597-z

Cui, R., Ban, N., Demory, M. E., Schar, C. (2023). Exploring hail and lightning diagnostics over the Alpine-Adriatic region in a km-scale climate model. *Weather and Climate Dynamics* (in preparation)

Curran, E. B., Holle, R. L., & Lopez, R. E. (2000). Lightning casualties and damages in the United States from 1959 to 1994.*Journal of Climate* , *13* (19), 3448–3464. https://doi.org/10.1175/1520-0442(2000)013<3448:LCADIT>2.0.CO;2

Dessens, J. (1998). A physical evaluation of a hail suppression project with silver iodide ground burners in southwestern France. *Journal of Applied Meteorology* , *37* (12), 1588–1599. https://doi.org/10.1175/1520-0450(1998)037<1588:APEOAH>2.0.CO;2

Dowdy, A. J., Fromm, M. D., & McCarthy, N. (2017). Pyrocumulonimbus lightning and fire ignition on Black Saturday in southeast Australia.*Journal of Geophysical Research: Atmospheres* , *122* (14), 7342–7354. https://doi.org/10.1002/2017JD026577

Dudhia, J. (1989). Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model.*Journal of the Atmospheric Sciences* , *46* (20), 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2

Ebert, E. E. (2008). Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteorological Applications* , *15* (1), 51–64. https://doi.org/10.1002/met.25

Federer, B., Schmid, W., & Waldvogel, A. (1978). The Design of Grossversuch IV, a Randomized Hail Suppression Experiment in Switzerland. *Atmosphere-Ocean* , *16* (1), 6–16. https://doi.org/10.1080/07055900.1978.9649009

Ferro, C. A. T., & Stephenson, D. B. (2011). Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events. *Weather and Forecasting* , *26* (5), 699–713. https://doi.org/10.1175/WAF-D-10-05030.1

Fiori, E., Comellas, A., Molini, L., Rebora, N., Siccardi, F., Gochis, D. J., Tanelli, S., & Parodi, A. (2014). Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: The Genoa 2011 case. *Atmospheric Research* , *138* , 13–29. https://doi.org/10.1016/j.atmosres.2013.10.007

Foote, G. A., Krauss, T. W., & Makitov, V. (2005). Hail metrics using convectional Radar. In *Proceedings of the 16th Conference on Planned and Inadvertent Weather Modification* , San Diego, CA, USA

Fowler, H. J., Ali, H., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Cabi, N. S., Chan, S., Dale, M., Dunn, R. J. H., Ekström, M., Evans, J. P., Fosser, G., Golding, B., Guerreiro, S. B., Hegerl, G. C., Kahraman, A., Kendon, E. J., ... Whitford, A. (2021). Towards advancing scientific knowledge of

climate change impacts on short-duration rainfall extremes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* ,*379* (2195). https://doi.org/10.1098/rsta.2019.0542

Franc, B., Filipović-Grčić, B., & Milardić, V. (2016). Lightning overvoltage performance of 110 kV air-insulated substation.*Electric Power Systems Research* , *138* , 78–84. https://doi.org/10.1016/J.EPSR.2015.12.002

Germann, U., Boscacci, M., Gabella, M., & Sartori, M. (2015).*Radar design for prediction in the Swiss Alps* .

Giaiotti, D., Nordio, S., & Stel, F. (2003). The climatology of hail in the plain of Friuli Venezia Giulia. *Atmospheric Research* ,*67 –68* , 247–259. https://doi.org/10.1016/S0169-8095(03)00084-X

Hentgen, L., Ban, N., Kröner, N., Leutwyler, D., & Schär, C. (2019). Clouds in Convection-Resolving Climate Simulations Over Europe.*Journal of Geophysical Research: Atmospheres* , *124* (7), 3849–3870. https://doi.org/10.1029/2018JD030150

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., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* , *146* (730), 1999–2049. https://doi.org/10.1002/QJ.3803

Holle, R. L., López, R. E., & Navarro, B. C. (2005). Deaths, Injuries, and Damages from Lightning in the United States in the 1890s in Comparison with the 1990s. *Journal of Applied Meteorology and Climatology* , *44* (10), 1563–1573. https://doi.org/10.1175/JAM2287.1

Hong, S., & Lim, J. (2006). The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). *Undefined* .

Horvath, K., Koracin, D., Vellore, R., Jiang, J., & Belu, R. (2012). Sub-kilometer dynamical downscaling of near-surface winds in complex terrain using WRF and MM5 mesoscale models. *Journal of Geophysical Research Atmospheres* , *117* (11). https://doi.org/10.1029/2012JD017432

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2019).*NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Prepared for: Global Precipitation Measurement (GPM) National Aeronautics and Space Administration (NASA)* . https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg

Jelić, D., Megyeri, O. A., Malečić, B., Belušić Vozila, A., Strelec Mahović, N., & Telišman Prtenjak, M. (2020). Hail Climatology Along the Northeastern Adriatic. *Journal of Geophysical Research: Atmospheres* , *125* (23). https://doi.org/10.1029/2020JD032749

Jelić, D., Prtenjak, M. T., Malečić, B., Vozila, A. B., Megyeri, O. A., & Renko, T. (2021). A New Approach for the Analysis of Deep Convective Events: Thunderstorm Intensity Index. *Atmosphere 2021, Vol. 12, Page 908* , *12* (7), 908. https://doi.org/10.3390/ATMOS12070908

Jewell, R., & Brimelow, J. (2009). Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States.*Weather and Forecasting* , *24* (6), 1592–1609. https://doi.org/10.1175/2009WAF2222230.1

Joe, P., Burgess, D., Potts, R., Keenan, T., Stumpf, G., & Treloar, A. (2004). The S2K severe weather detection algorithms and their performance. *Weather and Forecasting* , *19* (1), 43–63. https://doi.org/10.1175/1520-0434(2004)019<0043:TSSWDA>2.0.CO;2

Kain, J. S., & Kain, J. (2004). The Kain - Fritsch convective parameterization: An update. *Journal of Applied Meteorology* ,*43* (1), 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2

Kain, J. S., Weiss, S. J., Bright, D. R., Baldwin, M. E., Levit, J. J., Carbin, G. W., Schwartz, C. S., Weisman, M. L., Droegemeier, K. K., Weber, D., & Thomas, K. W. (2008). Some practical considerations

regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Weather and Forecasting* , 100804092600065. https://doi.org/10.1175/2008waf2007106.1

Kain, J. S., Weiss, S. J., LevIt, J. J., Baldwin, M. E., & Bright, D. R. (2006). Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Weather and Forecasting* , *21* (2), 167–181. https://doi.org/10.1175/WAF906.1

Kanata, J., Ametani, A., & Yamamoto, K. (2012). Threats of lightning current through an electric vehicle. *2012 31st International Conference on Lightning Protection, ICLP 2012* . https://doi.org/10.1109/ICLP.2012.6344299

Keller, M., Fuhrer, O., Schmidli, J., Stengel, M., Stöckli, R., & Schär, C. (2016). Evaluation of convection-resolving models using satellite data: The diurnal cycle of summer convection over the Alps.*Meteorologische Zeitschrift* , *25* (2), 165–179. https://doi.org/10.1127/METZ/2015/0715

Klasa, C., Arpagaus, M., Walser, A., & 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. https://doi.org/10.1002/QJ.3245

Kunz, M., Blahak, U., Handwerker, J., Schmidberger, M., Punge, H. J., Mohr, S., Fluck, E., & Bedka, K. M. (2018). The severe hailstorm in southwest Germany on 28 July 2013: characteristics, impacts and meteorological conditions. *Quarterly Journal of the Royal Meteorological Society* , *144* (710), 231–250. https://doi.org/10.1002/qj.3197

Lagasio, M., Parodi, A., Procopio, R., Rachidi, F., & Fiori, E. (2017a). Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: The Genoa 2014 event. *Journal of Geophysical Research* ,*122* (8), 4238–4257. https://doi.org/10.1002/2016JD026115

Lagasio, M., Parodi, A., Procopio, R., Rachidi, F., & Fiori, E. (2017b). Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: The Genoa 2014 event. In *Journal of Geophysical Research*(Vol. 122, Issue 8, pp. 4238–4257). https://doi.org/10.1002/2016JD026115

Latham, D., & Williams, E. (2001). Lightning and Forest Fires.*Forest Fires* , 375–418. https://doi.org/10.1016/B978-012386660-8/50013-1

Lavin-Gullon, A., Fernandez, J., Bastin, S., Cardoso, R. M., Fita, L., Giannaros, T. M., Goergen, K., Gutierrez, J. M., Kartsios, S., Katragkou, E., Lorenz, T., Milovac, J., Soares, P. M. M., Sobolowski, S., & Warrach-Sagi, K. (2021). Internal variability versus multi-physics uncertainty in a regional climate model.*International Journal of Climatology* , *41* (S1), E656–E671. https://doi.org/10.1002/joc.6717

Lee, J. Y., & Collins, G. J. (2017). Risk analysis of lightning effects in aircraft systems. *IEEE Aerospace Conference Proceedings* . https://doi.org/10.1109/AERO.2017.7943671

Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O., & Schär, C. (2017). Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres* ,*122* (10), 5237–5258. https://doi.org/10.1002/2016JD026013

Lynn, B. H., & Yair, Y. Y. (2008). Lightning Power Index: A new tool for predicting the lightning density and the potential for extreme rainfall. In *Geophysical Research Abstracts* (Vol. 10).

Malečić, B., Prtenjak, M. T., Horvath, K., Jelić, D., Jurković, P. M., Ćorko, K., & Mahović, N. S. (2022). Performance of HAILCAST and the Lightning Potential Index in simulating hailstorms in Croatia in a mesoscale model – Sensitivity to the PBL and microphysics parameterization schemes. *Atmospheric Research* ,*272* (March), 106143. https://doi.org/10.1016/j.atmosres.2022.106143

Manzato, A. (2008). A verification of numerical model forecasts for sounding-derived indices above Udine, northeast Italy. *Weather and Forecasting* , *23* (3), 477–495. https://doi.org/10.1175/2007WAF2007018.1

Manzato, A. (2012). Hail in Northeast Italy: Climatology and Bivariate Analysis with the Sounding-Derived Indices. *Journal of Applied Meteorology and Climatology* , *51* (3), 449–467. https://doi.org/10.1175/JAMC-D-10-05012.1

Manzato, A., Riva, V., Tiesi, A., & Marcello Miglietta, M. (2020). Observational analysis and simulations of a severe hailstorm in northeastern Italy. *Quarterly Journal of the Royal Meteorological Society* , *146* (732), 3587–3611. https://doi.org/10.1002/qj.3886

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research Atmospheres* , *102* (14), 16663–16682. https://doi.org/10.1029/97jd00237

Morel, S. (2014). *Verification of radar-based hail detection algorithms with insurance loss data in Switzerland* . 83.

Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology* , *119* (2), 397–407. https://doi.org/10.1007/s10546-005-9030-8

Nisi, L., Martius, O., Hering, A., Kunz, M., & Germann, U. (2016). Spatial and temporal distribution of hailstorms in the Alpine region: A long-term, high resolution, radar-based analysis. *Quarterly Journal of the Royal Meteorological Society* , *142* (697), 1590–1604. https://doi.org/10.1002/qj.2771

Park, S. H., Skamarock, W. C., Klemp, J. B., Fowler, L. D., & Duda, M. G. (2013). Evaluation of global atmospheric solvers using extensions of the jablonowski and williamson baroclinic wave test case. *Monthly Weather Review* , *141* (9), 3116–3129. https://doi.org/10.1175/MWR-D-12-00096.1

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., ... 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* (11–12), 3581–3602. https://doi.org/10.1007/s00382-021-05657-4

Pocakal, D. (2011). Hailpad data analysis for the continental part of Croatia. *Meteorologische Zeitschrift* , *20* (4), 441–447. https://doi.org/10.1127/0941-2948/2011/0263

Počakal, D., Večenaj, Ž., Mikuš Jurković, P., & Grisogono, B. (2018). Analysis of orographic influence on hail parameters in NW Croatia.*International Journal of Climatology* , *38* (15), 5646–5658. https://doi.org/10.1002/joc.5769

Počakal, D., Večenaj, Ž., & Štalec, J. (2009). Hail characteristics of different regions in continental part of Croatia based on influence of orography. *Atmospheric Research* , *93* (1–3), 516–525. https://doi.org/10.1016/j.atmosres.2008.10.017

Poolman, E. R., (1992). Die voorspelling van haelkorrelgroei in Suid-Afrika (The forecasting of hail growth in South Africa). M.S. thesis, Faculty of Engineering, University of Pretoria, 113 pp.

Prein, A. F., Holland, G. J., Rasmussen, R. M., Done, J., Ikeda, K., Clark, M. P., & Liu, C. H. (2013). Importance of Regional Climate Model Grid Spacing for the Simulation of Heavy Precipitation in the Colorado Headwaters. *Journal of Climate* , *26* (13), 4848–4857. https://doi.org/10.1175/JCLI-D-12-00727.1

Púcik, T., Castellano, C., Groenemeijer, P., Kühne, T., Rädler, A. T., Antonescu, B., & Faust, E. (2019). Large hail incidence and its economic and societal impacts across Europe. *Monthly Weather Review* , *147* (11), 3901–3916. https://doi.org/10.1175/MWR-D-19-0204.1

Punge, H. J., & Kunz, M. (2016). Hail observations and hailstorm characteristics in Europe: A review. *Atmospheric Research ,176 –177* , 159–184. https://doi.org/10.1016/j.atmosres.2016.02.012

Rachidi, F., Rubinstein, M., Montanyà, J., Bermúdez, J. L., Sola, R. R., Solà, G., & Korovkin, N. (2008). A review of current issues in lightning protection of new-generation wind-turbine blades. *IEEE Transactions on Industrial Electronics* , *55* (6), 2489–2496. https://doi.org/10.1109/TIE.2007.896443

Raupach, T. H., Martynov, A., Nisi, L., Hering, A., Barton, Y., & Martius, O. (2021). Object-based analysis of simulated thunderstorms in Switzerland: application and validation of automated thunderstorm tracking on simulation data. *Geoscientific Model Development Discussions* , *May* , 1–29. https://doi.org/10.5194/gmd-2021-105

Reinhardt, T., & Seifert, A. (2006). A Three-Category Ice Scheme For LMK. *COSMONewsletter* , *6* , 115–120.

Ritter, B., & Geleyn, J. F. (1992). A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations. Monthly Weather Review, 120, 303-325. - References - Scientific Research Publishing. *Monthly Weather Review* ,*120* , 303–325.

Romps, D. M., Seeley, J. T., Vollaro, D., & Molinari, J. (2014). Projected increase in lightning strikes in the united states due to global warming. *Science* , *346* (6211), 851–854. https://doi.org/10.1126/science.1259100

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., & Wernli, H. (2020). Kilometer-Scale Climate Models: Prospects and Challenges. *Bulletin of the American Meteorological Society* , *101* (5), E567–E587. https://doi.org/10.1175/BAMS-D-18-0167.1

Schuster, S. S., Blong, R. J., Leigh, R. J., & McAneney, K. J. (2005). Characteristics of the 14 April 1999 Sydney hailstorm based on ground observations, weather radar, insurance data and emergency calls. *Natural Hazards and Earth System Science* , *5* (5), 613–620. https://doi.org/10.5194/nhess-5-613-2005

Sioutas, M., Meaden, T., & Webb, J. D. C. (2009). Hail frequency, distribution and intensity in Northern Greece. *Atmospheric Research* , *93* (1–3), 526–533. https://doi.org/10.1016/J.ATMOSRES.2008.09.023

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., & Huang, X. (2019). A Description of the Advanced Research WRF Model Version 4. *Undefined* . https://doi.org/10.5065/1DFH-6P97

Skamarock, W. C. (2004). Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Monthly Weather Review* , *132* (12), 3019–3032. https://doi.org/10.1175/MWR2830.1

Smith, P. L., & Waldvogel, A. (1989). On Determinations of Maximum Hailstone Sizes from Hailpad Observations. *Journal of Applied Meteorology* , *28* (1), 71–76. https://doi.org/10.1175/1520-0450(1989)028<0071:odomhs>2.0.co;2

Sokol, Z., & Minářová, J. (2020). Impact of 1- and 2-moment cloud microphysics and horizontal resolution on lightning Potential Index within COSMO NWP model. *Atmospheric Research* , *237* . https://doi.org/10.1016/j.atmosres.2020.104862

Svabik, O. (1989). Review of meteorological aspects on hail defense activities in Austria. *Theoretical and Applied Climatology 1989 40:4* , *40* (4), 247–254. https://doi.org/10.1007/BF00865975

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research Atmospheres* ,*106* (D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Thornton, J. A., Virts, K. S., Holzworth, R. H., & Mitchell, T. P. (2017). Lightning enhancement over major oceanic shipping lanes. *Geophysical Research Letters* , *44* (17), 9102–9111. https://doi.org/10.1002/2017GL074982

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review* ,*117* (8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

Tiesi, A., LastNameMazzà, S., Conte, D., Ricchi, A., Baldini, L., Montopoli, M., Picciotti, E., Vulpiani, G., Ferretti, R., & Miglietta, M. M. (2022). Numerical Simulation of a Giant-Hail-Bearing Mediterranean Supercell in the Adriatic Sea. *Atmosphere 2022, Vol. 13, Page 1219* , *13* (8), 1219. https://doi.org/10.3390/ATMOS13081219

Trefalt, S., Martynov, A., Barras, H., Besic, N., Hering, A. M., Lenggenhager, S., Noti, P., Röthlisberger, M., Schemm, S., Germann, U., & Martius, O. (2018). A severe hail storm in complex topography in Switzerland - Observations and processes. *Atmospheric Research* ,*209* , 76–94. https://doi.org/10.1016/j.atmosres.2018.03.007

Vergara-Temprado, J., Ban, N., Panosetti, D., Schlemmer, L., & Schär, C. (2020). Climate Models Permit Convection at Much Coarser Resolutions Than Previously Considered. *Undefined* , *33* (5), 1915–1933. https://doi.org/10.1175/JCLI-D-19-0286.1

Waldvogel, A., Federer, B., & Grimm, P. (1979). Criteria for detection of hail cells. *Journal of Applied Meteorology* , *18* (12), 1521–1525. https://doi.org/10.1175/1520-0450(1979)018<1521:CFTDOH>2.0.CO;2

Willemse, S., & Furger, M. (2016). *From weather observations to atmospheric and climate sciences in Switzerland. Celebrating 100 years of the Swiss Society of Meteorology* .

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D., & Thomas, K. W. (1998). An enhanced hail detection algorithm for the WSR-88D. *Weather and Forecasting* , *13* (2), 286–303. https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2

Yair, Y., Lynn, B., Price, C., Kotroni, V., Lagouvardos, K., Morin, E., Mugnai, A., & Del Carmen Llasat, M. (2010). Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields.*Journal of Geophysical Research Atmospheres* , *115* (4). https://doi.org/10.1029/2008JD010868

Zeman, C., Wedi, N. P., Dueben, P. D., Ban, N., & Schär, C. (2021). Model intercomparison of COSMO 5.0 and IFS 45r1 at kilometer-scale grid spacing. *Geosci. Model Dev* , *14* , 1–23. https://doi.org/10.5194/gmd-14-1-2021

Hosted file

`supplementarymaterial_jgr_bmalecic_2023_cosmoxwrf_07_10_2022.docx` available at https://authorea.com/users/555430/articles/606029-simulating-hail-and-lightning-over-the-alpine-adriatic-region-a-model-intercomparison-study

B. Malečić[1], R. Cui[2], M. E. Demory[2], K. Horvath[3], D. Jelić[1], C. Schär[2], M. Telišman Prtenjak[1], P. Velasquez[2], N. Ban[4]

[1] University of Zagreb, Faculty of Science, Department of Geophysics, Zagreb, Croatia.

[2] Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland.

[3] Croatian Meteorological and Hydrological Service, Zagreb, Croatia.

[4] University of Innsbruck, Department of Atmospheric and Cryospheric Sciences, Innsbruck, Austria.

Corresponding author: Barbara Malečić (barbara.malecic@gfz.hr)

Key Points:

- Km-scale simulations over the Alpine-Adriatic region show that HAILCAST and LPI are successful in diagnosing observed hail and lightning.
- Results are presented using the COSMO and WRF models for a total of eight case studies covering different types of hail storms.
- Overall the two models yield similar results, but some systematic differences are found and tied to differences in model structure.

Abstract

Hail is a significant convective weather hazard, often causing considerable crop and property damage across the world. Although extremely damaging, hail still remains a challenging phenomenon to model and forecast, given the limited computational resolution and the gaps in understanding the processes involved in hail formation. Here, eight hailstorms occurring over the Alpine-Adriatic region are analyzed using Weather Research and Forecasting (WRF) and Consortium for Small Scale Modeling (COSMO) simulations, with embedded HAILCAST and Lightning Potential Index (LPI) diagnostics at kilometer-scale grid spacing (~2.2 km). In addition, a model intercomparison study is performed to investigate the ability of the different modeling systems in reproducing such convective extremes, and to further assess the uncertainties associated with simulations of such localized phenomena. The results are verified by hailpad observations over Croatia, radar estimates of hail over Switzerland and lightning measurements from the LINET network. The analysis revealed that both HAILCAST and LPI are able to reproduce the areas and intensities affected by hail and lightning. Moreover, the hail and lightning fields produced by both models are similar, although a slight tendency of WRF to produce smaller hail swaths with larger hailstones and higher LPI compared to COSMO is visible. It is found that these differences can be explained by systematic differences in vertical profiles of microphysical properties and updraft strength between the models. Overall, the promising results indicate that both HAILCAST and LPI could be valuable

tools for real-time forecasting and climatological assessment of hail and lightning in current and changing climates.

1 Introduction

Hail is a severe weather hazard that can produce significant crop and property damage across the world (Allen et al., 2020), especially when it occurs over highly populated areas with high-density assets (Kunz et al., 2018). In the literature, a large number of hailstorms causing more than US$1 billion in damage is reported across the world (Schuster et al., 2005; Changnon, 2009; Brown et al., 2015; Kunz et al., 2018; Púcik et al., 2019). As described by Punge and Kunz (2016) and Púcik et al. (2019), several hail hotspots can be found in Europe, including the pre-Alpine and Adriatic areas. Although large hail occurs less often over the highest mountain peaks in the central Alps, severe hailstorms frequently affect Switzerland with up to 4 large hail days per year (Nisi et al., 2016; Púcik et al., 2019). In this area, the maximum hail diameter can sometimes exceed 10 cm (e.g., see Figure 8 from Púčik et al., 2019). Furthermore, parts of Croatia (Počakal et al., 2018; Jelić et al., 2020) and broader northern Adriatic region (e.g. Manzato (2012)) have similar statistics of hail frequency as southern Germany or southeastern Austria (Punge & Kunz, 2016). Therefore, considering the high economic losses associated with (severe) hailstorms, and high frequencies of hail occurrence, it is very important to have successful hail forecasts, both for short-term numerical weather prediction (NWP) and long-term climate-change adaptation strategies.

One of the largest limitations in understanding processes involved in hail formation is the lack of dense and direct measurements of hail properties on the ground. Hailpads, which are simple meteorological devices consisting of a stand and a measuring plate, represent one of the few methods to detect and measure hailstones directly on the ground. Besides the number of falling hailstones and their diameters, hailpads can also detect the intensity (i.e., kinetic energy) of hail (Smith & Waldvogel, 1989). In Europe, hailpad networks exist in several regions including parts of Spain, France, Greece, northern Italy, eastern Austria and parts of Croatia (Svabik, 1989; Dessens, 1998; Giaiotti et al., 2003; Sioutas et al., 2009; Počakal et al., 2009; Berthet et al., 2011; Počakal, 2011) and have also been used in randomized hail suppression experiments in Switzerland (Federer et al., 1978). Although hailpads are one of the few sources of direct information on hail occurrence, they provide spatially discrete (but unique) information on hail occurrence as they only record hail at the point where they are installed.

Another source of information on hail occurrence is related to weather radars. As the abilities of weather radars to detect different kinds of hydrometeors such as rain, snow and hail progressed over the years, several hail detection algorithms have been developed (e.g. Waldvogel et al., 1979; Witt et al., 1998). At the present, hail detection algorithms are widely used as hail proxies and can provide spatially continuous information on various hail properties, e.g., probability of hail occurrence or maximum expected hailstone size. In Switzerland, two hail

detection algorithms are operational in real-time, namely, Probability of Hail (POH, Waldvogel et al., 1979; Foote et al., 2005) that indicates a probability of a hailstorm occurring at a certain location, and Maximum Expected Severe Hail Size (MESHS, Treloar, 1998; Joe et al., 2004) that estimates expected severe hail size at the ground over the Alpine region.

An additional challenge in understanding hail processes is the limited number of high-resolution modeling studies of hailstorms. With increasing computational power, it has become possible to run simulations at convection-permitting scales (horizontal grid spacing < 4 km). Several studies reported the benefits of using models at kilometer scales for more realistic representations of convective processes (Leutwyler et al., 2017), mean diurnal cycles of precipitation (Ban et al., 2014), spatial precipitation patterns and associated extreme values (Prein et al., 2013; Brisson et al., 2016, 2018; Fowler et al., 2021; Pichelli et al., 2021), better representation of convective clouds (Keller et al., 2016; Brisson et al., 2016; Hentgen et al., 2019), local wind systems like sea breeze (Belušić et al., 2018), and complex terrain winds (Horvath et al., 2012). Since models, when run at km scales, can produce a more realistic representation of convective processes, Adams-Selin and Ziegler (2016) integrated a physically improved 1D hail growth scheme – called HAILCAST (Poolman, 1992; Brimelow et al., 2002; Jewell & Brimelow, 2009) – with the km-scale WRF model. When HAILCAST is coupled with WRF, the model simulates the maximum expected hail size at the ground using the profiles of cloud liquid and ice water, vertical velocity, temperature, water vapor and pressure fields from a given model timestep. Several recent studies employed HAILCAST embedded in high-resolution numerical models, such as WRF or COSMO, to study hailstorms occurring over the United States and Europe. The studies found that the models can reproduce the atmospheric conditions and triggering mechanisms responsible for hailstorm formation, resulting in simulating comparable hailstorms to those observed over the complex terrain of the United States (Adams-Selin & Ziegler, 2016; Adams-Selin et al., 2019), Switzerland (Trefalt et al., 2018; Raupach et al., 2021; Cui et al., 2023), Italy (Manzato et al., 2020; Tiesi et al., 2022), and Croatia (Malečić et al., 2022).

Similar to hail, lightning poses a serious threat to human lives (Curran et al., 2000; Holle et al., 2005), wind turbines (Rachidi et al., 2008) and transportation (Kanata et al., 2012; Lee & Collins, 2017; Thornton et al., 2017). Moreover, lighting is a major cause of wildfires (Latham & Williams, 2001; Abatzoglou & Williams, 2016; Dowdy et al., 2017). Considering the hazards associated with lightning occurrence, the lightning potential index (LPI) was developed as a tool for diagnosing areas prone to lightning discharges (Lynn & Yair, 2008; Yair et al., 2010). With a better representation of convective processes km-scale simulations, LPI offers the possibility to use the parameterizations of lightning that describe the non-inductive process occurring inside a thundercloud (Yair et al., 2010; Brisson et al., 2021). It is defined as a potential for charge formation and separation inside a thundercloud and it relies on the presence of both solid and liquid hydrometeors. Even though LPI is not directly connected to the

observed number of lightning flashes, several studies found that LPI could be a valuable tool for implicit lightning forecasting in COSMO (Sokol & Minářová, 2020, Cui et al., 2023) and WRF (Yair et al., 2010; Lagasio et al., 2017a; Malečić et al., 2022) models. Recently, LPI was used in the climatological assessment of lightning over Germany (Brisson et al., 2021) and proved to be a better indicator of lightning occurrence than the commonly used convective available potential energy times precipitation (CAPE x PREC) parameterization (Romps et al., 2014).

Recently, an effort was made to utilize both HAILCAST and LPI to study hailstorms occurring in Croatia using WRF (Malečić et al., 2022) and over the broader Alpine-Adriatic region using COSMO (Cui et al., 2023) models. More specifically, Cui et al., (2023) selected eight days with severe convection over the Alpine-Adriatic region based on the observed impacts and underlying synoptic forcing to study the mechanisms responsible for severe weather effects. Adopting the process-oriented approach, the authors found that both HAILCAST and LPI successfully reproduced observed hail and lightning characteristics over a broad range of synoptic situations.

Motivated by the promising results, this study aims to complement the valuable previous research on hailstorm simulation over the topographically complex Alpine-Adriatic region (Figure 1a). More specifically, we employ these two diagnostic tools in km-scale models (COSMO and WRF) to simulate hailstorms analyzed by Cui et al. (2023), with a geographical focus on Croatia and Switzerland. By analyzing eight hail cases using two models, we aim (i) to conduct a systematic and quantitative evaluation of the model's performance for all cases, and (ii) to identify the robustness of HAILCAST and LPI results produced by two intrinsically different modeling systems. The outcome of this study reveals information about model biases and the origins of disagreements between the two models in simulating severe storms associated with hail and lightning over the complex Alpine-Adriatic region. Moreover, as for Cui et al. (2023), this study benefits from two valuable but intrinsically different datasets of hail observations, namely spatially discrete, but direct, hail measurements from the hailpad network (and two hailpad polygons) in Croatia and spatially continuous, but indirect, radar estimates on hail occurrence from Switzerland. Therefore, the objectives of this paper can be summarized as follows:

- To what extent can HAILCAST and LPI credibly diagnose the occurrence of hail and lightning?
- How do simulations with two distinct models but the same diagnostic hail and lightning modules differ from each other?

This paper is organized as follows: Section 2 describes the selected hail events and observational data used to evaluate COSMO and WRF models. An overview of the model setups and evaluation approach is indicated in Section 3. The results are presented and discussed in Section 4. The conclusions and remarks are given in Section 5.

2 Selected hailstorms and observational data

Eight days with severe convection occurring over the Alpine-Adriatic region already analyzed in Cui et al. (2023) are selected for the analysis. Out of eight selected days, hailstorms were observed in Croatia and/or Switzerland during seven of these days. During the day without observed hail in the central and eastern Alps, intense precipitation over the Alps leading to severe flooding is reported. Hailstorms are selected based on their intensity, measured by their impact or the kinetic energy recorded with hailpads. Besides that, an attempt was made to select hailstorms occurring over a set of synoptic and mesoscale situations to assess the abilities of both models to reproduce the observed convection during a variety of driving conditions. A detailed description of synoptic factors contributing to the occurrence of severe weather during the selected days is given by Cui et al. (2023). Moreover, by selecting a day where no hail is observed, the ability of both models to distinguish between convective days with and without hail is assessed. The type of observations available for eight selected events along with a brief description of their impacts are listed in Table 1.

Table 1. Selected convective events over the Alpine-Adriatic region. The type of available observations and a short description of the event's impact is indicated. Radar observations are covering Switzerland while hailpad observations are associated with the hailpad network in Croatia. More information on the impacts of some of these events can be obtained through http://www.sturmarchiv.ch/index.php/Hagel.

Date	Observations of hail	Impact
23 July 2009	Radar	A thunderstorm embedded in a cold front hit S
1 June 2013	Radar	Significant damage to vineyards in eastern Swi
18 June 2013	No hail observed in Croatia and Switzerland	Intense precipitation over the Alps led to sever
25 June 2017	Radar, Hailpads	Large mesoscale convective system observed in
8 July 2017	Radar	Damage to crops and vineyards in Switzerland
24 July 2017	Radar, Hailpads	Large hailstones were observed in Croatia. Ma
17 May 2018	Hailpads	Non-gradient pressure field over Croatia. Signi
30 May 2018	Radar	Significant damage to cars and buildings in cer

To assess the model's ability to reproduce the observed severe weather events, several datasets are used. To validate simulated precipitation over the Alpine-Adriatic region, the Final Run of Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) mission (Huffman et al., 2019) dataset is used. IMERG is a globally gridded precipitation product that estimates surface precipitation rates at 0.1° spatial and 30 min temporal resolution. IMERG incorporates satellite microwave precipitation estimates, microwave-calibrated infrared satellite estimates and rain gauge observations. Combining remote sensing and in situ observations, IMERG provides spatially and temporally contin-

uous and homogeneous precipitation estimates over the whole Alpine-Adriatic region.

Further, to assess the ability of LPI to reproduce the observed lightning activity, lightning data from the Lightning Detection Network (LINET) (Betz et al., 2009) is used. With 190 sensors in 31 countries that are up to 250 km away, LINET successfully detects cloud-to-ground and intracloud lightning flashes and differentiates between positive and negative discharges across Europe. Moreover, the LINET network can detect weaker stroke signals with a current amplitude lower than 1 kA. With continuous improvements of the LINET network, the median values of detected current amplitude values had decreased by half from 2009 (Franc et al., 2016), showing significant improvement in the sensitivity detection toward smaller stroke current amplitudes. For most of the European region (Franc et al., 2016; Jelić et al., 2021), the average minimum detectable signal is 0.7 kA, and the median location accuracy error is ±84 m. Here, we considered total lightning information, i.e., we did not differentiate between types or polarities of lightning flashes as LPI presents the overall potential for lightning activity without preferences to the type or polarity of lightning discharges. The total lightning for the examined cases was taken from the 2D database of lightning flashes at a 3 km x 3 km horizontal and 2 min temporal resolution (developed by Jelić et al., 2021) over the domain shown in Figure 1a.

Next, hail detection products from the Swiss radar network (Germann et al., 2015; Willemse & Furger, 2016) operated by MeteoSwiss are used to assess the HAILCAST results. Namely, operationally computed POH product is used. POH indicates the grid-based probability of hail reaching the ground. It is computed following Waldvogel et al. (1979) and Foote (2005) as a difference in height between the altitude of the center of the highest radar bin at which 45 dBZ echo (i.e., Echo Top of 45 dBZ) is found and the height of the freezing level retrieved from the forecasts of the operational numerical prediction model COSMO. POH has been verified using insurance loss data (Morel, 2014; Nisi et al., 2016) and a good agreement between hail damage and POH >= 80 % was found. The area in which this product is available is indicated in Figure 1b.

Finally, HAILCAST results are assessed against direct hail measurements from the Croatian hailpad network. It consists of (i) hail suppression stations in the continental region of Croatia, (ii) a specially designed hailpad polygon in northwestern Croatia, and (iii) hailpad stations in the northeastern (NE) Adriatic region (Figure 1c). Overall, 590 hailpads on hail suppression stations and 150 hailpads on the polygon with average spacing between hailpads of ~5.5 km and ~2 km, respectively, have been installed and maintained by the Croatian Meteorological and Hydrological Service (Počakal et al., 2009; Pocakal, 2011). Moreover, during the VITCLIC project (https://www.pmf.unizg.hr/geof/en/research/climatology/vitclic) 65 hailpads were installed in Istria (NE Adriatic) in the vicinity of the vineyards. Notably, the Istrian region is not a part of the hail suppression network; therefore, hail observations from these hailpads are not under the potential influence of hail

suppression activities.

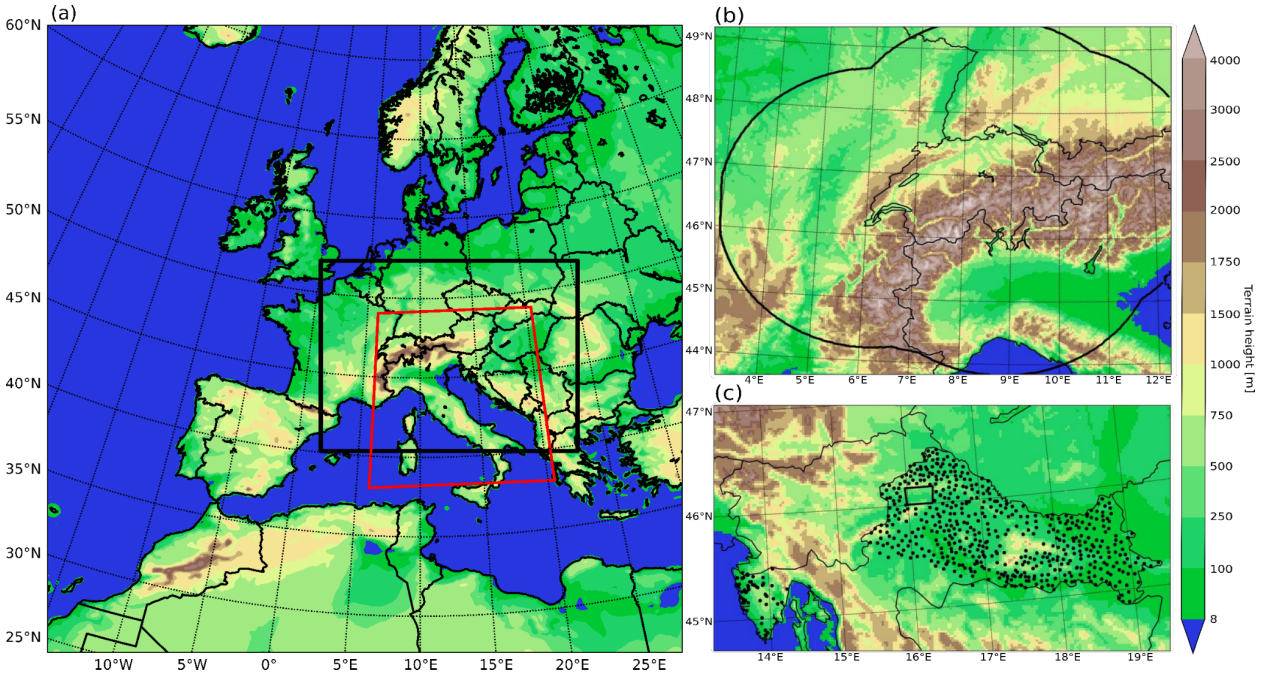

Figure 1. (a) Terrain height (above sea level) as represented in WRF for the outer 12 km domain. The inner 2.2 km domain is indicated with the black rectangle. The domain where LINET measurements are available is indicated with the red line. (b) Terrain height as represented in WRF for the 2.2 km domain over the Alpine region. The black line indicates the Swiss radar spatial coverage. (c) Terrain height as represented in WRF for the 2.2 km domain over Croatia. The black dots indicate the positions of hailpads.

3 Modeling setup and evaluation approach

3.1 COSMO and WRF setups

Selected hailstorms were simulated using an Advanced Research Weather Research and Forecasting (WRF, version 4.1.5) model (C. Skamarock et al., 2019), and the climate version of the Consortium for Small Scale Modelling (COSMO-crCLIM based on COSMO 5.0) model (Baldauf et al., 2011; Leutwyler et al., 2017; Schär et al., 2020) alongside HAILCAST and LPI. An attempt was made to make a setup of both models as similar as possible. Additionally, one of the hailstorms is simulated using the newest version of COSMO v6.0 model.

The modeling setup consisted of two one-way nested domains with horizontal grid spacing of approximately 12 km (0.11°, 361x361 grid points) and 2.2 km (0.02°, 800x600 grid points for COSMO and 801x601 grid points for WRF) (Figure 1a). Considering the importance of fine grid spacing in the vertical

direction (e.g. Fiori et al., 2014) and the sensitivity of sounding derived indices to vertical sampling (e.g. Manzato, 2008), the model setup consists of 65 vertical levels in WRF and 60 vertical levels in COSMO. WRF uses a hybrid sigma-pressure vertical coordinate (Park et al., 2013), while COSMO uses Gal-Chen coordinates. WRF's time step is set to 20 s and 4 s, while COSMO's time step is set to 90 s and 20 s for 12 km and 2.2 km simulations, respectively. The simulations were initialized and driven at the lateral boundaries using ERA5 reanalysis (Hersbach et al., 2020) at 12 UTC the day before severe convection was observed. It should be noted that unlike for WRF, soil moisture for COSMO was not taken directly from ERA5 reanalysis. Instead, each case simulation was initialized 7 days before the event using the equilibrated monthly mean soil profiles from a 10-year (1999-2008) 12 km COSMO climate simulation (Vergara-Temprado et al., 2020), and let run for 7-days. Then, the model integration started at 12 UTC the day before severe convection was observed using the new soil moisture conditions from the 7 days run. This approach allows for an adjustment of the top soil layers to the conditions of each event.

Grid spacing of approximately 2.2 km allows the model to represent convective processes explicitly (Kain et al., 2006, 2008), and therefore, in the inner domain, no cumulus parameterization is applied. In the outermost domain, convection in WRF is parameterized using the Kain-Fritsch scheme (Kain & Kain, 2004). Other physics options used include the rapid radiative transfer model scheme (RRTM) (Mlawer et al., 1997) for longwave radiation and the Dudhia scheme (Dudhia, 1989) for shortwave radiation. Regarding the planetary boundary layer (PBL) parameterization and microphysics parameterization schemes, MYNN2.5 (Nakanishi & Niino, 2006) and WRF single-moment six-class scheme (WSM6) (Hong & Lim, 2006) are used.

The convection scheme used for the outer COSMO 12 km domain is based on the Tiedtke (1989) with shallow convection parameterized version. Similar to WRF, for the inner domain, explicit convection is applied for shallow and deep convection. The single-moment microphysics scheme (Reinhardt & Seifert, 2006) represents the cloud microphysics with prognostic cloud water, cloud ice, graupel, rain and snow. The radiation scheme is provided by Ritter and Geleyn (1992) with a -two-stream approach.

For the simulation of hail and lightning, HAILCAST (Adams-Selin & Ziegler, 2016) and LPI (Lynn & Yair, 2008; Yair et al., 2010) are used in both models. HAILCAST is a time-dependent hail growth model that provides the forecast of the maximum hailstone diameter at the ground. In our setup, HAILCAST is activated every 5 min on the inner 2.2 km convection-permitting domain if the updraft in a particular grid point exceeds 10 ms^{-1} for more than 15 min. Similarly, we adopt the same formulation of LPI in both WRF and COSMO models. In those formulations the following conditions must be met so that the LPI for a particular grid point is nonzero: (i) vertical velocity in a particular grid point must be greater than 0.5 ms^{-1}, (ii) vertical velocity in adjacent grid points (within a five-grid radius) must be greater than 2 ms^{-1}, and finally (iii)

a particular grid point and its adjacent grid points must be in an unstable environment. An unstable environment is defined by analysis of a parameter similar to mixed-layer CAPE obtained by the integration over a 500 hPa layer starting at 50 hPa above ground. More details on these requirements can be found in Brisson et al. (2021). In this study, LPI is computed every 15 min and 15 min fields are stored for both models. More information on HAILCAST and LPI can be found in the Supplement.

3.2 Evaluation approach

When evaluating the results of diagnostic tools such as HAILCAST and LPI against observations, it should be considered that their performance relies on the skill of the convection-permitting model to represent the convection properly. For this reason, the results are evaluated in three sequential phases. First, the model's skill to represent the observed precipitation is evaluated. Second, simulated LPI is assessed against LINET lightning data using the minimum coverage neighborhood verification method (Ebert, 2008). Third, HAILCAST results are evaluated against radar estimates on hail occurrence from Switzerland and direct hail measurements from the Croatian hailpad network using a minimum coverage neighborhood verification method (Ebert, 2008) and upscaled neighborhood verification method proposed by Malečić et al. (2022), respectively.

The first step considered an evaluation of the skill with which models produced the observed convection. Specifically, precipitation simulated in both models is assessed against precipitation estimated by IMERG. Considering that simulated precipitation could be shifted in time compared to the observations, daily aggregated fields of both simulated and observed precipitation are evaluated. Evaluation is done by determining standard deviations, correlation coefficients and root mean square errors between observed and modeled fields. The results are summarized using Taylor diagrams (Taylor, 2001).

Second, LPI is assessed against lightning observations from the LINET network. Here, we had to consider that it is difficult for a high-resolution simulation to precisely match the observation in space and/or time. However, even simulations that do not precisely match observations in space, time, or even intensity, can still be useful (Ebert, 2008). For this reason, in our verification approach, we are using a minimum coverage neighborhood method. Thus, a useful forecast is defined as the one where lightning is simulated anywhere in the neighborhood of the point where it is observed. Based on this method, a contingency table is built and a symmetric extremal dependence index (SEDI) (Ferro & Stephenson, 2011) is computed. Moreover, here we are varying verification window sizes as well as thresholds for the number of lightning flashes to examine the scale-intensity combination at which high-resolution simulation is useful.

Third, HAILCAST results are assessed against radar products POH and MESHS from Switzerland and hail measurements from the Croatian hailpad network. To evaluate HAILCAST results against radar products, a minimum coverage

verification method with varying verification windows sizes is utilized and categorical skill score such as probability of detection (POD), false alarm ratio (FAR) and extremal dependence index (EDI) (Ferro & Stephenson, 2011) are determined. Next, HAILCAST results are assessed against hailpad observation from Croatian hailpad network. Here, to overcome challenges associated with the limited spatial information from hailpad networks and to limit the effect of double penalty that occurs when verifying slightly offset high-resolution forecasts of extremely rare events (Ebert, 2008), an upscaled neighborhood verification method is used (Malečić et al., 2022). This verification methodology is composed of the elements of methods such as point to point, upscaling and a minimum coverage verification method (as described by Ebert (2008)) and is further described by Malečić et al. (2022). Based on this method, a contingency table is built and categorical skill scores are determined.

4 Results and discussion

4.1 Precipitation

The first step in the analysis of results consists of the comparison between simulated and observed precipitation. The comparison refers to the period from 00 UTC to 24 UTC on the day when severe convection was observed. For the evaluation, the IMERG dataset is selected, since it covers the whole domain of interest (domain 2 from Figure 1). When comparing the daily accumulated fields of both observed and simulated precipitation, a generally good agreement between observed and simulated fields can be found (Figure 2). Therefore, both models reproduce the observed precipitation patterns fairly well, even though there are slight local variations. Moreover, the areas with more intense precipitation correspond well between simulated and observed fields. However, a tendency of both models to produce more peaked and more scattered precipitation objects compared to IMERG data can be found in all analyzed cases except 8 July 2017 and 18 June 2013. This tendency could partially be attributed to the horizontal resolution differences between simulated and observed fields (simulated fields are obtained at 2.2 km horizontal grid spacing, while observed fields are obtained at 11 km horizontal grid spacing). On the other hand, for 18 June 2013 and 8 June 2017, both models produced mostly smaller and less peaked objects than observed. In addition, comparing the fields produced by COSMO-crCLIM and WRF, it is noted that WRF tends to produce slightly less peaked precipitation objects than COSMO-crCLIM.

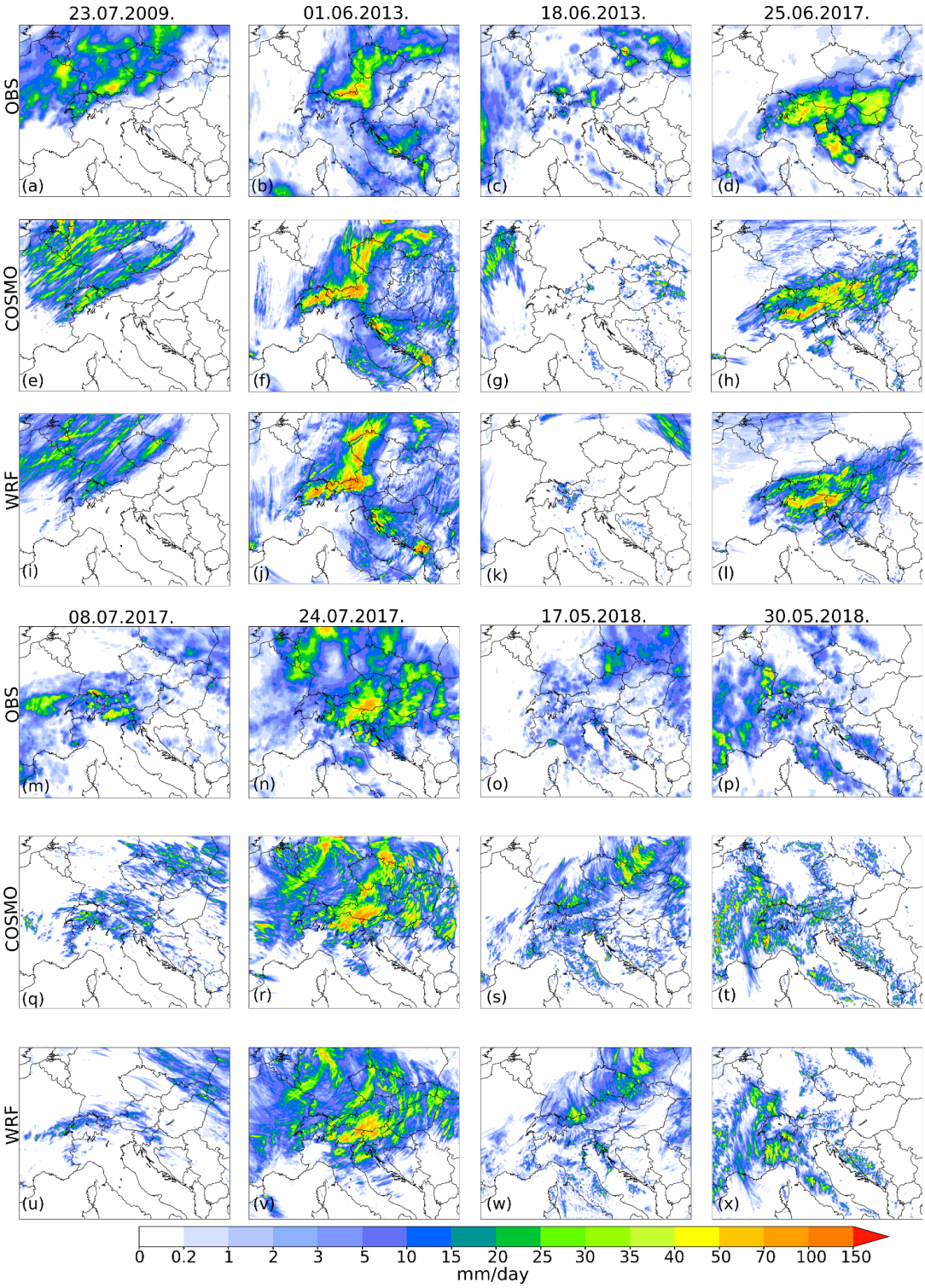
23.07.2009.
01.06.2013.
18.06.2013.
25.06.2017.
OBS
COSMO
WRF
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
08.07.2017.
24.07.2017.
17.05.2018.
30.05.2018.
(m)
(n)
(o)
(p)
(q)
(r)
(s)
(t)
(u)
(v)
(w)
(x)
0
0.2
1
2
3
5
10
15
20
25
30
35
40
50
70
100
150
mm/day

Figure 2. Accumulated precipitation for the period between 00 and 24 UTC on the day severe convection was observed for the eight case studies. Fields observed by (a-d; m-p) IMERG and simulated by (e-h; q-t) COSMO-crCLIM and (i-l; u-x) WRF are presented.

To further expand and complement this analysis, the hourly accumulated precipitation averaged over the whole inner domain (Figure 1a) is compared between the observations and the models (Figure 3). Considering the discrepancy between the model's and IMERG horizontal resolution, both observed and simulated fields are interpolated to the outer 12 km grid with bounds corresponding to the inner 2.2 km grid presented in Figure 1a. The comparison reveals that both models captured the temporal evolution of precipitation fairly well in all cases, except 25 June 2017. However, it should be noted that some underestimations/overestimations or time shifts might be present, depending on the case analyzed. For 25 June 2017, both models failed to represent the two local maximums of precipitation observed in the early morning and evening hours. Interestingly, during most cases, both models simulate comparable or slightly larger amounts of precipitation compared to the observations. An underestimation of precipitation is present only for 18 June 2013 and 8 July 2017 cases as already noted in the analysis of precipitation spatial patterns. Therefore, for the events analyzed, WRF produces, on average, less precipitation compared to COSMO-crCLIM.

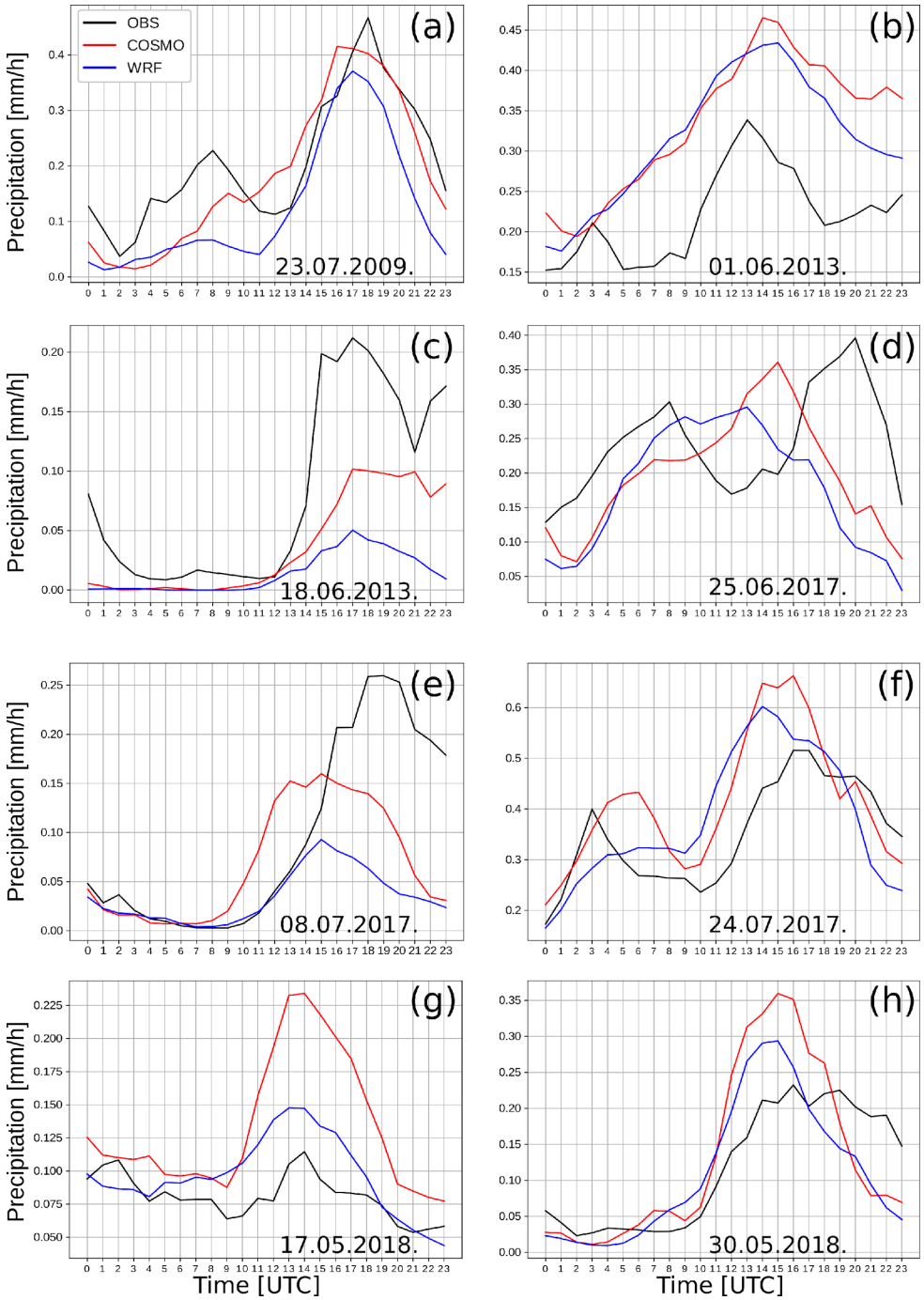

Figure 3. Hourly accumulated precipitation obtained from IMERG (black) observations and simulated by COSMO-crCLIM (red) and WRF (blue) for all

eight cases (a)-(h). The observed and simulated hourly precipitation amounts are interpolated to a 12 km grid and then averaged over the 2.2 km domain.

The precipitation results are further compared quantitatively using Taylor diagrams (Taylor, 2001b). To account for possible time shifts between observed and simulated convection, and different horizontal resolutions between IMERG and the models, Taylor diagrams are obtained by interpolating the daily accumulated observed and simulated fields to a common 12 km grid. As shown in Figure 4 both models perform similarly, although larger differences in standardized deviations are found for 18 June 2013 and 8 July 2017 case. Both models show similar correlation coefficients between simulated and observed fields. Looking at the median performance for all cases together, it can be concluded that both models perform similarly in terms of simulating the observed precipitation with standardized deviations of 1.14 and 1 mm/day, correlation coefficients of 0.48 and 0.46 and root mean square errors of 1.14 and 1.04 mm/day for COSMO-crCLIM and WRF, respectively.

Overall, given the presented analysis it can be concluded that both models successfully represent the observed precipitation, and WRF tends to simulate less precipitation than COSMO-crCLIM.

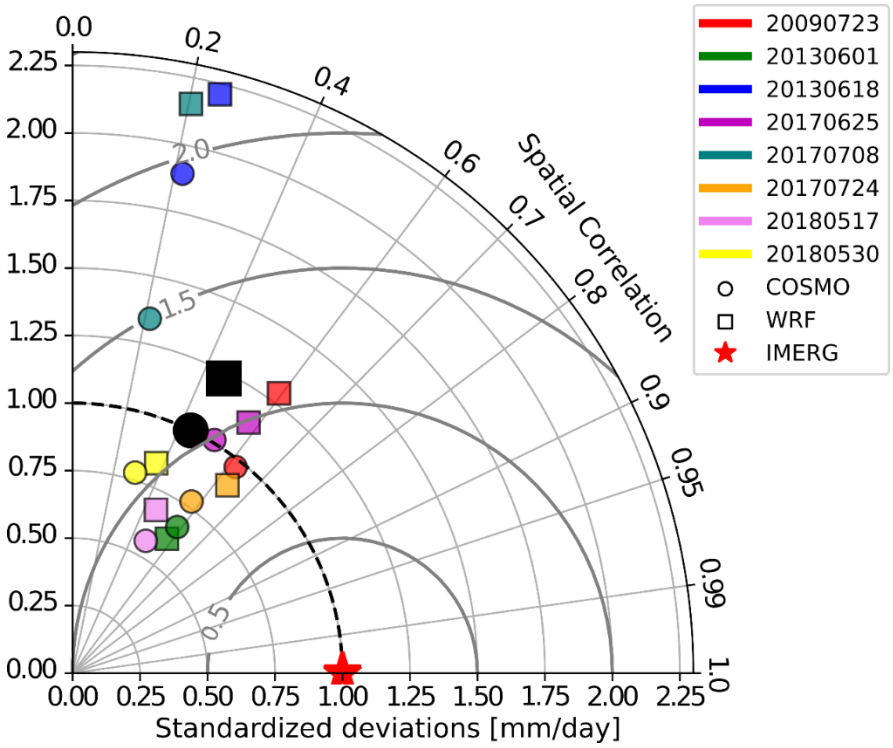

Figure 4. Taylor diagram showing the performance of COSMO-crCLIM (circle) and WRF (square) when simulating daily accumulated precipitation observed

by IMERG (red star). The performance for each case is indicated by colored markers, while the corresponding median values are indicated by black markers.

4.2 Lightning potential index results

The second step of the evaluation considers the assessment of LPI [J/kg] against the observed number of lightning flashes from the LINET network. LPI indicates the potential for lightning activity, and as such, it is not directly connected to the observed number of lightning flashes. To make a direct comparison between LPI and the observed number of lightning flashes, a conversion of LPI to the number of lightning flashes following Brisson et al. (2021) and Malečić et al. (2022) is done. Conversion assumes a linear relationship between the LPI and the observed number of lightning flashes as well as the existence of the threshold value of LPI for which a lightning flash is produced, such that:

$$\text{k LPI + l}, \quad \& \text{ LPI} > \text{t} \tag{1}$$

where LPI_{adj} [km^{-2} h^{-1}] denotes the adjusted LPI, i.e., LPI converted to the number of lightning flashes, parameter t denotes the minimum value of LPI for which a lightning flash is produced, k and l represent the parameters of a straight line. The parameters t, k and l are iterated across [0, 20], [0, 10], [-20, 20] intervals, respectively. For every combination of parameters t, k, and l, hourly means of LPI_{adj} are calculated. Then, a distribution function of both simulated and observed hourly means of lightning flashes during all cases is determined. Further, a root mean square error (RMSE) between the two discussed distributions is found. The optimal combination of parameters is the one that minimizes the RMSE. Here, a conversion is done by using $t = 0.045$, $k = 3.3$ and $l = 0.1$ for COSMO-crCLIM and $t = 0.65$, $k = 0.65$ and $l = -0.2$ for WRF for all cases. The discrepancy in optimal parameter values between WRF and COSMO highlights the discrepancies between LPI produced by COSMO-crCLIM and WRF. Namely, higher values of t and lower values of k associated with WRF indicate that WRF produces higher LPI, contrary to the results obtained for precipitation where WRF produced slightly lower precipitation amounts compared to COSMO. Since LPI is highly dependent on the updraft intensity and hydrometeor fields (as indicated by Equation S2.1), this discrepancy in t, k and l values between the two models could be due to the differences in simulated updrafts or differences in microphysical fields produced by COSMO-crCLIM and WRF. These discrepancies will be further discussed in Section 4.4.

Finally, to reduce uncertainties in simulating the temporal characteristics of observed convection (as seen in Figure 3), the daily sums of both LPI_{adj} produced by COSMO-crCLIM and WRF and the observed number of lightning flashes for each case analyzed are presented in Figure 5. Overall, it seems that the general spatial pattern of the observed lightning activity is well reproduced by both mod-

els although the simulated fields appear to be more scattered than the observed. This could be partially attributed to the fact that LPI is calculated every 15 min, while the LINET network detects lightning flashes continuously. Moreover, considering all cases, it is noted that the conversion of LPI to lightning flashes is better fitted towards less intense lightning activity. This is explained by the fact that the fit is performed on all grid points: as there are more grid points with low flash counts then intense lightning activity, the fit is intrinsically better for lower flash counts. The discrepancy in fit between lower and higher flash counts is more pronounced during the cases with more intense and widespread lightning activity, i.e., 25 June 2017 and 24 July 2017. Nonetheless, in general, the spatial distribution of lightning, i.e., the distribution of the areas with more and less intense lightning activity, corresponds well between simulated and observed fields, although local discrepancies could be present, depending on the case and model analyzed. Looking at the differences between fields produced by COSMO-crCLIM and WRF, a tendency of COSMO-crCLIM to produce more scattered and less peaked fields can be found.

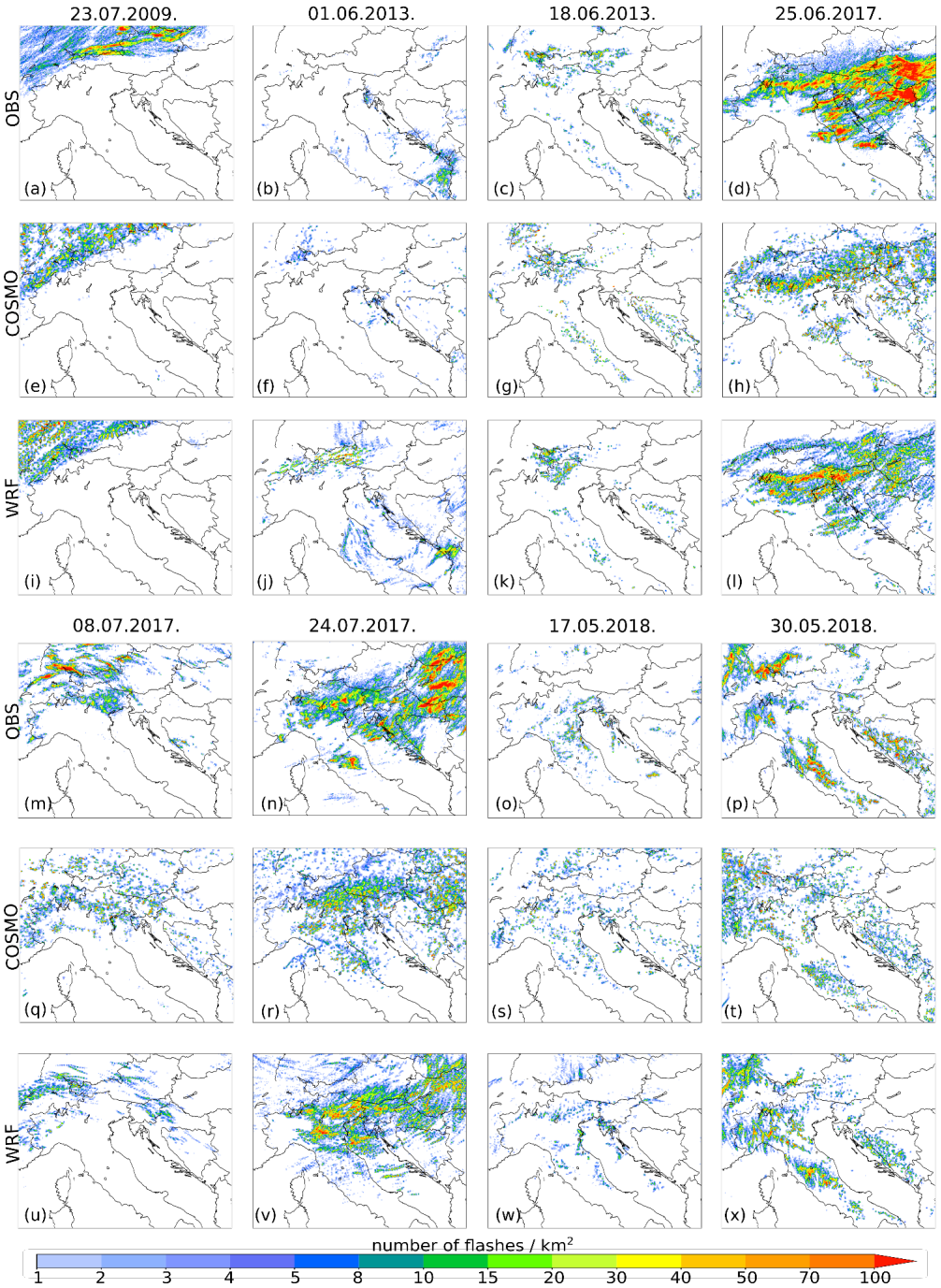

Figure 5. Simulated and measured lightning flash accumulation in the time window from 00 to 24 UTC on the day with severe convection. Columns denote

cases, while rows denote measurements from the (a-d; m-p) LINET network, and fields produced by (e-h; q-t) COSMO-crCLIM and (i-l; u-x) WRF models.

Furthermore, to quantitatively evaluate the capabilities of COSMO-crCLIM and WRF to simulate the observed lightning activity, a minimum coverage method is utilized. To get more robust results, the evaluation is done by aggregating all analyzed cases together and analyzing the daily sums of observed and simulated lightning activity (fields presented in Figure 5). Using the minimum coverage method combined with various radiuses of verification windows and various thresholds for the number of lightning flashes, contingency tables are constructed and SEDI index is calculated (Figure 6). Both models show similar performance which is better for the lower thresholds of lightning flashes. Moreover, we get good performance (SEDI > 0.6) even for more intense thresholds if we consider larger verification window sizes. WRF tends to have higher SEDI values than COSMO-crCLIM towards higher and more localized lightning flashes (bottom right side of diagrams), which confirms the previous findings that COSMO-crCLIM tends to produce more scattered lightning activity.

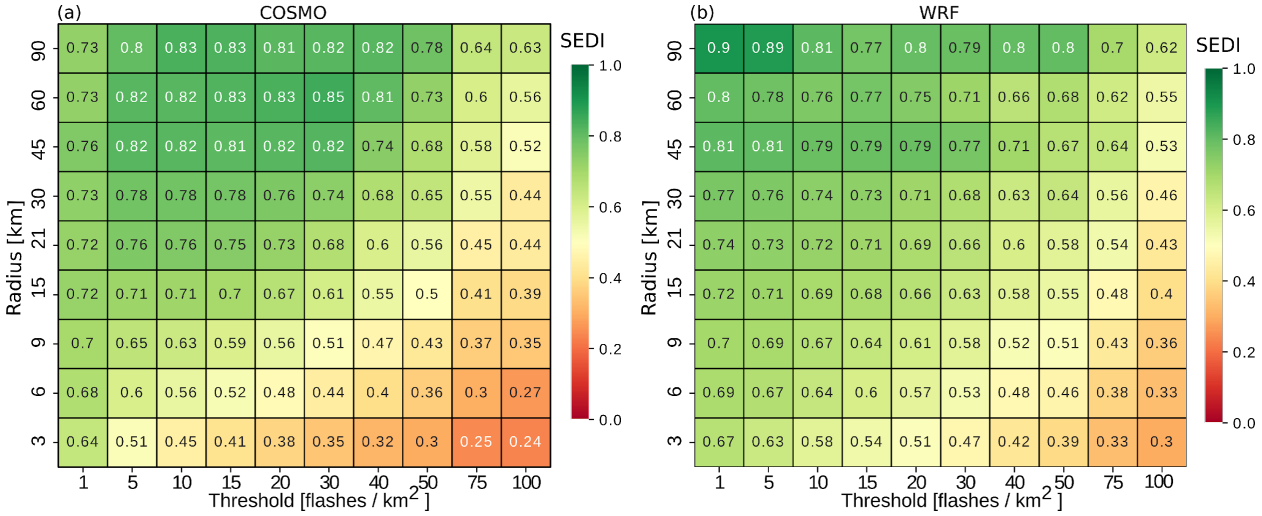

Figure 6. Performance of (a) COSMO-crCLIM and (b) WRF in simulating the observed lightning flashes. Performance depending on the threshold for the number of lightning flashes and verification window sizes (radius) is indicated in terms of SEDI skill score (shading). The higher/lower SEDI score means better/worse performance of the model, as reflected by the green/red colors.

4.2 HAILCAST results

HAILCAST results are assessed against remote-sensing and ground observations for a period from 00 UTC to 24 UTC on the day severe convection was observed. Moreover, to allow for possible temporal shifts between simulated and observed convection, we aggregate simulated and observed datasets over a 24 h period. First, we perform a qualitative comparison between hail swaths produced by

the two models. Figure 7 suggests that both models produced generally similar hail swaths over the same area of interest, although some local discrepancies between simulated hail swaths exist. Despite the overall similarity of the results, a tendency of COSMO-crCLIM to produce more hail in all analyzed cases is apparent. Interestingly, both models correctly reproduced heavy precipitation without hail over the Alps for 1 June 2013, which suggests that both models are able to distinguish intense precipitation events form hail events. However, it should be noted that both models still produced hail over only a few grid points over the Alps.

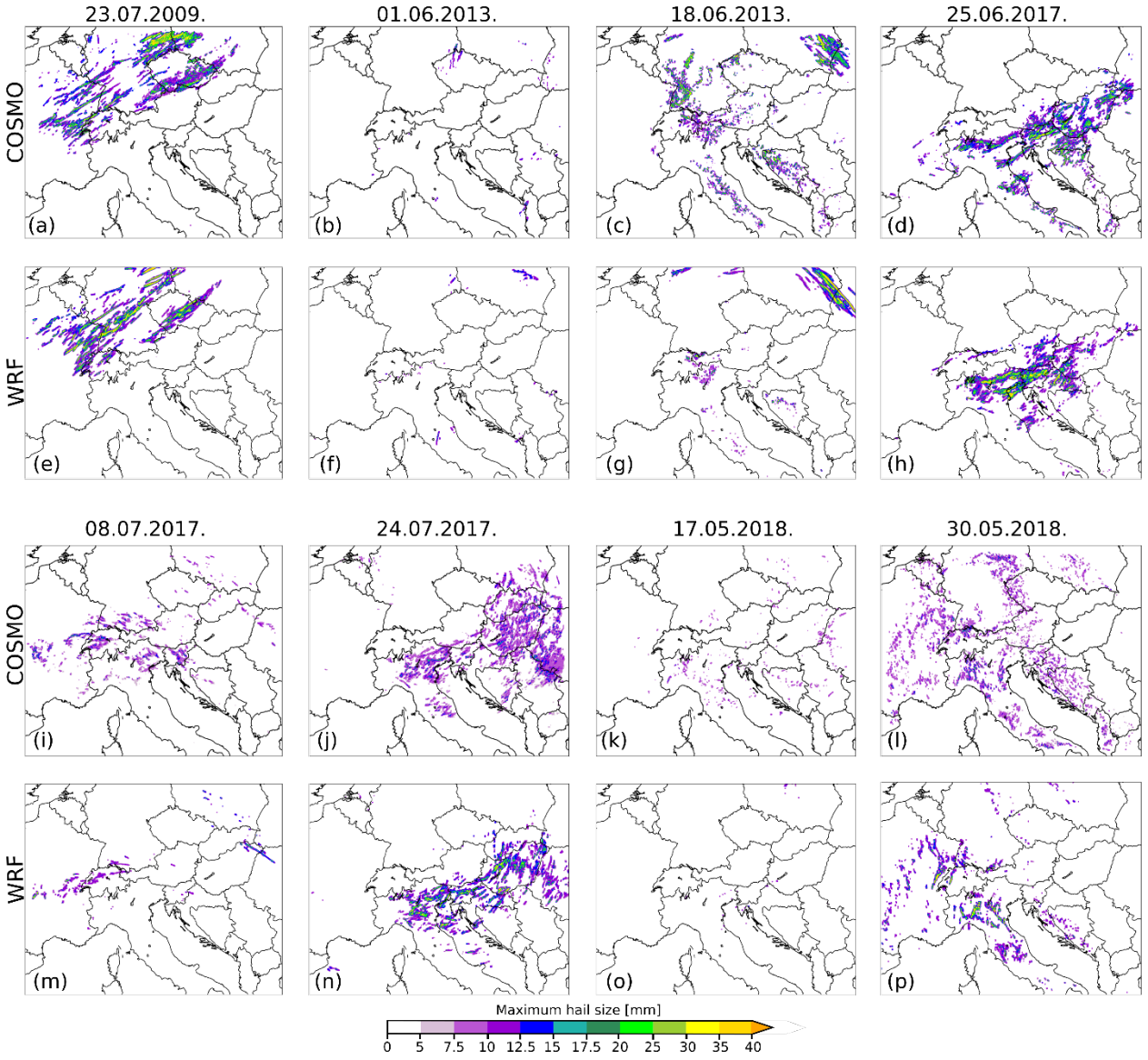

Figure 7. Maximum hailstone diameters in the time window from 00 UTC to 24 UTC on the day with severe convection simulated by COSMO-crCLIM (a)-(d); (i)-(l), and WRF (e)-(h); (m)-(p).

Further, we compare simulated fields against remote sensing observations of hail in Switzerland. Figure 8 shows the simulated and observed hail swaths over the Alpine region. It is clear that both models can produce hail swaths comparable to those observed, both in the context of the area affected by hail and the shapes of the observed hail swaths. Notably there is not an exact match between simulated and observed fields, as, some deviations are present. Interestingly, in most of the cases analyzed, WRF produces smaller hail swaths than COSMO-crCLIM. On the other hand, WRF simulates more grid points having maximum hailstone diameters greater than 35 mm.

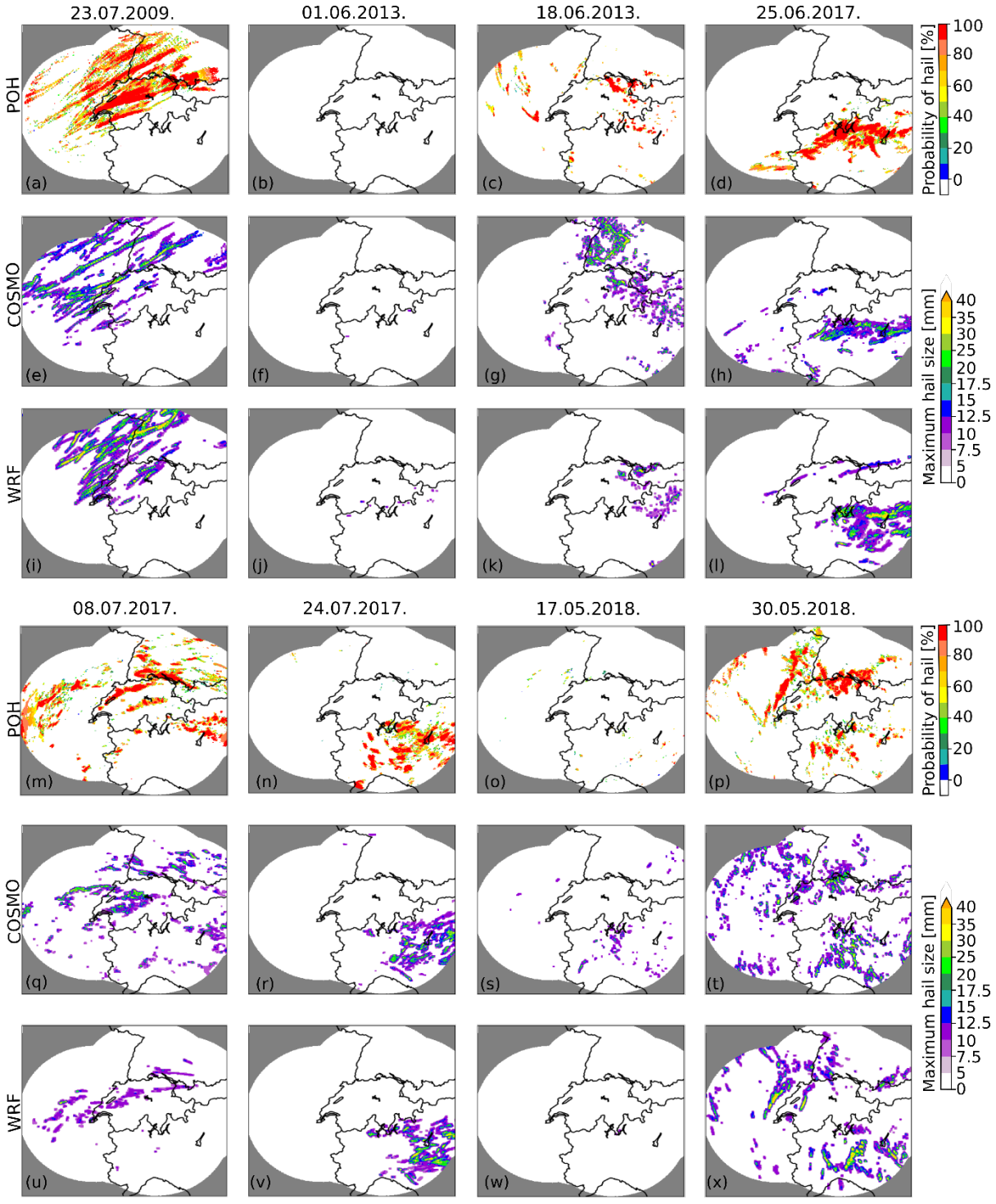
23.07.2009.
01.06.2013.
18.06.2013.
25.06.2017.
08.07.2017.
24.07.2017.
17.05.2018.
30.05.2018.
POH
COSMO
WRF
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
(p)
(q)
(r)
(s)
(t)
(u)
(v)
(w)
(x)
Probability of hail [%]
100
80
60
40
20
0
Maximum hail size [mm]
40
35
30
25
20
17.5
15
12.5
10
7.5
5
0

Figure 8. Daily maximums of the hailstone diameter simulated by (e)-(h); (q)-(t) COSMO-crCLIM and (i)-(l); (u)-(x) WRF and daily maximums of (a)-(d); (m)-(p) POH radar product.

Next, we assess simulated fields against another source of hail observations – hailpad measurements from the Croatian hailpad network. Out of eight cases with severe convection over the Alpine-Adriatic region, hailpads in Croatia recorded hail on only three of those days (25 June 2017, 24 July 2017, 17 May 2018). For these days, simulated hail swaths with indicated impacted hailpads are presented in Figure 9. There is a generally good agreement between observed and simulated hail produced by both models. Most of impacted hailpads lie in the area of simulated hail. However, both models exhibit a certain number of false alarms, i.e. hail is not observed, but the model simulated hail. Notably, some of these false alarms could be attributed to the limited spatial information on hail occurrence provided by the hailpad network. Unlike radars, the hailpad network provides information on hail occurrence only at the exact position where the hailpad is installed. In theory, hail could easily occur anywhere between the two hailpads and be left unrecorded. Nonetheless, the greatest number of false alarms is present for the 24 July 2017 case with WRF producing more false alarms than COSMO-crCLIM. Surprisingly, both models successfully reproduced even a highly localized hailstorm occurring on 17 May 2018 with a flat surface pressure distribution over the north-eastern Adriatic (Cui et al., 2023), although it should be noted that, unlike WRF, COSMO-crCLIM produces a few spurious false signals in the continental part of Croatia. Moreover, two of the analyzed cases, namely 25 June 2017 and 24 July 2017, were previously analyzed in Malečić et al. (2022). Even with different modeling settings compared to Malečić et al. (2022), i.e., different domains, horizontal resolutions, input data or HAILCAST activation time, WRF-HAILCAST produced similar hail swaths in both studies.

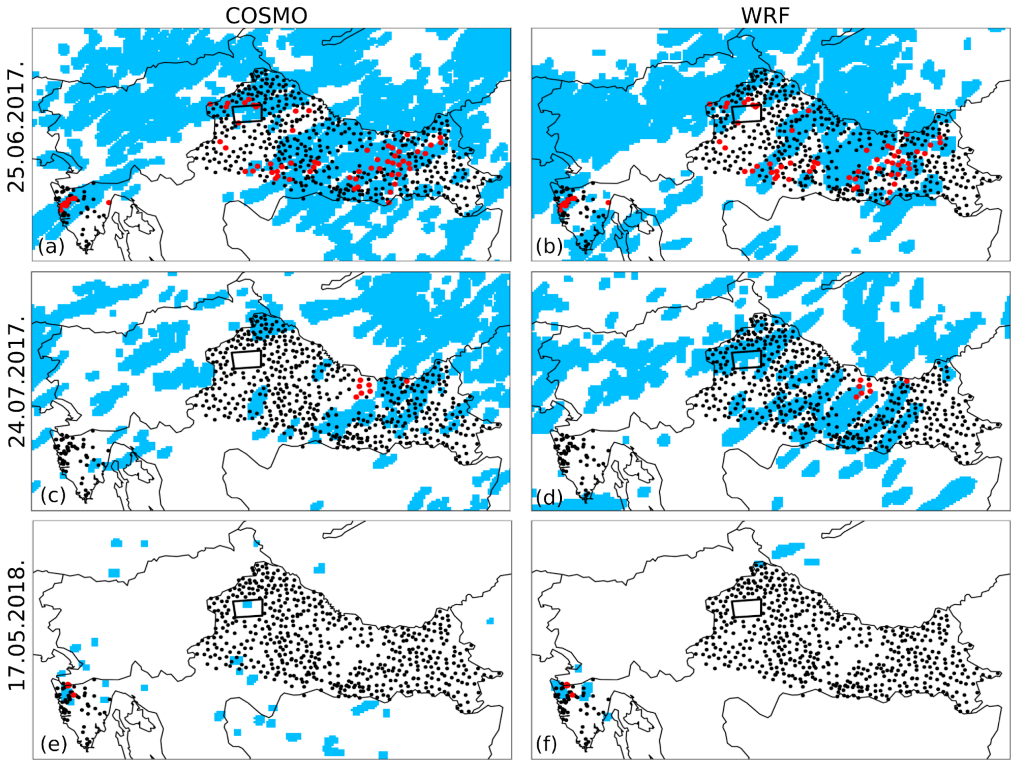

Figure 9. Simulated and recorded hail during the three cases with hail in Croatia. The shaded blue area represents simulated hail swaths (maximum hailstone diameter larger of equal to 5 mm) from 00 to 24 UTC on the day hail was observed. The position of hailpads is indicated with black dots. Impacted hailpads are marked with red circles. The position of a densely populated hailpad polygon is marked with a black rectangle and the stations within the polygon are colored only if the hail was observed at that specific station. Fields produced by (a, c, e) COSMO-crCLIM and (b, d, f) WRF are presented.

To quantitatively describe the results, simulated fields are evaluated against observations. To get more robust results, the evaluation is done by aggregating all analyzed cases together and analyzing the daily maximums of observed and simulated fields. More specifically, the fields presented in Figure 8 are assessed against POH $>$ 80 % signals, as the region corresponding to POH $>$ 80 % is highly probable to have hail on the ground (Nisi et al., 2016). For the analysis, POH product is interpolated to the model's grid and observed and simulated fields are evaluated using a minimum coverage verification method with varying verification window sizes, as already described in Section 3.3. Obtained performance diagrams presented in Figure 10a-b reveal that COSMO-crCLIM performs better in terms of POD and EDI skill scores for all considered verification window sizes. On the other hand, WRF performs better in terms of FAR for all verification window sizes except the one corresponding to 30 grid

points. These findings could be attributed to the fact that COSMO-crCLIM produces more hail compared to WRF which leads to higher POD and FAR values. Summarizing the insights obtained in Figure 8 and Figure 9a-b, it seems that COSMO-crCLIM produces hail swaths more similar to those observed over the Alpine region.

Next, simulated fields are evaluated against measurements from the hailpad network in Croatia using an upscaled verification method (as described in Section 3.3 and Malečić et al. (2022)). The obtained performance diagrams (Figure 10c-d) show similar performance between the models. High POD values for larger verification window sizes indicate that models simulated hail where it was observed. However, unlike the results connected with radars (Figure 10a-b), FAR values associated with hailpad network are much higher. That could be connected to the potential tendency of the model to overestimate the area affected by hail, if not also to the lack of spatially continuous information on hail occurrence in Croatia. Notably, there is a great contribution to the FAR values from the case on 24 July 2017 where both models produced a lot of false alarms. Interestingly, the same case, 24 July 2017, was also weakly represented in Malečić et al. (2022) with a lot of false alarms indicating a low predictability of the atmospheric conditions leading to the initiation and evolution of the observed convection.

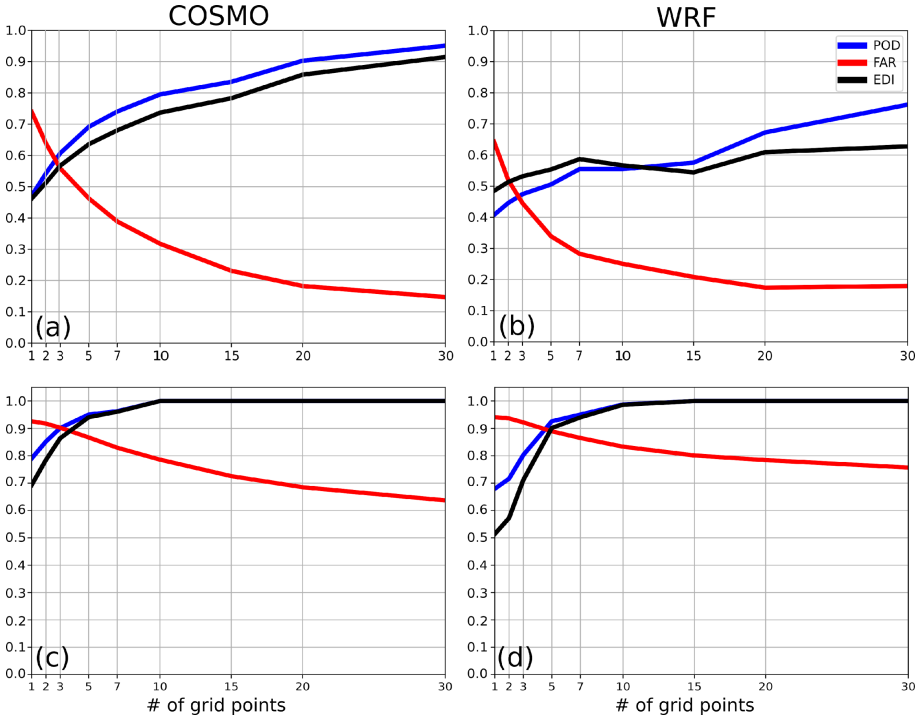

Figure 10. Performance of (a) COSMO-crCLIM and (b) WRF to simulate hail swaths as observed by the probability of hail (POH) radar product, and

performance of (c) COSMO-crCLIM and (d) WRF in simulating hail as observed by the Croatian hailpad network in terms of probability of detection (POD, blue), false-alarm rate (FAR, red) and extremal dependence index (EDI, black) skill scores. Performance depending on the verification window size is presented.

Hailpad networks, besides delivering information on hail occurrence, also provide information on hailstone sizes on the ground. Based on this information, a comparison of simulated and observed maximum hailstone diameters is performed. To account for possible spatial shifts between observed and simulated fields, a neighborhood inside a radius of 5 grid points (roughly corresponding to 12 km) of each impacted hailpad is scanned. The maximum simulated hailstone diameter inside this area is compared to the observed maximum hailstone diameter (Figure 11a). Both models underestimate the occurrence of smaller hailstones (diameters of 5-10 mm and 10-20 mm), and overestimate the occurrence of larger hailstones (diameters larger than 20 mm). However, when analyzing such results, one should proceed with caution, as it has been known from previous studies that hailpads are unlikely to record the largest hailstones given that they cover only 0.25 m^2 (e.g. Smith & Waldvogel, 1989). Indeed, to obtain a more realistic comparison between simulated and observed maximum hailstone diameter, it would be better to use the information on hailstone size observed by an observer – data which was not available for this study. Nonetheless, some tendencies could be extrapolated from the present comparison. Namely, COSMO-crCLIM mostly simulates hailstones in the 20-30 mm category, while WRF mostly simulates hailstones in the 30-50 mm category. Interestingly, WRF was able to reproduce an observed hailstone larger than 50 mm. Those differences between the two models are further confirmed if we compare the distribution of maximum hailstone sizes over the whole domain and all cases analyzed (Figure 11b). Here, it is clear that COSMO-crCLIM produces more hailstones in the 5-10 mm category than WRF, while WRF tends to produce more larger hailstones. These findings are consistent with Cui et al. (2023) who showed that COSMO tends to produce too many small hailstones and not enough larger ones compared to MESHS radar product.

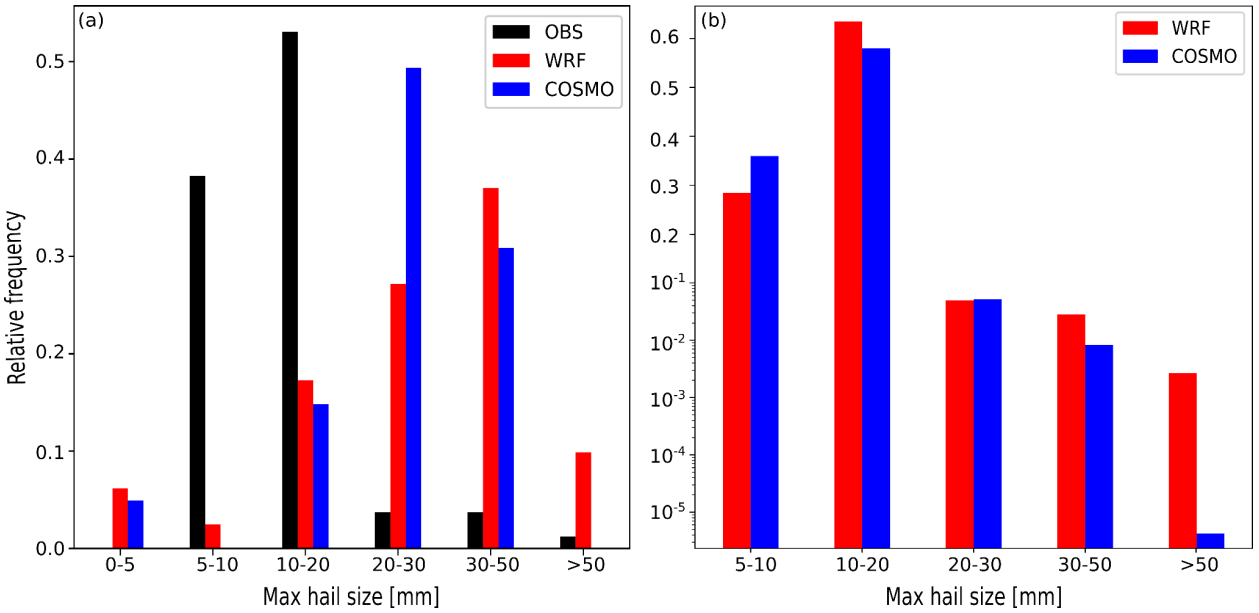

Figure 11. (a) Relative frequency of maximum recorded hailstone size from hailpads (black) and simulated maximum hailstone size by COSMO-crCLIM (blue) and WRF (red), and (b) relative frequency of simulated maximum hailstone sizes over the whole domain for COSMO-crCLIM (blue) and WRF (red). Histograms are normalized by dividing the count of hailstone sizes in each category with the total observed number of hailstones. To better depict differences between models and runs, the y-axis in (b) is partly linear and partly logarithmic.

To summarize the results, HAILCAST, whether integrated in WRF or COSMO, is able to reproduce observed hail swaths over the Alpine-Adriatic region as already reported by previous studies (Trefalt et al., 2018; Manzato et al., 2020; Raupach et al., 2021; Malečić et al., 2022; Tiesi et al., 2022; Cui et al., 2023). Further by comparing the simulated and observed maximum hailstone diameters, we find that both models tend to overestimate the maximum hailstone size compared to the Croatian hailpad network. Given the limitations of hailpad networks to record the largest hailstones, a detailed analysis with maximum hailstone sizes observed by an observer needs to be conducted to get more justifiable conclusions. Although the results are similar between the two models, some differences can be found. Overall, COSMO-crCLIM tends to produce larger areas covered by hail but smaller hailstones compared to WRF. Based on the analyzed cases, WRF tends to produce smaller areas covered by hail and wider spectra of hailstone sizes, while COSMO-crCLIM shows better performance in terms of simulating hail where it was observed over the whole domain, and the simulated hailstone sizes are somewhat closer to the observed ones in Croatia.

4.4 Differences between models and model internal variability

The analysis reveals that WRF tends to produce less precipitation, smaller hail swaths but higher values of LPI and more large hailstones compared to COSMO-crCLIM. Here, we study the potential origins of these differences, and consider the role of model internal variability in our results. For this reason, we form an ensemble of simulations with different initialization times for one of the cases with widespread hail and lightning across the Alpine-Adriatic region, namely the 30 May 2018. Both models were initialized at 06, 12 and 18 UTC the day before hail was observed (29 May 2018). Additionally, we utilize a simulation of the newest version of COSMO, namely COSMO 6.0, to further increase the ensemble size increase the robustness of our conclusions. The simulation using COSMO 6.0 is initialized at 12 UTC the day before hail was observed.

First, we analyze simulated daily precipitations fields between ensemble members (Figure 12). It is clear that all ensemble members produce precipitation patterns similar to the observed. Moreover, there is a greater difference in the fields produced by COSMO and WRF than between the members of the same model. This finding suggests that the differences in results are rather linked to systematic differences between models than to the model internal variability. However, it should be noted that differences between model members exist. Interestingly, both versions of COSMO produced comparable fields, although local differences are present that may be relevant for forecasting applications. For instance, the observed heavy precipitation along the French-German border is largely or completely missed in two of the COSMO ensemble members, but present in the member initialized at 06 UTC. This highlights the need for using ensemble techniques in numerical weather prediction (NWP) applications (Klasa et al., 2018).

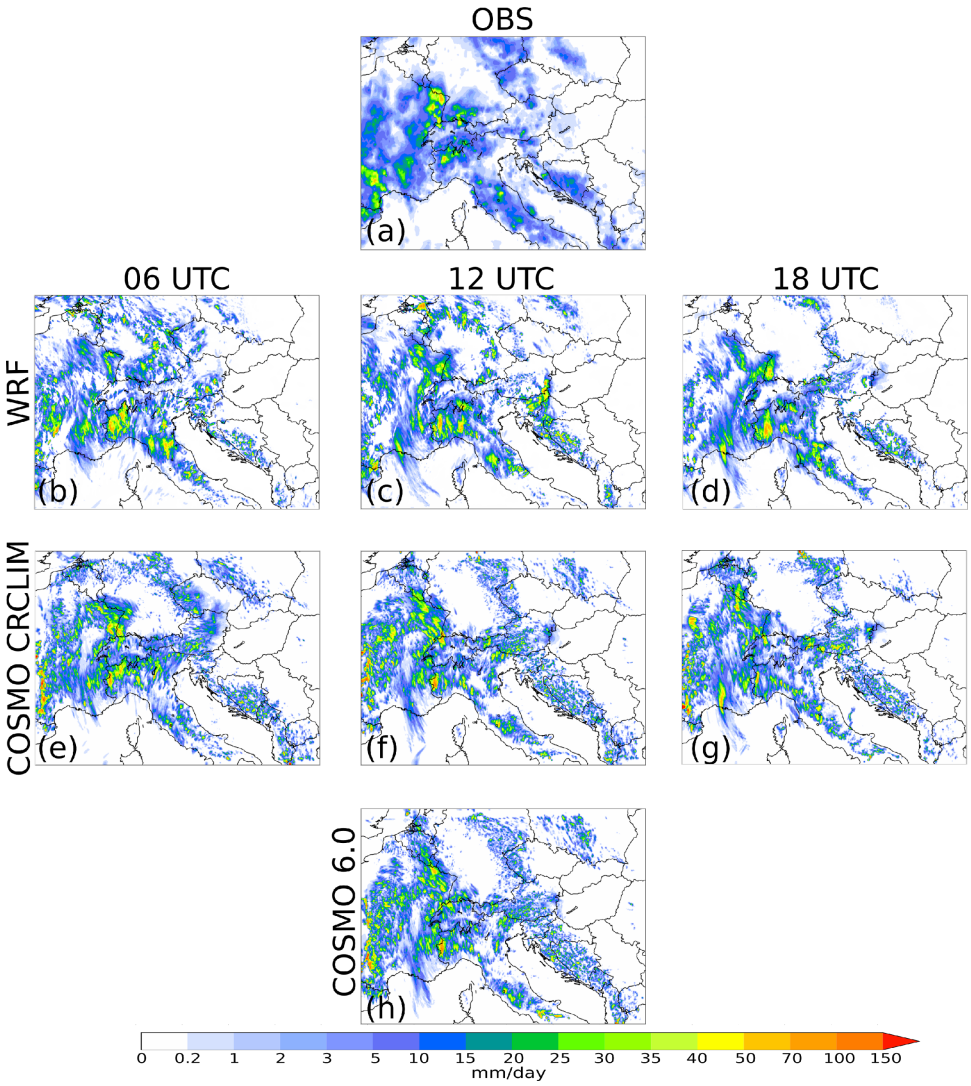

Figure 12. Accumulated precipitation for the period between 00 and 24 UTC on 30 May 2018. From top to bottom rows are (a) IMERG observation and simulated fields using (b-d) WRF, (e-g) COSMO-crCLIM, (h) COSMO 6.0. The columns from left to right represent the simulations initialized at 06, 12 and 18 UTC on the day before the event, respectively.

Similarly, daily maximums of LPI produced by each of the ensemble members are compared against daily sums of the number of lightning flashes from the LINET network. Here we use the raw LPI instead of the adjusted LPI because we want to avoid filtering the signals to have a direct comparison. It is clear that both models reproduced the area affected by lightning fairly well, although WRF produces higher values of LPI. This finding is consistent with the above

results for which we applied a higher threshold for WRF to convert LPI to the observed number of lightning flashes. Although there are differences in LPI between the model members, there are larger differences in LPI produced by the two different models. This suggests that the conclusions that WRF produces higher LPI than COSMO could be due to differences in simulating convection rather than to the model internal variability. Interestingly, COSMO 6.0 produced LPI in agreement with COSMO-crCLIM, although with slightly higher values.

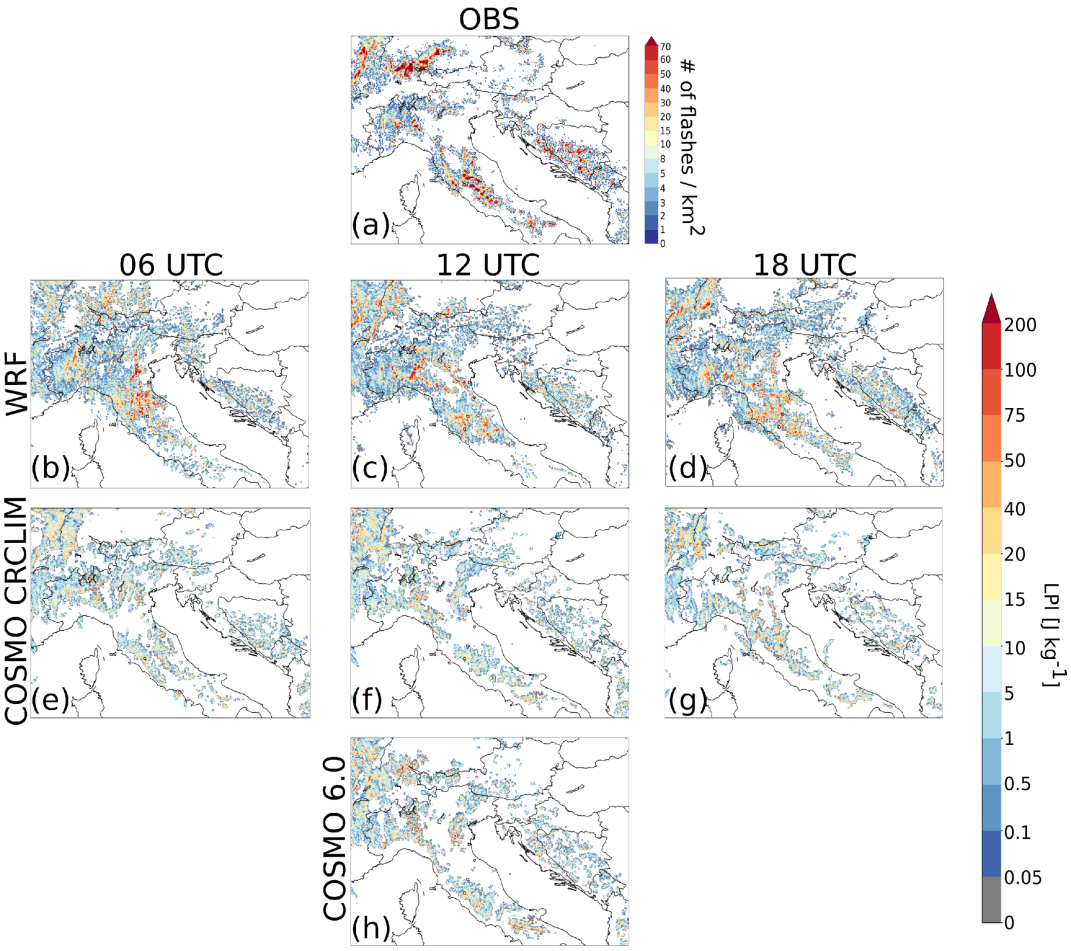

Figure 13. (a) Daily sum of the observed lightning flashes by the LINET network for the period between 00 and 24 UTC on 30 May 2018. Daily maximum of LPI produced by (b-d) WRF, (e-g) COSMO-crCLIM, and (h) COSMO 6.0. The columns represent simulations initialized at with 06, 12 and 18 UTC on the day before the event.

By comparing daily maximums of the hailstone diameters produced by ensemble

members for both spatial (Figure 14) and cumulative distribution (Figure 15), similar conclusions are found. Figure 14 shows that the simulated fields are overall similar, although WRF produces less hail compared to COSMO regardless of the initialization time. Hail produced by the same model but different initialization times (06, 12 and 18 UTC) and different model versions (COSMO-crCLIM and COSMO 6.0) is more similar than hail produced by different modeling systems (WRF vs COSMO). Additionally, when comparing simulated maximum hailstone diameters (Figure 15), we notice that, for hailstones smaller than 30 mm, the differences between COSMO and WRF are within each model internal variability. However, for hailstones larger than 30 mm, not only the differences between models becomes larger than model internal variability, but also it is clear that, out of all ensemble members, only WRF produced hailstones larger than 50 mm. This gives confidence to the overall conclusion that WRF produces more of the larger hailstones than COSMO (whatever its model version).

It should be noted that these findings are valid for one case only, and since the magnitude of the internal variability depends on the synoptic situation, model configuration, region and season (e.g. Lavin-Gullon et al., 2021), more cases should be analyzed to get more robust conclusions. Nonetheless, our results are encouraging and indicate that hailstorms occurring over the Alpine-Adriatic region can be simulated well using km-scale models, while the variability among different models and initialization times indicates that one could benefit from employing multi-model ensembles when simulating these events in an NWP context.

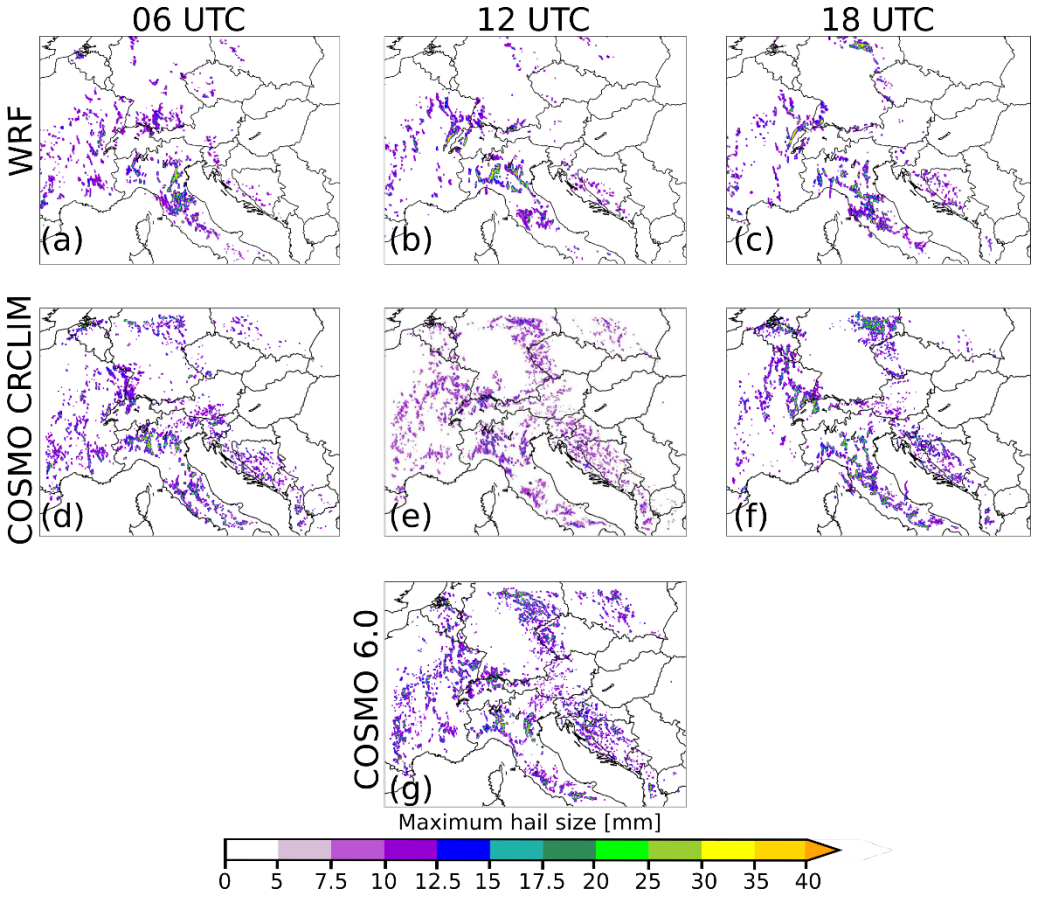

Figure 14. Daily maximum of hailstone size for the period between 00 and 24 UTC on 30 May 2018 simulated by (a-c) WRF, (d-f) COSMO-crCLIM, and (g) COSMO 6.0. The columns represent the simulations initialized at 06, 12 and 18 UTC on the day before the event.

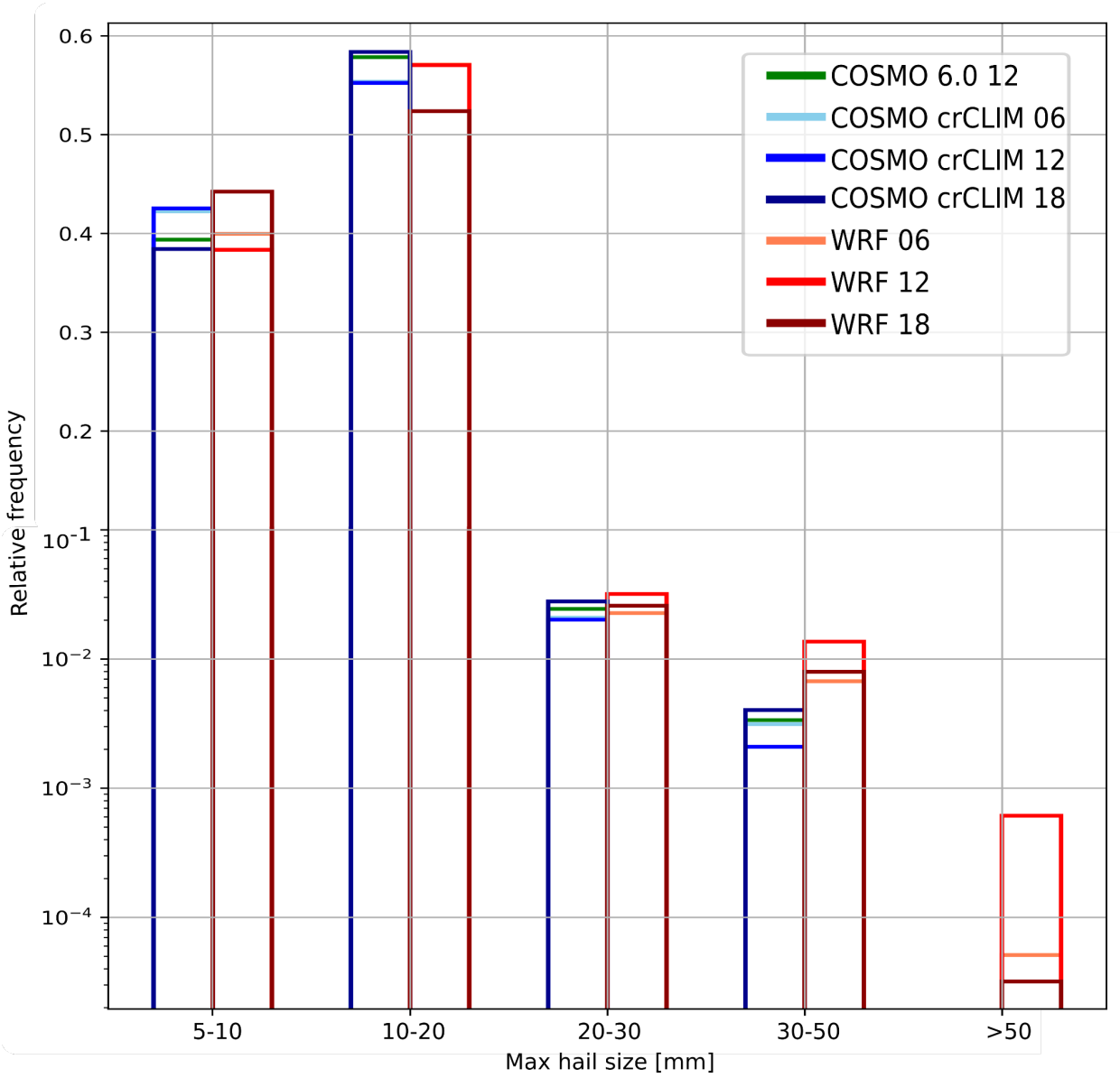

Figure 15. Relative frequency of the simulated maximum hailstone diameters over the whole domain for COSMO-crCLIM initialized at 06, 12 and 18 UTC (shades of blue), WRF initialized at 06, 12 and 18 UTC (shades of red) and COSMO 6.0 initialized at 12 UTC (green). The histograms are normalized by dividing the count of hailstone sizes in each category with the total number of grid points where hail occurs. To better depict differences between models and members, the y-axis is partly linear and partly logarithmic.

Further, considering the importance of the updrafts, as well as solid and liquid hydrometeors in the LPI and HAILCAST formulations, vertical profiles of these variables are compared among the models. Figure 16 presents vertical profiles averaged over time and grid points that have LPI greater than 0 during the 30 May 2018 case. The models produced different distributions of solid and liquid hydrometeors inside thunderclouds. Namely, WRF produces higher cloud water mixing ratios compared to both versions of COSMO and higher rain water mixing ratios compared to COSMO-crCLIM, but lower rain water mixing ratios compared to COSMO 6.0. Similarly, both versions of COSMO produce higher ice and graupel water mixing ratios, but lower snow water mixing ratios. Since in the LPI formulation the ratio between solid and liquid hydrometeors inside a thundercloud is more important than their exact values, total liquid water and ice fractional liquid ratio terms from the LPI formulation (Q_L and Q_I terms from Equation S2.2; S2.3) are computed alongside a dimensionless parameter representing the scaling factor for the updraft in the LPI formulation (Equation S2.1). obtains maximum values when total liquid water and ice fractional mixing ratios are equal (Equation S2.2). Surprisingly, analysis of shows no apparent discrepancies between the models even though there are some differences in Q_L and Q_I parameters. Interestingly, there is a difference in the simulated updrafts, i.e., both COSMO versions simulate, on average, weaker and higher updraft cores compared to WRF. WRF simulates stronger updrafts with the updraft core exactly at the position of the maximum value of coinciding with the region with nearly equal amounts of solid and liquid hydrometeors. Since the presence of both solid and liquid hydrometeors is important for lightning and hail growth processes, it is not surprising that WRF simulates higher LPI and more large hailstones compared to COSMO. On the other hand, the updraft core in both versions of COSMO is in the region with much more solid than liquid hydrometeors which is not as favorable for lightning or hail growth processes. This could be the reason why COSMO simulates lower LPI values and higher amounts of smaller hailstones. Interestingly, COSMO 6.0 simulates stronger updraft cores compared to COSMO-crCLIM which could explain why COSMO 6.0 produces slightly higher LPI values than COSMO-crCLIM (Figure 13) and shows a tendency to produce slightly larger hailstones compared to COSMO-crCLIM (Figure 15a).

Several studies reported a sensitivity of hail and lightning related variables such as updrafts and graupel mixing ratios on the choice of microphysics (Lagasio et al., 2017b; Trefalt et al., 2018; Manzato et al., 2020; Sokol & Minářová, 2020; Raupach et al., 2021), a combination of microphysics and planetary boundary layer parameterization scheme (Malečić et al., 2022), and large-scale forcing and initialization time (Manzato et al., 2020). Thus, different models with different configurations can produce large variability for different cases which encourages the use of a multi-model and/or multi-physics ensemble to analyze or forecast such events and to investigate the response of such events to further warming of the atmosphere.

Likewise, differences in updrafts strength and updraft structure strongly depend

upon the dynamical core of the models. Such differences can objectively be assessed using kinetic energy spectra (W. C. Skamarock, 2004). While both models considered here have similar dynamical cores using the split-explicit approach, there are significant differences in terms of advection schemes. The role of model formulation for heavy summer convection over Europe has recently been investigated in an intercomparison of the COSMO and the ECMWF-IFS models (Zeman et al., 2021). Results revealed a strong sensitivity with respect to the dynamical core (split-explicit versus spectral) but also with respect to time-step size as well as (explicit or implicit) numerical diffusion.

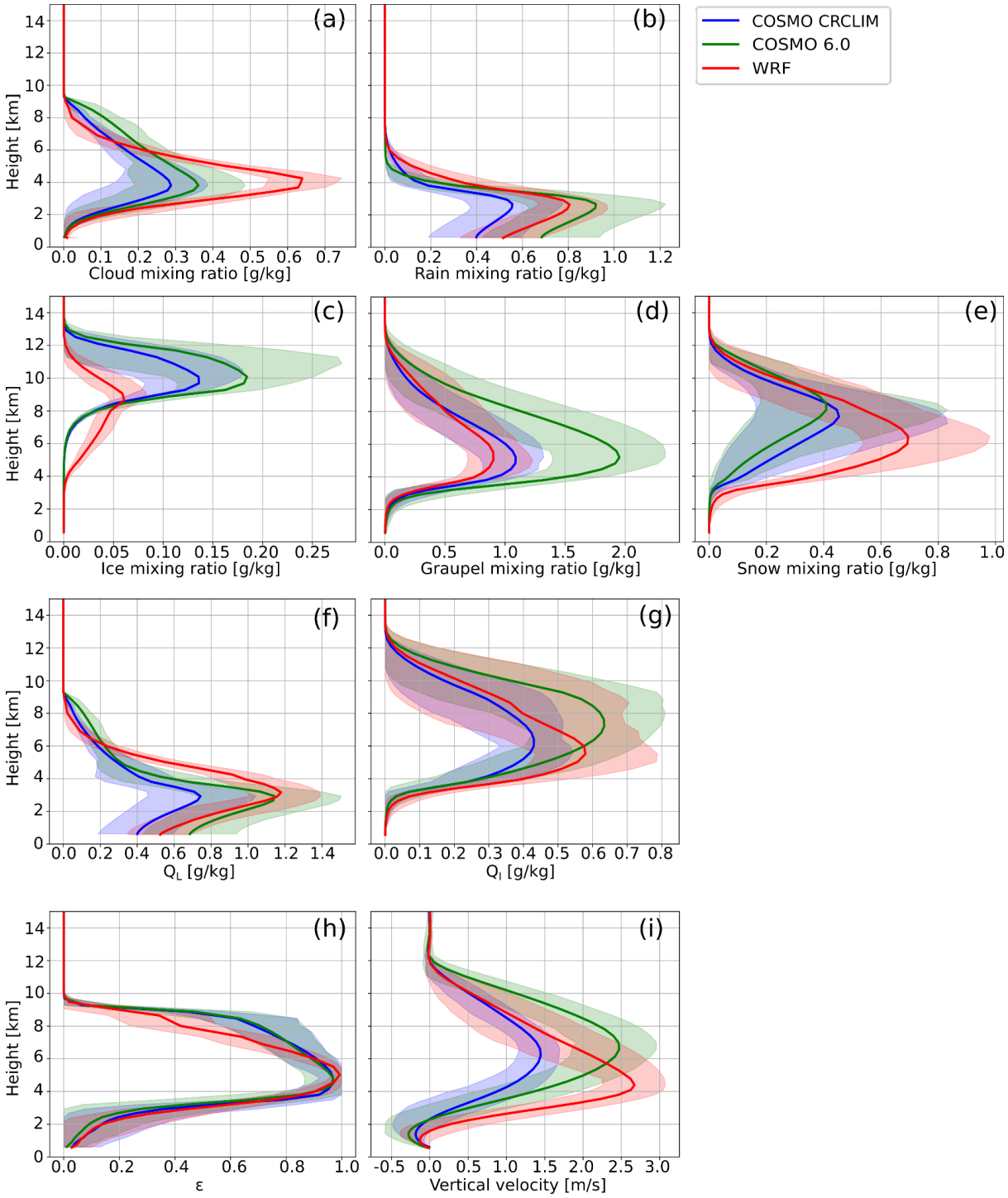
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
COSMO CRCLIM
COSMO 6.0
WRF
Height [km]
Cloud mixing ratio [g/kg]
Rain mixing ratio [g/kg]
Ice mixing ratio [g/kg]
Graupel mixing ratio [g/kg]
Snow mixing ratio [g/kg]
Q_L [g/kg]
Q_I [g/kg]
ε
Vertical velocity [m/s]

Figure 16. Vertical profiles of (a) cloud, and (b) rain mixing ratio; (c) ice, (d) graupel, (e) snow, (f) total liquid water, and (g) ice fractional mixing ratio; (h) updraft scaling parameter; and (i) vertical velocity as simulated by COSMO-crCLIM (blue), COSMO 6.0 (green) and WRF (red) models. The lines indicate the mean values across all grid points with LPI > 0, while the shading indicates the range between 5^{th} and 95^{th} percentile.

5 Conclusions

Hail and lightning, which are damaging and relatively frequent phenomena over the Alpine-Adriatic region, still remain difficult to model. Thus, this study employed two km-scale models, namely COSMO and WRF, with hail (HAILCAST) and lightning (LPI) diagnostic tools to simulate eight severe convective events occurring over the Alpine-Adriatic region. This study has been encouraged by and complement the two studies performed on WRF (Malečić et al., 2022) and COSMO (Cui et al., 2023) separately. The main aim was to analyze the robustness of HAILCAST and LPI results produced by the two different modeling systems, to explore their differences and to systematically and quantitatively evaluate the performance of each model. The main conclusions from this analysis can be summarized as follows:

- Both models reproduced the observed precipitation patterns and amounts, but WRF tended to produce slightly lower precipitation amounts than COSMO. Moreover, temporal evolution of precipitation is captured well in most cases.
- Both models showed good performance in reproducing the observed lightning activity despite WRF's tendency to simulate higher LPI values.
- Simulated hail swaths produced by the two models are overall similar, but a tendency of COSMO to produce more hail than WRF was found. Both models showed a good performance in reproducing hail observed by radar estimates over Switzerland and in-situ measurements over Croatia, although COSMO performed slightly better than WRF. Moreover, both models, on average, overestimated observed maximum hailstone diameters. Interestingly, a tendency of WRF to produce larger hailstones compared to COSMO was revealed, which is attributed to differences in model structures.

Furthermore, the potential origins of differences between the models and their internal variability are studied by employing an ensemble of simulations for one case with varying initialization times (06, 12, 18 UTC) using both COSMO and WRF models. Additionally, a simulation initialized at 12 UTC performed with the newest version of COSMO, namely COSMO 6.0, is added to the ensemble. The results for precipitation, lightning and hail are overall similar, but differences depending on the initialization time and modeling system exist. Moreover, the tendencies of WRF to produce less precipitation, smaller hail swaths but higher values of LPI and more large hailstones are present regardless of the initialization time. A careful analysis of the most important variables for hail and

lightning formation, namely updrafts and hydrometeor mixing ratios, revealed that the analyzed modeling systems produce, on average, different distributions of updrafts, solid and liquid hydrometeors inside thunderclouds. The different distributions of updrafts, solid and liquid hydrometeors inside thunderclouds could be linked to differences in model structures.

In conclusion, we show that atmospheric conditions leading to hailstorm formation and evolution are well simulated using state-of-the-art km-scale modeling systems. Moreover, diagnostic tools such as HAILCAST and LPI have a great potential for real-time forecasting and climatological assessment of hail and lightning in current and future climate. However, the variability of the results depending on the modeling system used encourages the use of a multi-model and/or multi-physics ensemble when modeling such events. Additionally, in an operational setting, data assimilation could further improve the predictability of such extreme events. Despite the promising results, it should be noted that this study is based on a small number of cases. To get statistically more robust conclusions a larger number of hailstorms needs to be analyzed. Moreover, this study would highly benefit from employing other data sources of hail observations covering the whole Alpine-Adriatic region. Nonetheless, given all limitations, this study represents the first attempt to systematically analyze and evaluate the performance of two intrinsically different km-scale modeling systems to reproduce the main characteristics of multiple hailstorms occurring over the Alpine-Adriatic region.

Acknowledgments

This research is enabled by the SWALDRIC (IZHRZO-180587) project, which is financed within the Croatian-Swiss Research Program of the Croatian Science Foundation and the Swiss National Science Foundation with funds obtained from the Swiss-Croatian Cooperation Programme. Lightning data were obtained from the Lightning Detection Network in Europe (LINET) (https://www.nowcast.de/en/solutions/linet-data; 5 June 2022). Authors acknowledge MeteoSwiss for providing radar observations (POH), Croatian Meteorological and Hydrological Service for providing hailpad observations and NASA for providing IMERG data. Hersbach et al., (2020) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. This research was supported by the WRF model (freely available at www.wrf-model.org/index.php). The WRF simulations were run using Isabella cluster (https://www.srce.unizg.hr/en/). RC, PV, NB, MED and CS acknowledge the Partnership for advanced computing in Europe (PRACE) for awarding access to Piz Daint at ETH Zürich at the Swiss National Supercomputing Centre (CSCS, Switzerland). The authors also acknowledge the Federal Office for Meteorology and Climatology (MeteoSwiss), CSCS, the Center for Climate Systems Modeling (C2SM) and ETH Zürich for their contributions to the

development and maintenance of the GPU-accelerated version of COSMO.

Open Research

Data for this research were obtained from three sources. Lightning data were obtained from the Lightning Detection Network in Europe (LINET) https://www.nowcast.de/en/solutions/linet-data (Betz et al., 2009). ERA 5 reanalysis fields used as initial and boundary conditions can be obtained through the following link (https://cds.climate.copernicus.eu/#!/home) while the IMERG dataset can be obtained through (https://gpm.nasa.gov/data/imerg). Hail measurements from the hailpad network in Croatia are available through inquiries of the Croatian Meteorological and Hydrological Service (usluge@cirus.dhz.hr) while radar products from Switzerland are available through inquiries at MeteoSwiss (Contact form).

Author contributions

BM, RC and PV did run the WRF, COSMO-crCLIM and COSMO 6.0 simulations, respectively. BM drafted the paper, with all co-authors providing input.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(42), 11770–11775. https://doi.org/10.1073/PNAS.1607171113/-/DCSUPPLEMENTALAdams-Selin, R. D., Clark, A. J., Melick, C. J., Dembek, S. R., Jirak, I. L., & Ziegler, C. L. (2019). Evolution of WRF-HAILCAST during the 2014-16 NOAA/Hazardous Weather Testbed Spring Forecasting Experiments. *Weather and Forecasting*, *34*(1), 61–79. https://doi.org/10.1175/WAF-D-18-0024.1Adams-Selin, R. D., & Ziegler, C. L. (2016). Forecasting hail using a one-dimensional hail growth model within WRF. *Monthly Weather Review*, *144*(12), 4919–4939. https://doi.org/10.1175/MWR-D-16-0027.1Allen, J. T., Giammanco, I. M., Kumjian, M. R., Jurgen Punge, H., Zhang, Q., Groenemeijer, P., Kunz, M., & Ortega, K. (2020). Understanding Hail in the Earth System. In *Reviews of Geophysics* (Vol. 58, Issue 1). Blackwell Publishing Ltd. https://doi.org/10.1029/2019RG000665Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., & Reinhardt, T. (2011). Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Monthly Weather Review*, *139*(12), 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1Ban, N., Schmidli, J., & Schär, C. (2014). Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research*, *119*(13), 7889–7907. https://doi.org/10.1002/2014JD021478Belušić, A., Prtenjak, M. T., Güttler, I., Ban, N., Leutwyler, D., & Schär, C. (2018). Near-surface wind variability over the broader Adriatic region: insights from an ensemble of regional climate models. *Climate Dynamics*,

50(11–12), 4455–4480. https://doi.org/10.1007/S00382-017-3885-5Berthet, C., Dessens, J., & Sanchez, J. L. (2011). Regional and yearly variations of hail frequency and intensity in France. *Atmospheric Research*, *100*(4), 391–400. https://doi.org/10.1016/J.ATMOSRES.2010.10.008Betz, H. D., Schmidt, K., Laroche, P., Blanchet, P., Oettinger, W. P., Defer, E., Dziewit, Z., & Konarski, J. (2009). LINET-An international lightning detection network in Europe. *Atmospheric Research*, *91*(2–4), 564–573. https://doi.org/10.1016/j.atmosres.2008.06.012Brimelow, J. C., Reuter, G. W., & Poolman, E. R. (2002). Modeling maximum hail size in Alberta thunderstorms. *Weather and Forecasting*, *17*(5), 1048–1062. https://doi.org/10.1175/1520-0434(2002)017<1048:MMHSIA>2.0.CO;2Brisson, E., Blahak, U., Lucas-Picher, P., Purr, C., & Ahrens, B. (2021). Contrasting lightning projection using the lightning potential index adapted in a convection-permitting regional climate model. *Climate Dynamics 2021*, *1*, 1–15. https://doi.org/10.1007/S00382-021-05791-ZBrisson, E., Brendel, C., Herzog, S., Ahrens, B., Brisson, E., Brendel, C., Herzog, S., & Ahrens, B. (2018). Lagrangian evaluation of convective shower characteristics in a convection-permitting model;Lagrangian evaluation of convective shower characteristics in a convection-permitting model. *MetZe*, *27*(1), 59–66. https://doi.org/10.1127/METZ/2017/0817Brisson, E., Van Weverberg, K., Demuzere, M., Devis, A., Saeed, S., Stengel, M., & van Lipzig, N. P. M. (2016). How well can a convection-permitting climate model reproduce decadal statistics of precipitation, temperature and cloud characteristics? *Climate Dynamics*, *47*(9–10), 3043–3061. https://doi.org/10.1007/S00382-016-3012-ZBrown, T. M., Pogorzelski, W. H., & Giammanco, I. M. (2015). Evaluating hail damage using property insurance claims data. *Weather, Climate, and Society*, *7*(3), 197–210. https://doi.org/10.1175/WCAS-D-15-0011.1Changnon, S. A. (2009). Increasing major hail losses in the U.S. *Climatic Change*, *96*(1), 161–166. https://doi.org/10.1007/s10584-009-9597-zCui, R., Ban, N., Demory, M. E., Schar, C. (2023). Exploring hail and lightning diagnostics over the Alpine-Adriatic region in a km-scale climate model. *Weather and Climate Dynamics* (in preparation) Curran, E. B., Holle, R. L., & Lopez, R. E. (2000). Lightning casualties and damages in the United States from 1959 to 1994. *Journal of Climate*, *13*(19), 3448–3464. https://doi.org/10.1175/1520-0442(2000)013<3448:LCADIT>2.0.CO;2Dessens, J. (1998). A physical evaluation of a hail suppression project with silver iodide ground burners in southwestern France. *Journal of Applied Meteorology*, *37*(12), 1588–1599. https://doi.org/10.1175/1520-0450(1998)037<1588:APEOAH>2.0.CO;2Dowdy, A. J., Fromm, M. D., & McCarthy, N. (2017). Pyrocumulonimbus lightning and fire ignition on Black Saturday in southeast Australia. *Journal of Geophysical Research: Atmospheres*, *122*(14), 7342–7354. https://doi.org/10.1002/2017JD026577Dudhia, J. (1989). Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*(20), 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2Ebert, E. E. (2008). Fuzzy verification of high-resolution gridded forecasts:

A review and proposed framework. *Meteorological Applications*, *15*(1), 51–64. https://doi.org/10.1002/met.25Federer, B., Schmid, W., & Waldvogel, A. (1978). The Design of Grossversuch IV, a Randomized Hail Suppression Experiment in Switzerland. *Atmosphere-Ocean*, *16*(1), 6–16. https://doi.org/10.1080/07055900.1978.9649009Ferro, C. A. T., & Stephenson, D. B. (2011). Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events. *Weather and Forecasting*, *26*(5), 699–713. https://doi.org/10.1175/WAF-D-10-05030.1Fiori, E., Comellas, A., Molini, L., Rebora, N., Siccardi, F., Gochis, D. J., Tanelli, S., & Parodi, A. (2014). Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: The Genoa 2011 case. *Atmospheric Research*, *138*, 13–29. https://doi.org/10.1016/j.atmosres.2013.10.007Foote, G. A., Krauss, T. W., & Makitov, V. (2005). Hail metrics using convectional Radar. In *Proceedings of the 16th Conference on Planned and Inadvertent Weather Modification*, San Diego, CA, USAFowler, H. J., Ali, H., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Cabi, N. S., Chan, S., Dale, M., Dunn, R. J. H., Ekström, M., Evans, J. P., Fosser, G., Golding, B., Guerreiro, S. B., Hegerl, G. C., Kahraman, A., Kendon, E. J., … Whitford, A. (2021). Towards advancing scientific knowledge of climate change impacts on short-duration rainfall extremes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2195). https://doi.org/10.1098/rsta.2019.0542Franc, B., Filipović-Grčić, B., & Milardić, V. (2016). Lightning overvoltage performance of 110 kV air-insulated substation. *Electric Power Systems Research*, *138*, 78–84. https://doi.org/10.1016/J.EPSR.2015.12.002Germann, U., Boscacci, M., Gabella, M., & Sartori, M. (2015). *Radar design for prediction in the Swiss Alps.*Giaiotti, D., Nordio, S., & Stel, F. (2003). The climatology of hail in the plain of Friuli Venezia Giulia. *Atmospheric Research*, *67–68*, 247–259. https://doi.org/10.1016/S0169-8095(03)00084-XHentgen, L., Ban, N., Kröner, N., Leutwyler, D., & Schär, C. (2019). Clouds in Convection-Resolving Climate Simulations Over Europe. *Journal of Geophysical Research: Atmospheres*, *124*(7), 3849–3870. https://doi.org/10.1029/2018JD030150Hersbach, 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., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/QJ.3803Holle, R. L., López, R. E., & Navarro, B. C. (2005). Deaths, Injuries, and Damages from Lightning in the United States in the 1890s in Comparison with the 1990s. *Journal of Applied Meteorology and Climatology*, *44*(10), 1563–1573. https://doi.org/10.1175/JAM2287.1Hong, S., & Lim, J. (2006). The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). *Undefined.*Horvath, K., Koracin, D., Vellore, R., Jiang, J., & Belu, R. (2012). Sub-kilometer dynamical downscaling of near-surface winds in complex terrain using WRF and MM5 mesoscale models. *Journal of Geophysical Research Atmospheres*, *117*(11). https://doi.org/10.1029/2012JD017432Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J.,

& Xie, P. (2019). *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Prepared for: Global Precipitation Measurement (GPM) National Aeronautics and Space Administration (NASA)*. https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpgJelić, D., Megyeri, O. A., Malečić, B., Belušić Vozila, A., Strelec Mahović, N., & Telišman Prtenjak, M. (2020). Hail Climatology Along the Northeastern Adriatic. *Journal of Geophysical Research: Atmospheres*, *125*(23). https://doi.org/10.1029/2020JD032749Jelić, D., Prtenjak, M. T., Malečić, B., Vozila, A. B., Megyeri, O. A., & Renko, T. (2021). A New Approach for the Analysis of Deep Convective Events: Thunderstorm Intensity Index. *Atmosphere 2021, Vol. 12, Page 908*, *12*(7), 908. https://doi.org/10.3390/ATMOS12070908Jewell, R., & Brimelow, J. (2009). Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Weather and Forecasting*, *24*(6), 1592–1609. https://doi.org/10.1175/2009WAF2222230.1Joe, P., Burgess, D., Potts, R., Keenan, T., Stumpf, G., & Treloar, A. (2004). The S2K severe weather detection algorithms and their performance. *Weather and Forecasting*, *19*(1), 43–63. https://doi.org/10.1175/1520-0434(2004)019<0043:TSSWDA>2.0.CO;2Kain, J. S., & Kain, J. (2004). The Kain - Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, *43*(1), 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2Kain, J. S., Weiss, S. J., Bright, D. R., Baldwin, M. E., Levit, J. J., Carbin, G. W., Schwartz, C. S., Weisman, M. L., Droegemeier, K. K., Weber, D., & Thomas, K. W. (2008). Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Weather and Forecasting*, 100804092600065. https://doi.org/10.1175/2008waf2007106.1Kain, J. S., Weiss, S. J., LevIt, J. J., Baldwin, M. E., & Bright, D. R. (2006). Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Weather and Forecasting*, *21*(2), 167–181. https://doi.org/10.1175/WAF906.1Kanata, J., Ametani, A., & Yamamoto, K. (2012). Threats of lightning current through an electric vehicle. *2012 31st International Conference on Lightning Protection, ICLP 2012*. https://doi.org/10.1109/ICLP.2012.6344299Keller, M., Fuhrer, O., Schmidli, J., Stengel, M., Stöckli, R., & Schär, C. (2016). Evaluation of convection-resolving models using satellite data: The diurnal cycle of summer convection over the Alps. *Meteorologische Zeitschrift*, *25*(2), 165–179. https://doi.org/10.1127/METZ/2015/0715Klasa, C., Arpagaus, M., Walser, A., & 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. https://doi.org/10.1002/QJ.3245Kunz, M., Blahak, U., Handwerker, J., Schmidberger, M., Punge, H. J., Mohr, S., Fluck, E., & Bedka, K. M. (2018). The severe hailstorm in southwest Germany on 28 July 2013: characteristics, impacts and meteorological conditions. *Quarterly Journal of the Royal Meteorological Society*, *144*(710), 231–250. https://doi.org/10.1002/qj.3197Lagasio, M., Parodi, A., Procopio, R., Rachidi,

F., & Fiori, E. (2017a). Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: The Genoa 2014 event. *Journal of Geophysical Research*, *122*(8), 4238–4257. https://doi.org/10.1002/2016JD026115Lagasio, M., Parodi, A., Procopio, R., Rachidi, F., & Fiori, E. (2017b). Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: The Genoa 2014 event. In *Journal of Geophysical Research* (Vol. 122, Issue 8, pp. 4238–4257). https://doi.org/10.1002/2016JD026115Latham, D., & Williams, E. (2001). Lightning and Forest Fires. *Forest Fires*, 375–418. https://doi.org/10.1016/B978-012386660-8/50013-1Lavin-Gullon, A., Fernandez, J., Bastin, S., Cardoso, R. M., Fita, L., Giannaros, T. M., Goergen, K., Gutierrez, J. M., Kartsios, S., Katragkou, E., Lorenz, T., Milovac, J., Soares, P. M. M., Sobolowski, S., & Warrach-Sagi, K. (2021). Internal variability versus multi-physics uncertainty in a regional climate model. *International Journal of Climatology*, *41*(S1), E656–E671. https://doi.org/10.1002/joc.6717Lee, J. Y., & Collins, G. J. (2017). Risk analysis of lightning effects in aircraft systems. *IEEE Aerospace Conference Proceedings*. https://doi.org/10.1109/AERO.2017.7943671Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O., & Schär, C. (2017). Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres*, *122*(10), 5237–5258. https://doi.org/10.1002/2016JD026013Lynn, B. H., & Yair, Y. Y. (2008). Lightning Power Index: A new tool for predicting the lightning density and the potential for extreme rainfall. In *Geophysical Research Abstracts* (Vol. 10).Malečić, B., Prtenjak, M. T., Horvath, K., Jelić, D., Jurković, P. M., Ćorko, K., & Mahović, N. S. (2022). Performance of HAILCAST and the Lightning Potential Index in simulating hailstorms in Croatia in a mesoscale model – Sensitivity to the PBL and microphysics parameterization schemes. *Atmospheric Research*, *272*(March), 106143. https://doi.org/10.1016/j.atmosres.2022.106143Manzato, A. (2008). A verification of numerical model forecasts for sounding-derived indices above Udine, northeast Italy. *Weather and Forecasting*, *23*(3), 477–495. https://doi.org/10.1175/2007WAF2007018.1Manzato, A. (2012). Hail in Northeast Italy: Climatology and Bivariate Analysis with the Sounding-Derived Indices. *Journal of Applied Meteorology and Climatology*, *51*(3), 449–467. https://doi.org/10.1175/JAMC-D-10-05012.1Manzato, A., Riva, V., Tiesi, A., & Marcello Miglietta, M. (2020). Observational analysis and simulations of a severe hailstorm in northeastern Italy. *Quarterly Journal of the Royal Meteorological Society*, *146*(732), 3587–3611. https://doi.org/10.1002/qj.3886Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research Atmospheres*, *102*(14), 16663–16682. https://doi.org/10.1029/97jd00237Morel, S. (2014). *Verification of radar-based hail detection algorithms with insurance loss data in Switzerland*. 83.Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*,

119(2), 397–407. https://doi.org/10.1007/s10546-005-9030-8Nisi, L., Martius, O., Hering, A., Kunz, M., & Germann, U. (2016). Spatial and temporal distribution of hailstorms in the Alpine region: A long-term, high resolution, radar-based analysis. *Quarterly Journal of the Royal Meteorological Society*, *142*(697), 1590–1604. https://doi.org/10.1002/qj.2771Park, S. H., Skamarock, W. C., Klemp, J. B., Fowler, L. D., & Duda, M. G. (2013). Evaluation of global atmospheric solvers using extensions of the jablonowski and williamson baroclinic wave test case. *Monthly Weather Review*, *141*(9), 3116–3129. https://doi.org/10.1175/MWR-D-12-00096.1Pichelli, 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., … 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*(11–12), 3581–3602. https://doi.org/10.1007/s00382-021-05657-4Pocakal, D. (2011). Hailpad data analysis for the continental part of Croatia. *Meteorologische Zeitschrift*, *20*(4), 441–447. https://doi.org/10.1127/0941-2948/2011/0263Počakal, D., Večenaj, Ž., Mikuš Jurković, P., & Grisogono, B. (2018). Analysis of orographic influence on hail parameters in NW Croatia. *International Journal of Climatology*, *38*(15), 5646–5658. https://doi.org/10.1002/joc.5769Počakal, D., Večenaj, Ž., & Štalec, J. (2009). Hail characteristics of different regions in continental part of Croatia based on influence of orography. *Atmospheric Research*, *93*(1–3), 516–525. https://doi.org/10.1016/j.atmosres.2008.10.017Poolman, E. R., (1992). Die voorspelling van haelkorrelgroei in Suid-Afrika (The forecasting of hail growth in South Africa). M.S. thesis, Faculty of Engineering, University of Pretoria, 113 pp.Prein, A. F., Holland, G. J., Rasmussen, R. M., Done, J., Ikeda, K., Clark, M. P., & Liu, C. H. (2013). Importance of Regional Climate Model Grid Spacing for the Simulation of Heavy Precipitation in the Colorado Headwaters. *Journal of Climate*, *26*(13), 4848–4857. https://doi.org/10.1175/JCLI-D-12-00727.1Púcik, T., Castellano, C., Groenemeijer, P., Kühne, T., Rädler, A. T., Antonescu, B., & Faust, E. (2019). Large hail incidence and its economic and societal impacts across Europe. *Monthly Weather Review*, *147*(11), 3901–3916. https://doi.org/10.1175/MWR-D-19-0204.1Punge, H. J., & Kunz, M. (2016). Hail observations and hailstorm characteristics in Europe: A review. *Atmospheric Research*, *176–177*, 159–184. https://doi.org/10.1016/j.atmosres.2016.02.012Rachidi, F., Rubinstein, M., Montanyà, J., Bermúdez, J. L., Sola, R. R., Solà, G., & Korovkin, N. (2008). A review of current issues in lightning protection of new-generation wind-turbine blades. *IEEE Transactions on Industrial Electronics*, *55*(6), 2489–2496. https://doi.org/10.1109/TIE.2007.896443Raupach, T. H., Martynov, A., Nisi, L., Hering, A., Barton, Y., & Martius, O. (2021). Object-based analysis of simulated thunderstorms in Switzerland: application and validation of automated thunderstorm tracking on simulation data. *Geoscientific Model Development Discussions*, *May*, 1–29. https://doi.org/10.5194/gmd-2021-105Reinhardt, T., & Seifert, A. (2006). A Three-Category Ice Scheme For LMK. *COS-*

MONewsletter, *6*, 115–120.Ritter, B., & Geleyn, J. F. (1992). A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations. Monthly Weather Review, 120, 303-325. - References - Scientific Research Publishing. *Monthly Weather Review*, *120*, 303–325.Romps, D. M., Seeley, J. T., Vollaro, D., & Molinari, J. (2014). Projected increase in lightning strikes in the united states due to global warming. *Science*, *346*(6211), 851–854. https://doi.org/10.1126/science.1259100Schär, 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., & Wernli, H. (2020). Kilometer-Scale Climate Models: Prospects and Challenges. *Bulletin of the American Meteorological Society*, *101*(5), E567–E587. https://doi.org/10.1175/BAMS-D-18-0167.1Schuster, S. S., Blong, R. J., Leigh, R. J., & McAneney, K. J. (2005). Characteristics of the 14 April 1999 Sydney hailstorm based on ground observations, weather radar, insurance data and emergency calls. *Natural Hazards and Earth System Science*, *5*(5), 613–620. https://doi.org/10.5194/nhess-5-613-2005Sioutas, M., Meaden, T., & Webb, J. D. C. (2009). Hail frequency, distribution and intensity in Northern Greece. *Atmospheric Research*, *93*(1–3), 526–533. https://doi.org/10.1016/J.ATMOSRES.2008.09.023Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., & Huang, X. (2019). A Description of the Advanced Research WRF Model Version 4. *Undefined*. https://doi.org/10.5065/1DFH-6P97Skamarock, W. C. (2004). Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Monthly Weather Review*, *132*(12), 3019–3032. https://doi.org/10.1175/MWR2830.1Smith, P. L., & Waldvogel, A. (1989). On Determinations of Maximum Hailstone Sizes from Hallpad Observations. *Journal of Applied Meteorology*, *28*(1), 71–76. https://doi.org/10.1175/1520-0450(1989)028<0071:odomhs>2.0.co;2Sokol, Z., & Minářová, J. (2020). Impact of 1- and 2-moment cloud microphysics and horizontal resolution on lightning Potential Index within COSMO NWP model. *Atmospheric Research*, *237*. https://doi.org/10.1016/j.atmosres.2020.104862Svabik, O. (1989). Review of meteorological aspects on hail defense activities in Austria. *Theoretical and Applied Climatology 1989 40:4*, *40*(4), 247–254. https://doi.org/10.1007/BF00865975Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research Atmospheres*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719Thornton, J. A., Virts, K. S., Holzworth, R. H., & Mitchell, T. P. (2017). Lightning enhancement over major oceanic shipping lanes. *Geophysical Research Letters*, *44*(17), 9102–9111. https://doi.org/10.1002/2017GL074982Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*(8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2Tiesi, A., LastNameMazzà, S., Conte, D., Ricchi, A., Baldini, L., Montopoli, M., Picciotti, E., Vulpiani, G., Ferretti, R., & Miglietta, M. M. (2022). Numerical Simulation of a Giant-Hail-Bearing

Mediterranean Supercell in the Adriatic Sea. *Atmosphere 2022, Vol. 13, Page 1219*, *13*(8), 1219. https://doi.org/10.3390/ATMOS13081219Trefalt, S., Martynov, A., Barras, H., Besic, N., Hering, A. M., Lenggenhager, S., Noti, P., Röthlisberger, M., Schemm, S., Germann, U., & Martius, O. (2018). A severe hail storm in complex topography in Switzerland - Observations and processes. *Atmospheric Research*, *209*, 76–94. https://doi.org/10.1016/j.atmosres.2018.03.007Vergara-Temprado, J., Ban, N., Panosetti, D., Schlemmer, L., & Schär, C. (2020). Climate Models Permit Convection at Much Coarser Resolutions Than Previously Considered. *Undefined*, *33*(5), 1915–1933. https://doi.org/10.1175/JCLI-D-19-0286.1Waldvogel, A., Federer, B., & Grimm, P. (1979). Criteria for detection of hail cells. *Journal of Applied Meteorology*, *18*(12), 1521–1525. https://doi.org/10.1175/1520-0450(1979)018<1521:CFTDOH>2.0.CO;2Willemse, S., & Furger, M. (2016). *From weather observations to atmospheric and climate sciences in Switzerland. Celebrating 100 years of the Swiss Society of Meteorology*.Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D., & Thomas, K. W. (1998). An enhanced hail detection algorithm for the WSR-88D. *Weather and Forecasting*, *13*(2), 286–303. https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2Yair, Y., Lynn, B., Price, C., Kotroni, V., Lagouvardos, K., Morin, E., Mugnai, A., & Del Carmen Llasat, M. (2010). Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields. *Journal of Geophysical Research Atmospheres*, *115*(4). https://doi.org/10.1029/2008JD010868Zeman, C., Wedi, N. P., Dueben, P. D., Ban, N., & Schär, C. (2021). Model intercomparison of COSMO 5.0 and IFS 45r1 at kilometer-scale grid spacing. *Geosci. Model Dev*, *14*, 1–23. https://doi.org/10.5194/gmd-14-1-2021