Atmospheric Research 272 (2022) 106143

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

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

Barbara Malečić [a,*], Maja Telišman Prtenjak [a], Kristian Horvath [b], Damjan Jelić [a], Petra Mikuš Jurković [b], Karol Ćorko [c], Nataša Strelec Mahović [d]

[a] Department of Geophysics, Faculty of Science, University of Zagreb, Zagreb, Croatia
[b] Croatian Meteorological and Hydrological Service (DHMZ), Zagreb, Croatia
[c] Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[d] EUMETSAT, Darmstadt, Germany

ARTICLE INFO

Keywords:
Hail
Lightning
WRF
Numerical weather prediction
Severe weather
Hailpad

ABSTRACT

Hailstorms, although extremely damaging severe weather hazards, remain a very challenging phenomenon to predict. To better understand dynamic processes and model performance, which can be helpful in forecasting hailstorms, three selected hailstorms in Croatia are simulated with the WRF model at convection-permitting (1 km) grid spacing using the HAILCAST module. In addition, the performance of the Lightning Potential Index (LPI) algorithm in representing the observed lightning activity during the selected hailstorms is analyzed. A multiphysics ensemble of 12 sensitivity simulations with the combinations of four different microphysics and three different planetary boundary layer parameterization (PBL) schemes is adopted to assess the forecasting ability of HAILCAST and LPI and their sensitivity to the choice of microphysics and PBL parameterization schemes. First, the model's ability to reproduce surface measurements of 2-m temperature, 2-m relative humidity, 2-m equivalent potential temperature and 10-m wind are examined using root mean square error (RMSE) decomposition. Then, the LPI is assessed against lightning observations via the object-based Structure-Amplitude-Location (SAL) method. Finally, an upscaled neighborhood verification method is proposed to assess HAILCAST against hail observations from the Croatian hailpad network. The results show that the observed hail and lightning activity is represented well by the model. There is a greater sensitivity to the choice of microphysics scheme than the PBL scheme, with National Severe Storms Laboratory double-moment scheme (NSSL2) microphysics scheme differing the most among the entire sensitivity ensemble. Nonetheless, both HAILCAST and LPI show promising performance in simulating observed hail and lightning activity, although HAILCAST tends to overestimate the area affected by hail. Nonetheless, the discrepancies between model configurations highlight the importance of simulating convection correctly to obtain a meaningful forecast of hail and lightning.

1. Introduction

Progress in predicting the occurrence and intensity of hailstorms is of great importance in limiting the harmful consequences of this severe weather hazard. Every year, hailstorms cause considerable damage to buildings, agriculture, and vehicles, resulting in substantial economic and insured losses. The crop and property damage even from individual hailstorms can exceed $1 billion U.S. dollars (Brown et al., 2015; Changnon, 2009; Kunz et al., 2018; Púcik et al., 2019; Schuster et al., 2005). Although the importance of successful hail forecasting cannot be stressed enough, both for damage mitigation and investigation of hail properties in a changing climate, hail remains a difficult phenomenon to model. This is due to existing gaps in understanding the microphysical and dynamic processes involved in hail formation, which stem from limited direct observations of hail and difficulties associated with running models at sufficiently high resolutions to provide valuable hail information (Raupach et al., 2021).

Most hail-forecasting methods are currently based on the time extrapolation of hailstorm characteristics observed by remote sensing methods, mostly radar and lightning data (Farnell et al., 2018, 2017;

* Corresponding author.
E-mail address: barbara.malecic@gfz.hr (B. Malečić).

https://doi.org/10.1016/j.atmosres.2022.106143
Received 22 September 2021; Received in revised form 16 February 2022; Accepted 12 March 2022
Available online 18 March 2022

Nisi et al., 2014). In recent years, machine learning methods for hail forecasting have been developed (Czernecki et al., 2019; Gagne et al., 2017; Manzato, 2013; Marzban and Witt, 2001). However, machine learning methods trained on observational data can generally make useful predictions only up to a few hours ahead. Although predictions for one or more days in advance are possible (but rare), such machine learning models require large datasets of observations, and historical numerical weather prediction (NWP) model runs for training purposes (Gagne et al., 2017).

On the other hand, over the years, hail-forecasting methods based on a combination of approximations of the convective environment responsible for hailstorm formation and local hail climatology (Johns and Doswell, 1992) have been developed. Most existing hail-forecasting models use the estimation of updraft strength based on sounding observations to provide forecasts on hailstone size (Brimelow et al., 2002; Fawbush and Miller, 1953; Moore and Pino, 1990). This approach has practical limitations since sounding observations are mostly available only twice a day at discrete locations.

As computational power has advanced over the years, it has become possible to replace sounding observations with spatially continuous (i.e., modelled) vertical profiles. A physically based one-dimensional hail model called HAILCAST is one of the sounding-based models (Brimelow et al., 2002; Jewell and Brimelow, 2009). It showed a greater skill in providing a more reliable hail size forecast compared with the other sounding methods (Jewell and Brimelow, 2009). Recently, Adams-Selin and Ziegler (2016) coupled a physically improved version of HAILCAST with the Weather Research and Forecasting (WRF) model. Based on simulated cloud liquid and ice water, vertical velocity, temperature, water vapor and pressure fields, WRF-HAILCAST provides the forecast of maximum hailstone diameter at the ground. Adams-Selin and Ziegler (2016) found that WRF-HAILCAST can forecast hail sizes within 12 mm of the observed sizes in 66% of cases. Moreover, WRF-HAILCAST was tested and improved during the 2014–2016 NOAA Hazardous Weather Testbed Spring Forecasting Experiments (Jirak et al., 2014). As a result of testing and improvements, HAILCAST matched the results relatively consistently with the best performing storm surrogate products, such as updraft helicity and column-integrated graupel. Moreover, unlike analyzed storm-surrogate fields, HAILCAST produced consistent results even when used across different model configurations and horizontal grid spacings (Adams-Selin et al., 2019). Recently, Trefalt et al. (2018) tested WRF-HAILCAST for one particular thunderstorm case (6 June 2015) over the Alpine region and obtained a reasonable match of the model output with observed hail (based on radar products and reported damage). Further, Manzato et al. (2020) successfully simulated a severe hailstorm occurring over the northeastern Italy and compared the results with hailpad measurements. Authors found that WRF-HAILCAST was able to reproduce the severity of the hailstorm and the location of the areas affected with the largest hailstones. However, a strong variability of the results depending on large-scale forcing and initialization time is reported. On the other hand, the variability of results when considering different microphysics parameterization schemes is found to be smaller.

Another storm-related phenomenon, i.e., lightning, is one of the indicators of severe weather and extremely damaging phenomena. Explicit modelling of electric fields in the atmosphere, as well as lightning discharges within an NWP model, is possible (Altaratz, 2005; Barthe et al., 2012; Barthe and Pinty, 2007; Fierro et al., 2013; Helsdon et al., 1992; MacGorman et al., 2001; Mansell, 2005; Mansell et al., 2010; Pinty and Barthe, 2008; Tsenova and Mitzeva, 2009), but still, it is rather highly time- and resource-consuming. Therefore, indirect methods are often applied. These methods are based on the relationship between lightning flashes and various nonelectrical parameters/variables that characterize convective activity, such as convective available potential energy (CAPE) and precipitation rate (Romps et al., 2014), convective cloud top heights, convective precipitation and upward convective mass flux (Allen and Pickering, 2002) and upward cloud ice flux at 440 hPa (Finney et al., 2014). Recently (Lopez, 2016; Lopez, 2018) developed a new lightning parameterization for the global convection-parameterizing ECMWF IFS model. This parameterization considers CAPE and diagnosed vertical profiles of graupel and supercooled liquid from convection parameterization. It is shown that there is a good agreement between results of this lightning parameterization in deterministic forecasts on temporal and spatial scales above 6 h and 50 km. However, since this parametrization considers CAPE, it is challenging to use the parametrization in convection permitting models due to CAPE-removal by explicit convection.

Another approach is based on use of the lightning potential index (LPI) developed by Yair et al. (2010), which represents a measure for charge generation and separation inside a thundercloud. The LPI (Jkg^{-1}) is defined as an updraft kinetic energy scaled by the potential for charge separation based on the ratios between ice and liquid water content in the main charging zone of a thundercloud. Several studies have confirmed that LPI is a suitable tool for implicit forecasting of lightning in WRF (Lagasio et al., 2017; Lynn and Yair, 2008; Yair et al., 2010) and COSMO (Sokol and Minářová, 2020) NWP models.

Within this context, this study aims to investigate the predictive ability of the convection-permitting WRF model to reproduce atmospheric conditions and WRF-HAILCAST and LPI to reproduce hail and lightning tracks observed during three hail events in Croatia. Testing of the hail and LPI forecasts contributes to the improvement of timely forecasts and warning systems both for the public and specific users (e.g., the energy sector) and can further lower economic losses. An ensemble of twelve WRF simulations is formed, with a model setup varying among three planetary boundary layer (PBL) and four microphysics schemes. The results of the ensemble of sensitivity simulations are evaluated against the measured lightning flashes (Betz et al., 2009; Franc et al., 2016; Jelić et al., 2021; Jurković et al., 2015; Počakal et al., 2018) and direct hail measurements from the Croatian hailpad networks (e.g., Jelić et al., 2020; Počakal et al., 2009).

The paper is organized as follows. In Section 2, a brief description of selected hail events and observational data used to evaluate WRF-HAILCAST and LPI predictive ability is presented. Section 3 describes the modelling setup, WRF sensitivity experiments and proposed verification approach. The results are presented and discussed in Section 4. Some concluding remarks are given in Section 5.

2. Observational data and hail events

2.1. Hail events

For the simulations, three hail events over Croatia are selected based on their severity and data availability. The first chosen case occurred on 25 June 2017, representing one of the most intense hailstorms in our database. On that day, central Europe was under the influence of an upper-level trough. As the upper-level trough passed eastward, a shallow cyclone developed in the northern Adriatic. Several mesoscale convective systems formed over Croatia during this event, resulting in widespread hail occurrence across the Croatian territory. On that day, 72 hailpads recorded hail, with a maximum measured hailstone diameter of 31.4 mm. The hail event was accompanied by southwesterly (SW) large-scale flow, which is the most common upper-level flow type associated with hail/lightning occurrences in Croatia (Jelić et al., 2020; Mikuš et al., 2012). Moreover, during this event, the maximum total daily precipitation of 34.9 mm with the maximum precipitation rate of 1.82 mm/min was recorded at the Zagreb Maksimir station.

The second case occurred on 24 July 2017, when central Europe was under the influence of a cyclone resulting in SW flow over Croatia. As the cyclone moved to the south, a cold front passed over Croatia, resulting in the formation of hailstorms. Nine hailpads measured hail, and the maximum measured hailstone diameter was 54.2 mm. Moreover, during this event, a maximum total daily precipitation of 34.8 mm is recorded at Križevci station with the maximum precipitation rate of 1.43 mm/min.

A similar situation developed during the third case on 16 September 2017. An intense upper-level trough developed over the European area with the center in southern Europe. While the upper-level trough passed the Alps, a shallow cyclone developed over Genoa. This cyclone moved eastward across the Italian Peninsula and along the Adriatic coast. This hail event was accompanied by upper-level SW flow as well. Hail was reported in the evening and night time hours, from 17:00 to 00:00 UTC, across central Croatia. During the event, 22 hailpads were impacted, and the maximum measured hailstone diameter was 23.1 mm while the maximum recorded daily precipitation was 77.7 mm at Puntijarka station with 11.45 mm/min maximum precipitation rate.

2.2. Observational data

The first step in this analysis was to estimate the model's ability to reproduce general surface conditions on the selected hail days using standard meteorological measurements from automatic stations maintained by the Croatian Meteorological and Hydrological Service (DHMZ). We used hourly values of temperature, relative humidity and hourly maximum wind speed from the stations across Croatia (Fig. 1). Moreover, considering the importance of equivalent potential temperature (θ_e) to pseudo-adiabatic lifting processes and storm formation, simulated and observed surface values of θ_e are also evaluated.

To assess the ability of LPI to reproduce general lightning patterns during the selected hail events, we used lightning data from the Lightning Detection Network (LINET) (Betz et al., 2009). The LINET network detects both cloud-to-ground (CG) and intracloud (IC) lightning flashes and differentiates between positive and negative discharges. With 90 sensors in 17 countries that are up to 250 km away, the LINET network successfully detects weaker stroke signals with a current amplitude lower than 10 kA. From 2009, the median values of detected current amplitude values had decreased by half (Franc et al., 2016), showing significant improvement in the sensitivity detection towards 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. The total lightning for the examined cases was taken from the 2D database of lightning flashes at a 3 km × 3 km horizontal resolution and 2 min intervals (developed by Jelić et al., 2021). Higher spatial and temporal resolution is possible, but it exceeds our computational and storage capacities.

The third observational dataset included hailpad measurements from (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. 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, 2011; Počakal et al., 2009). Moreover, 65 hailpads were installed in Istria (NE Adriatic) in the vicinity of the vineyards during the VITCLIC project (https://www.pmf.unizg.hr/geof/en/research/climatology/vitclic). It is important to note that 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. Positions of hailpads are indicated in Fig. 1.

3. Modelling setup and verification approach

3.1. WRF sensitivity experiments

Selected hail events were simulated using an Advanced Research Weather Research and Forecasting (WRF, version 3.8.1) model (Skamarock et al., 2008a) alongside HAILCAST and LPI. The model setup consisted of three one-way nested domains with horizontal grid resolutions of 9 km (204 × 180 grid points), 3 km (328 × 232 grid points) and 1 km (535 × 334 grid points) (Fig. 2). Taking into account the complexity of the terrain, the position of domains represent a compromise among several factors: (i) a coverage of the basic driving forces that affect the appearance of convective activity, (ii) the inflow/outflow wind relationship (here dominant SW synoptic inflow), (iii) important topographic features, (iv) area of interest (i.e., northern-eastern Adriatic and hinterland of Croatia) and (v) sufficiently distanced smaller domain from the relaxation zone in the larger domain. Similar domain setup has already been studied and tested in previous studies focusing on convection, in Horvath et al. (2019), Kehler-Poljak et al. (2017), Orlić et al. (2010), Renko et al. (2018) and Šepić et al. (2009). 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

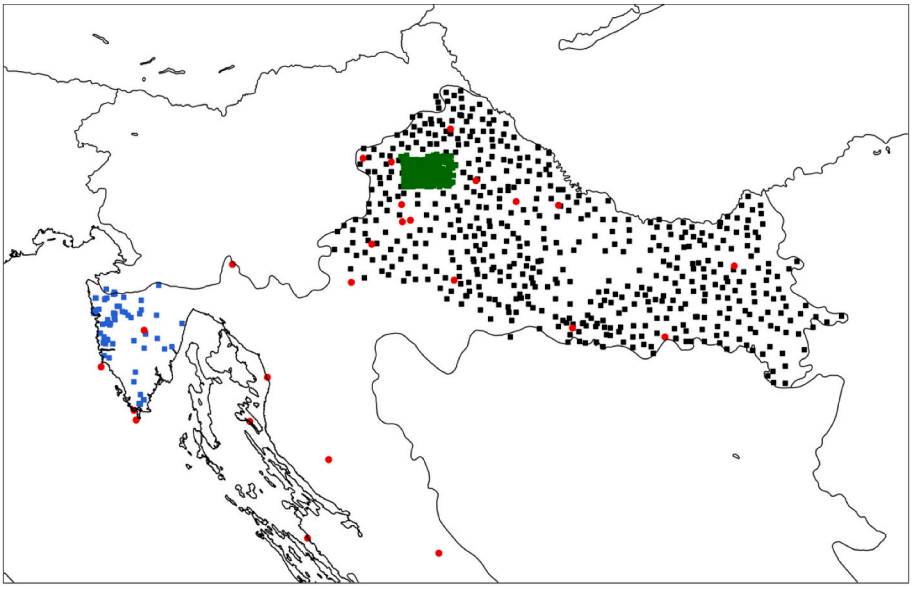

Fig. 1. Positions of automatic weather stations across Croatia used in this analysis (red dots) and hailpads from: (i) hail suppression stations (black squares), (ii) specially designed polygon in the northwestern Croatia (green square), and (iii) stations in Istria (the northeastern Adriatic) not connected to hail suppression activities (blue squares). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

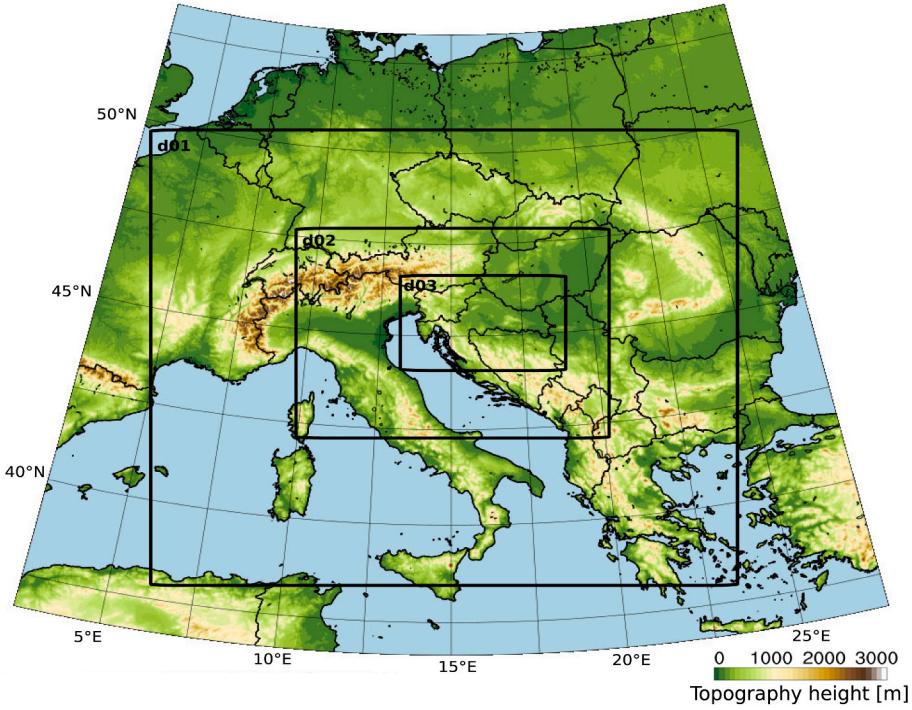

Fig. 2. Position of one-way nested domains at 9, 3, and 1 km grid resolution adopted for WRF simulations (rectangles). Shading indicates topography height as represented in 9 km domain.

sampling (e.g. Manzato, 2008), the model setup consisted of 97 vertical levels with 61 levels in the first 10 km above the surface. The model time step is set to 40 s, 13.33 s and 4.44 s, respectively.

Grid resolutions of 3 and 1 km allowed the model to represent many convective processes explicitly (Kain et al., 2008, 2006), and therefore, on smaller domains, no cumulus parameterization was applied. In the outermost domain, convection was parameterized using the Kain-Fritsch scheme (Kain and Kain, 2004). Other physics options used included 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 PBL parameterization and microphysics (MP) parameterization schemes, we used a multiphysics ensemble approach with combinations of three different PBL and four different MP parameterization schemes.

The chosen MP schemes were the Pardue Lin (Chen and Sun, 2002; Lin et al., 1983) scheme (LIN), WRF single-moment six-class scheme (WSM6) (Hong and Lim, 2006), Morrison double-moment scheme (MORR) (Morrison et al., 2009) and National Severe Storms Laboratory double-moment scheme (NSSL2) (Mansell et al., 2010). They ranged from single-moment (NSSL2 and WSM6) to double-moment ones (LIN and MORR) with the aim of reducing the uncertainty in predicting key microphysical processes responsible for hail and lightning formation. Moreover, we used three different PBL schemes: Yonsei University PBL (YSU) (Hong et al., 2006), MYNN2.5 (MYNN) (Nakanishi and Niino, 2006) and the Bougeault-Lacarrère scheme (BouLac) (Bougeault and Lacarrere, 1989).

The initial and boundary conditions were obtained from the European Center for Medium-Range Weather Forecasts (ECMWF) model analysis (Manzato et al., 2020) with a spatial resolution of 0.125° x 0.125°. All simulations were initialized at 12 UTC the day before the hail event, providing a necessary spin-up time of 12 h (Skamarock et al., 2008b), while the boundary conditions were updated every 6 h.

3.2. WRF - HAILCAST algorithm

WRF-HAILCAST is a time-dependent hail growth model that is integrated into the advanced research version of the WRF (WRF-ARW) model. The original HAILCAST was developed by Poolman (1992) and improved by Brimelow et al. (2002), and it consisted of a steady-state cloud model coupled with a time-dependent hail growth model.

WRF model, when run at a horizontal grid scale of 4 km and smaller, can successfully reproduce the dominant, larger-scale circulations, as well as associated hydrometeor fields (Kain et al., 2006; Weisman et al., 1997). Moreover, WRF can simulate more physically valuable information regarding the intracloud environment than a one-dimensional steady-state cloud model. Therefore, it is useful to couple HAILCAST with WRF as described in Adams-Selin and Ziegler (2016). The new physically improved version of HAILCAST developed by Adams-Selin and Ziegler (2016) is therefore coupled with WRF and it uses the vertical updraft, liquid and ice water content and temperature profiles from a given WRF time step and grid column, to forecast the maximum expected hailstone diameter at the ground. If any grid column has a vertical velocity that exceeds 10 ms^{-1} and such vertical velocities last at least 15 min, the profiles of vertical velocity, temperature, cloud water, cloud ice, and snow-mixing ratio from that grid column are forwarded to WRF-HAILCAST. Then, two embryos of sizes 5 and 7.5 mm are inserted at the level of −8 °C, and three embryos of sizes 5, 7.5 and 10 mm are inserted at the level of −13 °C. These embryos are tracked in a time-varying updraft until the maximum vertical velocity falls below 10 ms^{-1} or the updraft duration exceeds 2000 s. At that point, the hailstones are assumed to have exited the cloud, and the information on the maximum hailstone diameter is passed back to WRF, which stores it and proceeds to its next time step. In our setup HAILCAST is activated only on the smallest domain with 1 km horizontal resolution and HAILCAST activates every model's internal time step, that is every 4.44 s, while accumulated 15-min fields are stored.

Here, hail forecast is obtained using the latest publicly available version of HAILCAST described in Adams-Selin et al. (2019). Initially, HAILCAST is activated if the vertical velocity anywhere in the grid exceeds 10 ms^{-1}. However, the results presented in Adams-Selin et al. (2019) were obtained by simulations at horizontal resolutions of 4 and 3 km. Since we are using 1 km grid spacing and recent studies (e.g., Bryan

and Morrison, 2012; Prein et al., 2021) have pointed out that vertical updrafts are highly resolution-dependent, here, a new threshold of 18 ms^{-1} for HAILCAST activation is examined and introduced. A new threshold is determined by inspecting model's performance in terms of probability of detection and false alarm rate for four activation thresholds (10, 15, 18, 20 ms^{-1}). This choice is further discussed in Section 4.3.1.

3.3. LPI algorithm

Lightning potential index (LPI) developed by Lynn and Yair (2008) and Yair et al. (2010) is formulated as a measure of the potential for charge generation and separation that leads to lightning flashes in a thunderstorm. LPI is defined as a volume integral of the total mass flux of ice and liquid water within the zone between the freezing level and −20 °C isotherm, as this region is where the noninductive ice-graupel mechanism is the most effective due to the presence of supercooled liquid water and ice and graupel particles (Saunders, 2008).

The LPI [J/ kg] is defined as:

$$LPI = \frac{1}{V}\iiint \varepsilon w^2 dxdydz \quad (1)$$

where V [m^{-3}] is a model unit volume, dx, dy, and dz [m] are model grid increments, w [m/ s] vertical velocity and ε [kg kg^{-1}] a dimensionless number dependent on the mixing ratios of graupel, snow, cloud ice and liquid water that obtains values between 0 and 1. A maximum value of 1 is obtained when the mixing ratios of combined ice species and supercooled liquid water are equal, indicating that all species are required in the charge separation process. For more information about ε please refer to Lynn and Yair (2008) and Yair et al. (2010).

In the formulation of the LPI algorithm used in this study, there are several requirements that must be met so that LPI for a particular grid point is nonzero: (i) vertical velocity in that particular grid point must be greater than 1.1 ms^{-1}, (ii) vertical velocity in adjacent grid points (within a five-grid radius) must be greater than 1.1 ms^{-1}, and finally (iii) that particular grid point and its adjacent grid points must be in an unstable environment. Unstable environment is defined by inspection of a parameter similar to mixed layer CAPE obtained with the integration over a 500 hPa layer starting at 50 hPa above ground. A detailed description of these requirements can be found in Brisson et al. (2021). Here, LPI is computed every 2 min and 2 min fields are stored.

3.4. Verification approach

Simulations of selected hail events are verified in three sequential phases: (i) estimation of the model's ability to reproduce surface temperature, wind, relative humidity and equivalent potential temperature measurements from automatic meteorological stations in Croatia, (ii) assessment of the LPI against LINET lightning data using the Structure-Amplitude-Location (SAL) verification method (Wernli et al., 2008), and (iii) verification of HAILCAST results against hailpad observations using a proposed upscaled neighborhood verification method.

First, hourly values of temperature, relative humidity, equivalent potential temperature and hourly maximums of wind speed from automatic meteorological stations (Fig. 1) were used to assess the model's ability to simulate general surface conditions on the selected hail days. Simulated and measured values were compared for all stations considered, and root mean square error (RMSE) decomposition (Murphy, 1988; Takacs, 1985; Taylor, 2001) was performed. We used RMSE decomposition to isolate the influence of dispersion errors on the biases of the mean and standard deviation and consequently to isolate the influence of space and time uncertainties between simulated and observed fields. The practical application of the method using convective-permitting numerical simulations with the WRF model is given in Horvath et al. (2012). Additionally, simulated surface values are compared against ECMWF Integrated Forecasting System (IFS) forecast. Here, 3-hourly values of 2 m temperature, 2 m relative humidity and 2 m equivalent potential temperature obtained by the ECMWF IFS forecast initialized at 12 UTC the day before hail event were used.

Second, LPI is assessed against lightning observations using the SAL method (Wernli et al., 2008). Here, a formulation of SAL method from R's SpatialVx package is used (https://rdrr.io/cran/SpatialVx/man/SpatialVx-package.html). Although primarily designed to evaluate simulated precipitation fields, the object-based SAL method is used for the first time to assess the LPI algorithm's ability to reproduce the main characteristics of observed lightning activity in simulated hail events. The SAL method consists of three independent components: amplitude (A), structure (S) and location (L). These components are obtained considering simulated and observed fields over the entire domain. Amplitude component A is the only component that does not require identification of objects and considers normalized difference between the simulated and observed domain-average numbers of lightning flashes. The values of A are between −2 and 2, while 0 denotes a perfect forecast in amplitude. Positive values indicate that simulated fields overestimate the observed number of flashes and vice versa. Location, L, and structure, S, components require identification, but not matching, of objects in simulated and observed fields. Here, objects are identified after smoothing the field using convolution smoothing with a disk kernel of radius equal to 4 grid points as proposed by Davis et al. (2006). A threshold (inclusive) for defining objects is set to one flash while the minimal size of the object is set to 10 grid points. Location component, L, considers displacement of the simulated and observed field's center of mass and weighted-average distance of the simulated and observed objects from the total field's center of mass. This component can reach values between 0 and 2, and 0 indicates a perfect forecast. Lastly, structure component, S, provides information on whether the simulated objects are too small, too peaked or too large, and/or too flat. Similar to A, S values are also between −2 and 2, while 0 denotes a perfect forecast.

Finally, HAILCAST results are assessed against hailpad measurements. Unlike HAILCAST, hailpad network provides us with spatially limited information on hail occurrence since hailpads can record hail just at the point where they are installed. Potentially, hail could occur in the vicinity of a hailpad or anywhere between two hailpads and be left unrecorded. Unfortunately, Croatia is not yet fully covered with radars, and thus, spatially continuous information on hail occurrence derived from remote sensing fields, such as radars, was not available for this analysis. Moreover, interpolation of hail measurements from the Croatian hailpad network was not performed, as the optimal distance between neighboring hailpads for well-diagnosed interpolated fields should be in the range of 3–3.5 km (Dalezios et al., 2002), and the average distance between the majority of hailpads in Croatia is ~5.5 km. Thus, when evaluating HAILCAST results against hailpad measurements, we had to evaluate spatially continuous fields against point observations. We also had to consider that hail is a small-scale feature, and it is difficult for a high-resolution forecast to match precisely, in space, time, or intensity, observed small-scale features. However, even forecasts that do not precisely match observations in space, time, or even intensity can still be useful. Consequently, in our verification approach, we define a useful forecast as the one where hail is simulated within an allowable displacement radius of 11 km from the point where it is observed. In this way, we do not require the forecast to exactly match the observations, and we still give credit to close forecasts. The choice of a verification window of 11 km is further explained in Section 4.3.2 and Supplement (Fig. S5, S6).

To overcome challenges associated with the limited spatial information from point measurements and to limit the effect of double penalty that occurs when verifying slightly offset high-resolution forecasts of extremely rare events (Ebert, 2008), our proposed methodology to assess spatially continuous hail forecasts against point measurements of this extreme binary event considers an upscaled neighborhood verification method. Unlike the point-to-point verification method,

neighborhood verification methods do not require forecasts to precisely match observations at fine scales. Instead, neighborhood verification methods require that forecasts are in approximate agreement with observations in space, time, intensity or other important aspects (Ebert, 2008). The proposed verification methodology is composed of the elements of already existing verification methods such as point to point, upscaling and a minimum coverage verification method (as described in Ebert (2008)). Although, elements of already existing methods are used, to authors knowledge, this is the first time that this specific formulation of discussed methods is employed. The proposed verification method is schematically illustrated in Fig. 3 and can be described as follows.

Fig. 3a illustrates the starting point of a verification process – we have a spatially continuous hail forecast and nonevenly distributed impacted and unimpacted hailpads. The first step is to use point-to-point verification method to match point observations with the forecast from the closest model grid point (Fig. 3b). This leaves us with the contingency table indicated in Fig. 3c. Note that the contingency table is obtained by summarizing the number of hailpads that are considered hits (hailpad is impacted and hail is simulated at the closest model grid point), false alarms (hailpad is not impacted but hail is simulated), misses (hailpad is impacted but hail is not simulated) and correct negatives (hailpad is not impacted and hail is not simulated), while the grid points not covered with hailpads are not a part of the contingency table, as they lack information on hail occurrence. Then, we move to the second verification step (Fig. 3d), where a correction of point-to-point verification method is performed using the neighborhood verification method with an allowable displacement radius of 11 km. Here, we are only examining impacted hailpads and checking if hail is simulated inside a verification window. If it is, we consider it a hit, and if it is not, we consider it a miss. Therefore, we are decreasing the number of misses and increasing the number of hits compared to the results of a point-to-point method used in the previous step. The neighborhood of only impacted hailpads is scanned, as we cannot be certain if hail did or did not occur in the vicinity of an unimpacted hailpad. In this way, we are allowing the model to offset the observed hail slightly, but we are not penalizing it for simulating hail in the vicinity of an unimpacted hailpad because we cannot be certain that hail truly did not occur somewhere near the unimpacted hailpad. Note that Fig. 3d shows squared verification window rather than circular verification window that is used in our verification approach, as it is easier to formulate schematic overview of verification on rectangular grid. The next step (Fig. 3e) considers filtering false alarms in the area within 11 km of the hailpad that reports hail. Every station where hail is simulated and not reported and is within 11 km from station that did report hail, is ignored in the contingency table. Therefore, in this step, we are decreasing the number of false alarms around the area that reported hail. Hence, by completing this step we do not penalize the model for simulating hail where none was observed, as long as neighborhood hailpads reported hail. Therefore, we are allowing the model to simulate hail anywhere near (up to 11 km) the hailpad that reported hail and thereby we are representing the uncertainty in observations and considering the inability of the high-resolution models to exactly reproduce such small-scale phenomena. Finally, the last step (Fig. 3f) considers constructing a contingency table and computing categorical skill scores, such as probability of detection (POD), false alarm ratio (FAR) and extremal dependence index (EDI) (Ferro and Stephenson, 2011).

4. Results and discussion

4.1. Evaluation of surface conditions

The first phase of evaluation is a comparison between simulated and observed hourly values of 2-m temperature (T2), 2-m relative humidity (RH2), 2-m equivalent potential temperature (Θ_e2) and hourly maximums of 10-m wind speed (V10) at automatic meteorological stations across Croatia. The comparison is made for a period between 00 UTC and 24 UTC on the day hail was observed and for the grid point closest to the meteorological station's position. RMSE is determined and decomposed into three components following Murphy (1988) and Takacs (1985): (i) bias of the mean (MBIAS), (ii) bias of the standard deviation

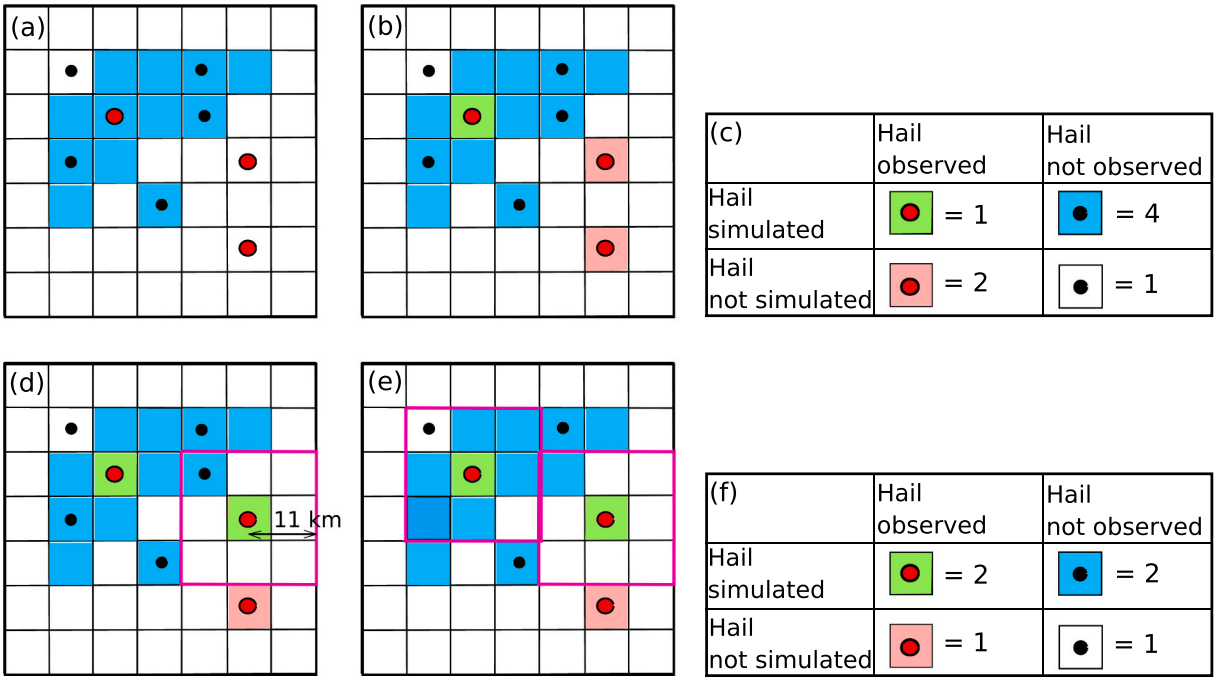

Fig. 3. Schematic description of a verification methodology. (a) Starting point of a verification process. Blue shading indicates grids where hail was simulated. Unimpacted hailpads are represented with black dots, and impacted hailpads are represented with red dots. (b) Point-to-point verification method. Hits are shaded in green, and misses are shaded in red. (c) Contingency table as the result of a point-to-point verification method. (d) Neighborhood verification method. Verification window is indicated with a magenta square. (e) Upscaling of hail observations. Verification windows are represented with magenta squares. (f) Final contingency table. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(SBIAS), and (iii) dispersion or phase error (DISP) in line with the following relation:

$$\frac{1}{MN}\sum_{k=1}^{M}\sum_{i=1}^{N}\left(x_{i,k}-y_{i,k}\right)^2=\frac{1}{M}\sum_{k=1}^{M}\left(\overline{x_k}-\overline{y_k}\right)^2+\left[\sigma_k(x)-\sigma_k(y)\right]^2+2\sigma_k(x)\sigma_k(y)\left[1-r_k(x,y)\right] \quad (2)$$

where x and y are modelled and measured data, respectively; k and i are indices in space and time, respectively; M is the number of stations; N is the number of time steps; σ is the standard deviation; r is the coefficient of correlation between modelled and measured data; and bars denote time means. The left-hand side represents the square of RMSE (mean square error), while the three terms on the right-hand side of Eq. (2) are the squares of MBIAS, SBIAS and DISP.

RMSE components averaged over all stations in Fig. 4 show similar results across all ensemble members and all cases considered. RMSE does not exceed 3.32 °C, 14.3%, 4.94 m/s and 6.36 K for T2, RH2, V10, and Θ_e2, respectively. Moreover, for 25 June 2017 and 24 July 2017, the greatest contribution to RMSE is from phase errors, indicating that there are temporal and/or spatial shifts present between simulated and observed fields. Although RMSE is generally smaller for the 16 September 2017 case, there is a greater contribution of biases of the mean to RMSE compared with the other cases. A positive MBIAS of T2 and a negative MBIAS of RH2 suggest that the model generally overestimated the observed temperature and, consequently, underestimated the observed relative humidity. Interestingly, MBIAS of Θ_e2 shows that equivalent potential temperature is underestimated in 25 June 2017 and 24 July 2017 cases, but overestimated in 16 September 2017 case.

Further, WRF's performance in simulating T2, RH2 and Θ_e2 is compared to that of ECMWF Integrated Forecasting System (IFS) (Fig. 4). Interestingly, WRF seems to perform better compared to ECMWF IFS in terms of simulating T2 in all three cases considered (except the NSSL2 members for 25 June 2017 case), but similar in

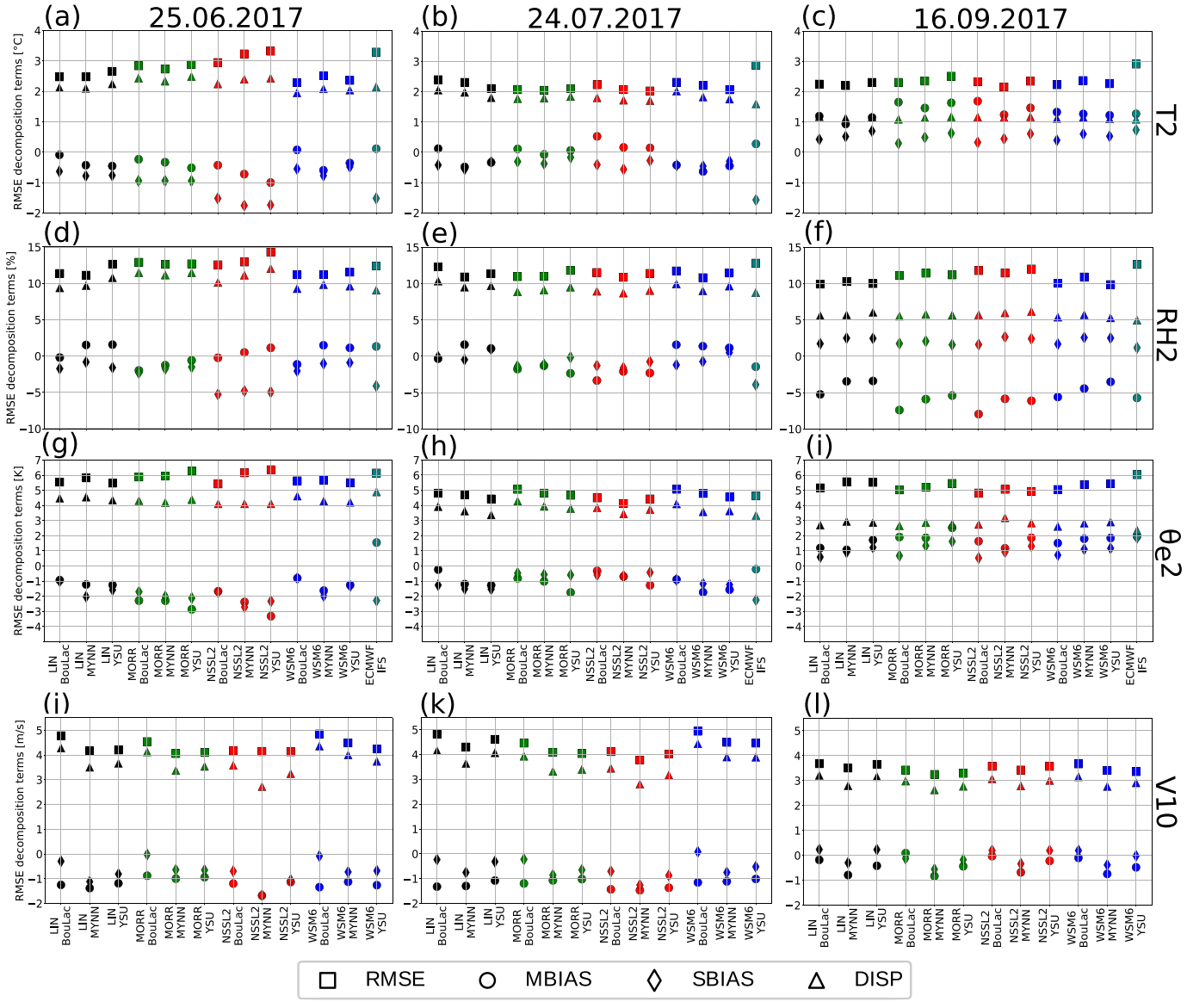

Fig. 4. Performance diagrams indicating RMSE decomposition terms across ensemble members. Each member is denoted on the x axis with the acronym indicating a combination of microphysics and PBL parameterization scheme. RMSE denotes total RMSE, MBIAS bias of the mean, SBIAS bias of the standard deviation and DISP dispersion or phase error. Performance for 2-m temperature (a), (b), and (c); 2-m relative humidity (d), (e), and (f); 2-m equivalent potential temperature (g), (h), and (i); 10-m wind speed(j), (k), and (l); for the three hail events analyzed (columns) is given. Every symbol represents the average values of RMSE, MBIAS, SBIAS and DISP across all stations considered. Colors represent members with the same microphysics scheme. Teal markers represent the values associated with ECMWF IFS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

simulating observed values of RH2 for 25 June 2017 and 24 July 2017 or slightly better in 16 September 2017 case. Overall, both models show similar performance in simulating observed $\Theta_e 2$ values although ECMWF IFS shows positive MBIAS for the 25 June 2017 case opposed to the negative MBIAS produced by all WRF ensemble members. Moreover, note that WRF's performance is evaluated using 1-hourly fields, while ECMWF IFS's performance is evaluated using 3-hourly fields as 1-hourly outputs are not available for ECMWF IFS model. However, similar remarks are obtained when evaluating 3-hourly fields from WRF and 3-hourly fields from ECMWF IFS (not shown).

Overall, general atmospheric conditions are well represented by all ensemble members in all cases considered, and they are in the range of the errors obtained in other numerical studies (e.g., Kehler-Poljak et al., 2017; Milovac et al., 2016; Poljak et al., 2014). It seems that different combinations of parameterization schemes did not show a systematic tendency to overestimate or underestimate surface observations (considering all cases together), which can be partially attributed to the averaging of deviations within the domain. Locally, the deviations at a particular station may be greater. Considering the importance of evaluating surface observations at stations affected with severe weather as

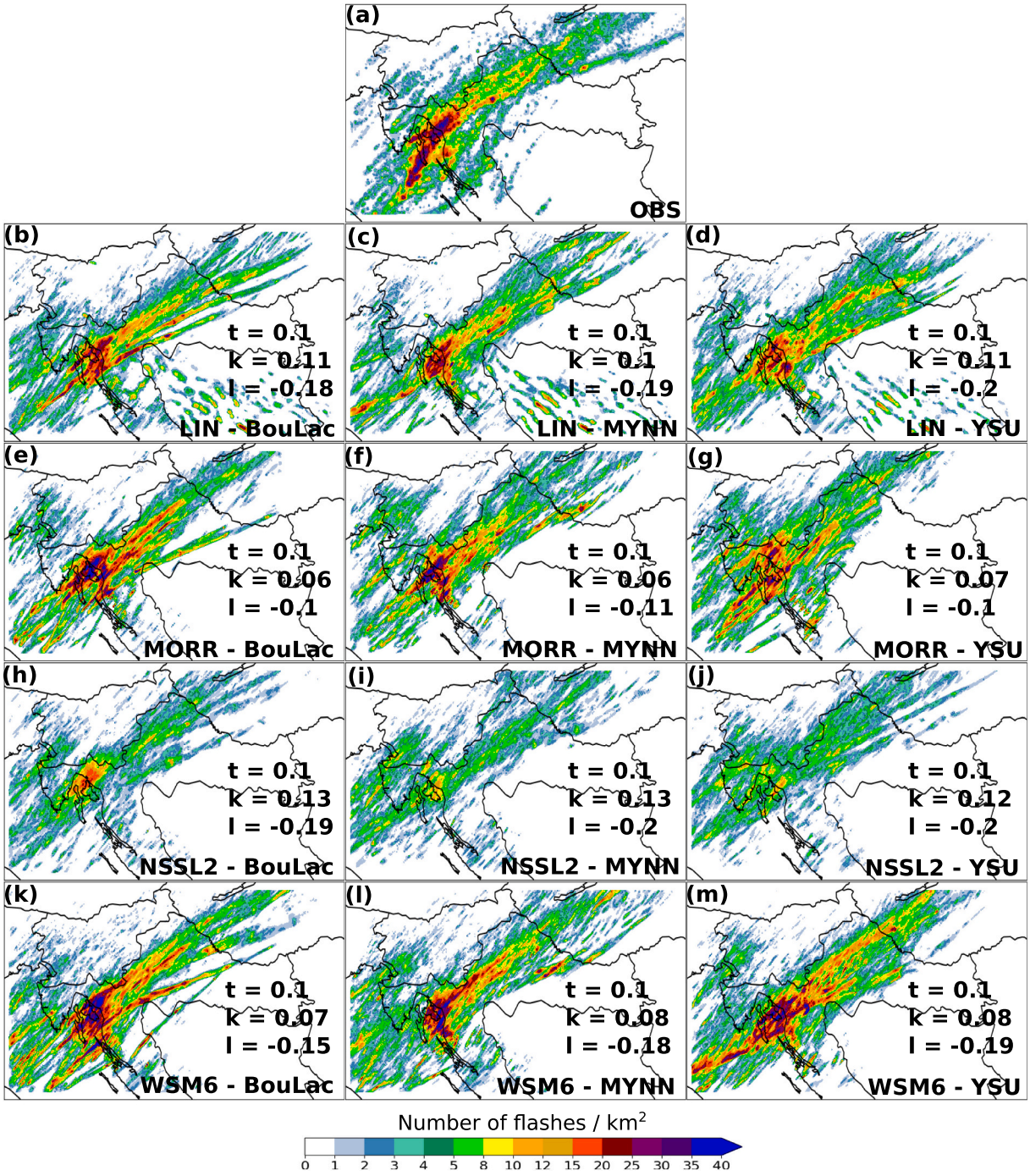

Fig. 5. Comparison of measured and simulated lightning flash accumulation in the time window from 00 UTC to 24 UTC on 16 September 2017 for (a) LINET network, (b) LIN-BouLac, (c) LIN-MYNN, (d) LIN-YSU, (e) MORR-BouLac, (f) MORR-MYNN, (g) MORR-YSU, (h) NSSL2-BouLac, (i) NSSL2-MYNN, (j) NSSL2-YSU, (k) WSM6-BouLac, (l) WSM6-MYNN, and (m) WSM6-YSU. As a reference, the values of *t*, *k*, *l* that optimize each member are indicated above ensemble member acronym. However, note that the presented fields are obtained using median *t*, *k*, *l* values for all ensemble members and all cases considered.

opposed to averaging the results in the whole domain, two representative stations for each hail event analyzed are chosen, and for those stations, the measured and simulated time series of surface conditions are inspected. The results are shown in Supplement (Fig. S1, S2 and S3). Overall, the majority of ensemble members captured the time evolution of observed surface values well, although some overestimation or underestimation may be present, depending on the case analyzed. The worst performance is obtained for maximum wind speeds as model tends to shift the observed maximums in time (and possibly in space).

4.2. Lightning Potential Index results

Lightning potential index [J/kg] highlights the areas where the potential for electrical activity exists, and as such, it is not directly connected to the observed number of lightning flashes. Therefore, to quantitatively compare these two fields, the conversion of LPI to the number of lightning flashes needs to be performed. Here, we convert LPI to the number of lightning flashes, similar to Brisson et al. (2021), by first defining a minimum value of LPI, t, for which a lightning flash was produced. Second, a linear relationship between LPI and the number of lightning flashes is assumed, such that:

$$LPI_{adj} = \begin{cases} 0, when\ LPI \leq t \\ k \bullet LPI + l, when\ LPI > t \end{cases}$$

where LPI_{adj} [$km^{-2}\ h^{-1}$] denotes adjusted LPI, i.e., LPI converted to the number of lightning flashes. Furthermore, the parameters t, k and l are sampled across [0.001, 10], [0, 5], [−20,20] domain, respectively, and for every combination of these parameters, hourly means of LPI_{adj} are calculated. Then, a distribution function associated with the hourly means of LPI_{adj} is compared with the distribution function associated with the observed hourly means of lightning flashes. The values of parameters k, l, and t, for which the RMSE between the two discussed distribution functions is minimal, are considered optimal. For every model configuration and every hail case considered, a unique set of parameters, k, l, and t, minimize RMSE between the discussed distribution functions, leading to a set of 36 values for each parameter. In the subsequent analysis, we choose the median values of obtained sets of parameters to investigate potential discrepancies across different model configurations and different hail cases. The medians and standard deviations are $t = 0.1 \pm 0.09$, $k = 0.11 \pm 0.04$ and $l = -0.3 \pm 0.35$.

Daily sums of both LPI_{adj} (Fig. 5b-m) and the observed number of total lightning flashes from LINET network (Fig. 5a) gridded on the model grid for one of the analyzed hail cases (16 September 2017) are presented. Overall, it seems that all ensemble members capture the general spatial pattern of observed lightning activity well, although there are some local discrepancies between simulated and observed lightning activity. However, the ensemble members with NSSL2 microphysics scheme seem to underestimate the intensity of the observed lightning activity. On the other hand, the most intense lightning activity seems to be produced by the members with the WSM6 and MORR microphysics scheme. Additionally, Fig. 5b–m shows the values of t, k and l that optimize each ensemble member separately even though the presented fields are obtained by considering median values of t, k and l. Note that the parameters t, k and l that optimize each ensemble member separately do not change considerably between members (this is also true in the two cases not shown) indicating that standard deviations associated with these parameters are coming from differences among cases, rather than differences among ensemble members. This

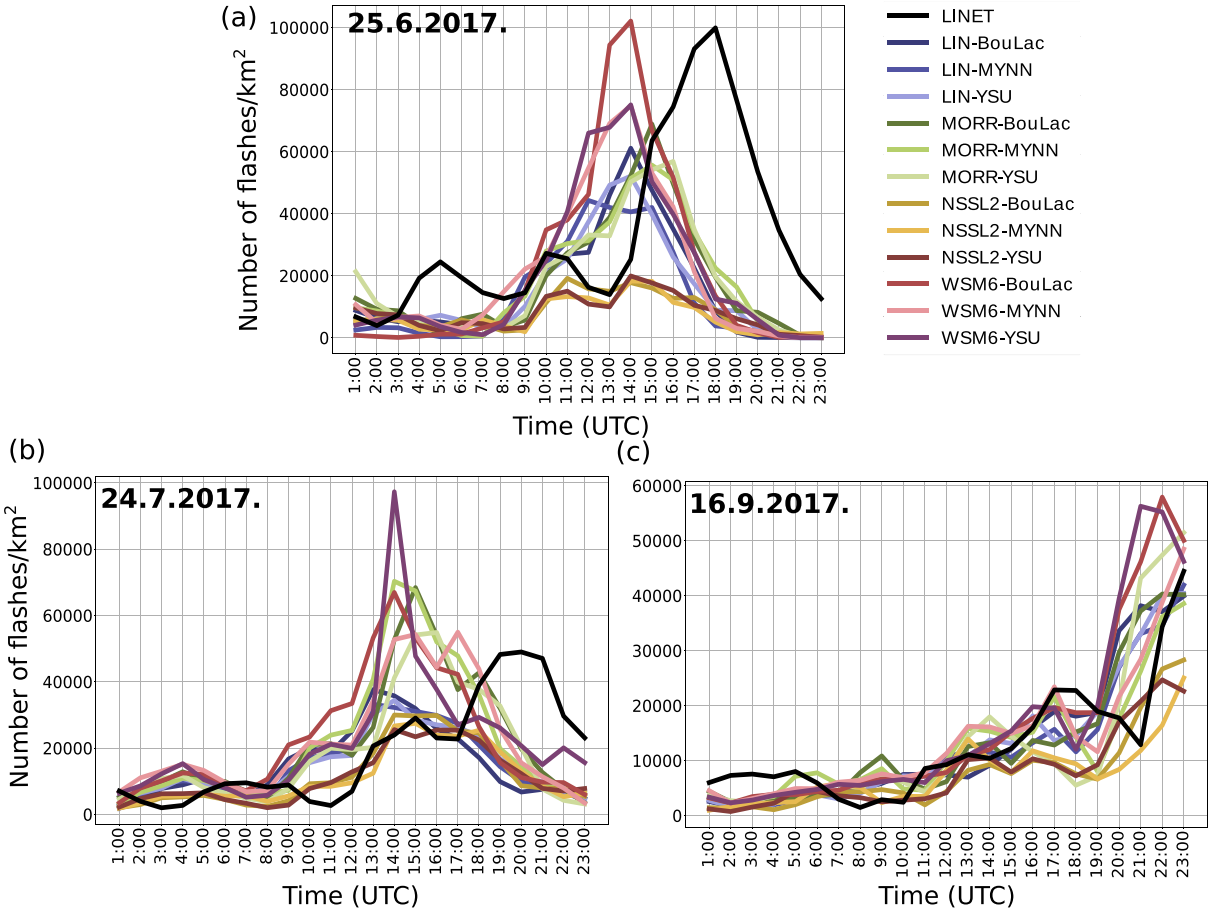

Fig. 6. Hourly sums of the number of lightning flashes across the entire domain (d03 from Fig. 2) for the three hail events analyzed: (a) 25 June 2017, (b) 24 July 2017, and (c) 16 September 2017. Both observed (black) and simulated lightning activities for each ensemble member (colors in legend) are shown.

highlights the fact that more cases would be needed to obtain a set of parameters that are optimal in majority of thunderstorm environments.

Looking at temporal evolution of lightning, similar remarks can be found (Fig. 6). Fig. 6 shows the hourly sums of the simulated and observed numbers of lightning flashes over the entire innermost domain shown in Fig. 2. It is clear that the members with NSSL2 microphysics systemically underestimate the observed lightning activity. At the same time, the members with the most intense lightning activity mostly have a WSM6 microphysics scheme.

Moreover, the observed lightning activity on 25 June 2017 is underestimated in all ensemble members except for the WSM6-BouLac combination. On 24 July 2017, the observed number of lightning flashes is overestimated in members with WSM6, as well as the MORR microphysics parameterization scheme. Interestingly, a time shift of up to 6 h, depending on the parameterization considered, between the onset of the maximum observed and simulated lightning activity is present in simulations for both 25 June 2017 and 24 July 2017. Convection occurring too early may occur due to various reasons including lack of data assimilation but also too strong coupling with daily cycle of convection and associated surface fluxes. While this issue has been a common property of convection-parametrized models (e.g. Ban et al., 2015; Manzato et al., 2018), considerable uncertainty of timing of convective precipitation exists also for convection-permitting model simulations (e.g. Bechtold et al., 2014; Clark et al., 2016).

The best matching between the simulated and observed onset of lightning activity and the number of lightning flashes is obtained for the nocturnal convection, hailstorm on 16 September 2017. Nevertheless, a discrepancy between the number of flashes simulated with WSM6 and NSSL2 microphysics schemes is present in this case as well. Since the lightning activity for this case peaks at the end of the simulation time, observed lightning activity for the extended period until 03 UTC the next day is shown in Supplement (Fig. S3).

Furthermore, to obtain a more detailed and quantitative description of ensemble member performance, the SAL verification method is performed. SAL analysis is performed on daily accumulated fields of both observed and simulated lightning activity to mitigate the effects of temporal shifts in the onset of observed and simulated convection (Fig. 7). SAL diagrams show that the ensemble members for 25 June 2017 exhibit lightning objects that are too small or too peaked, as indicated by the S component's negative values. Some of the members slightly overestimate, and others underestimate observed amplitudes. The object locations are relatively well reproduced in most members except those with the NSSL2 microphysics scheme. For the 24 July 2017 case, both structure, S, and amplitude, A, are well reproduced by all ensemble members, except for the WSM6-YSU member, which reproduces overly peaked objects, as indicated by the S component being close to a value of −2. Again, the members with the NSSL2 microphysics scheme exhibit the smallest A, indicating the underestimation of lightning activity. Interestingly, members with WSM6 scheme show large differences in performance for different PBL schemes with MYNN scheme performing the best and YSU scheme preforming the worst in terms of S component. The values of all components for case 16 September 2017 are close to 0, indicating good performance of all ensemble members. Furthermore, the discrepancies between lightning activity simulated with WSM6 and NSSL2 members are highlighted mostly in the A and S components.

Overall, analyzing the performance in all cases together, we can see that majority of ensemble members show similar performance, except those with NSSL2 microphysics and WSM6-YSU combination. The members with NSSL2 microphysics show amplitudes furthers from zero

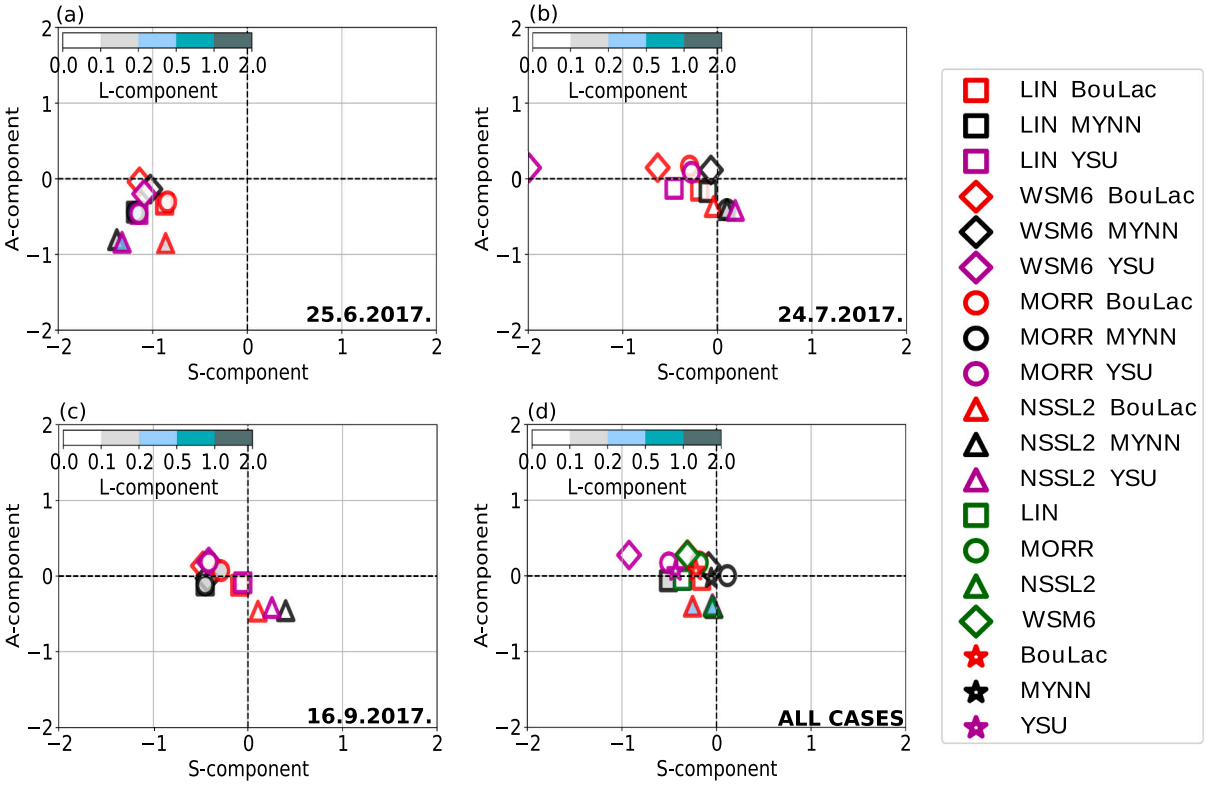

Fig. 7. SAL diagrams for the daily accumulated number of lightning flashes for all ensemble members (marked with different symbols) obtained for: (a) 25 June 2017, (b) 24 July 2017, (c) 16 September 2017, and (d) all cases together. Every symbol shows the values of all three SAL components; amplitude (A) and structure (S) are indicated on the *y*-axis and *x*-axis, respectively, while the value of the location component (L) is indicated by the fill colour of the symbols. The optimal values (all three components equal to 0) are indicated with horizontal and vertical lines. Additionally, green markers and stars show medians of A, S and L in all the cases analyzed for every MP and PBL parameterization scheme selected. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the SAL diagrams, indicating that these members did not reproduce the intensity of the observed lightning activity. Consequently, for our case studies this microphysics parameterization scheme seems to be the least appropriate choice of all schemes considered. On the other hand, looking at the performance of all cases together, WSM6-YSU member produces S component close to −1 indicating that this combination produces overly peaked objects. All the other considered configurations, except those with NSSL2 scheme, show similar performance in the corresponding median values and all of them could be suitable choices for future simulations of lightning events. However, it should be noted that WSM6 scheme slightly overestimates observed amplitudes and produces the most intense lightning activity of all microphysics schemes analyzed. Further, the performance of analyzed PBL schemes is similar, although MYNN seems to perform slightly better compared to BouLac and YSU, while YSU seems to be the worst choice.

Interestingly, single- and double-moment microphysics schemes yield similar performance, and no prevalence of either of the two approaches is apparent. These results are in agreement with the analysis conducted by Lagasio et al. (2017), who analyzed LPI performance in multi-microphysical ensemble simulation of a back-building mesoscale convective system over Genoa, Italy. They found no distinction in performance between single- and double-moment microphysics schemes. Moreover, the greatest LPI values and consequently the most intense lightning activity were produced by the WSM6 member, while the NSSL2 scheme with predicted cloud condensation nuclei (CCN) produced lower LPI values (note that we use NSSL2 scheme with steady background CCN). On the other hand, Sokol and Minářová (2020) compared the impact of single- and double-moment microphysics schemes on LPI results within the COSMO NWP model during 10 thunderstorm days in central Europe and concluded that LPI showed better performance while using double-moment cloud microphysics scheme. However, it should be mentioned that the cloud microphysics used in Sokol and Minářová (2020) is described differently in COSMO NWP than in WRF. Nevertheless, similar to previous studies (Lagasio et al., 2017; Sokol and Minářová, 2020; Yair et al., 2010), LPI was found to be a suitable tool in simulating the observed lightning activity during the selected hail events.

4.3. Hail results

4.3.1. HAILCAST tuning

Firstly, to avoid a considerable number of false alarms obtained by activating HAILCAST in every grid cell where the vertical velocity exceeds 10 ms^{-1}, four different thresholds for activating HAILCAST are examined: 10, 15, 18, and 20 ms^{-1}. For the selected thresholds, a point-to-point comparison between simulated and observed hail is performed, and hit rate ($a/(a + c)$) and false alarm rate ($b/(b + d)$) values are determined for all cases and all ensemble members (Fig. 8), where a denotes for observed-forecasted, b not observed-forecasted, c observed-not forecasted and d not observed-not forecasted event in the

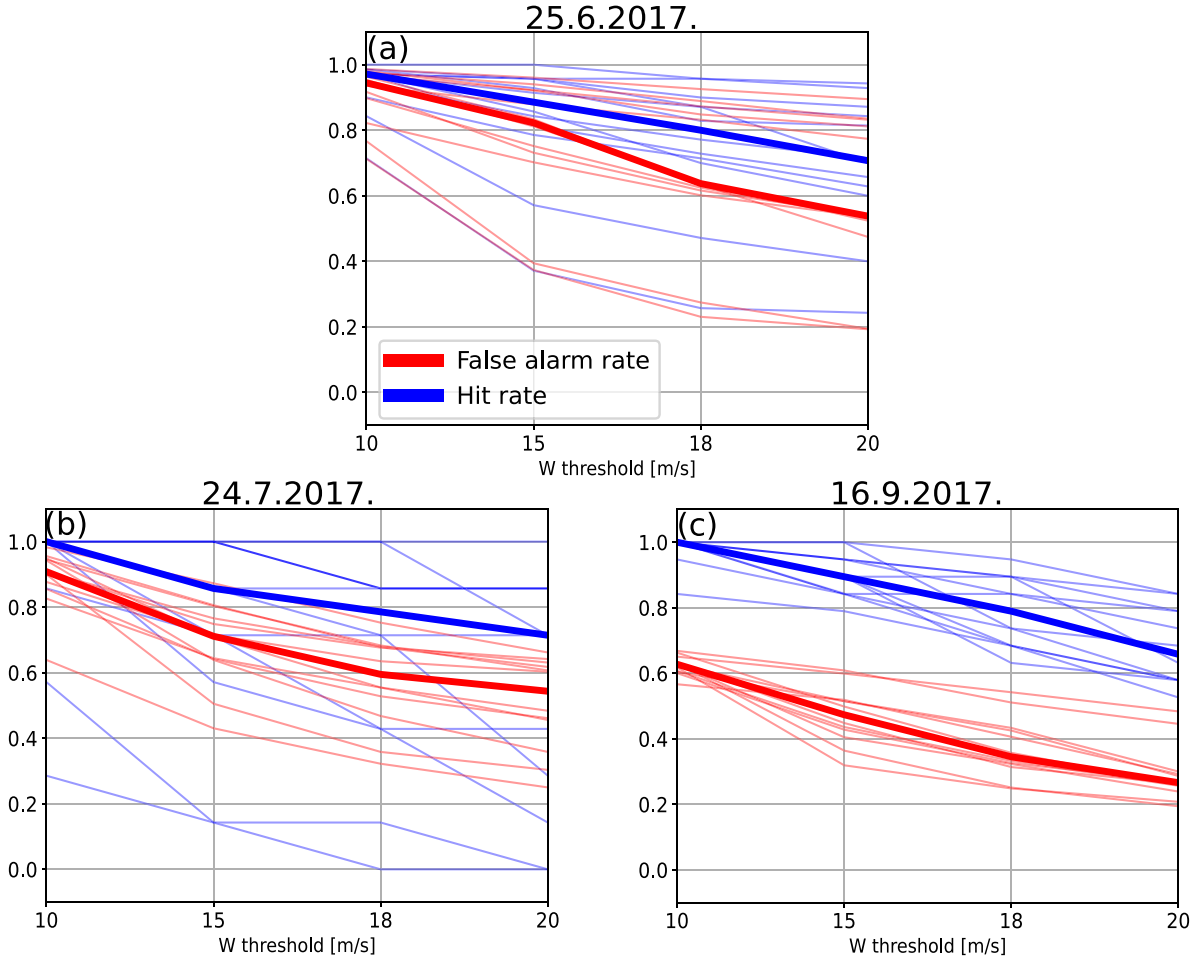

Fig. 8. Performance diagrams depicting hit rate and false alarm rate for different values of HAILCAST activation thresholds for all ensemble members and all cases: (a) 25 June 2017, (b) 24 July 2017, and (c) 16 September 2017. Thresholds of 10, 15, 18 and 20 ms^{-1} are examined. Hit rates for all ensemble members are shown as blue lines, while the thick blue line represents ensemble medians. False alarm rates for all ensemble members are shown as red lines, while the thick red line represents ensemble medians. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contingency table. It is clear that if the threshold for HAILCAST activation was 10 ms^{-1}, HAILCAST would activate in too many grid points, producing a forecast with many false alarms. By choosing a greater activation threshold, false alarm rates are reduced, but so are hit rates. Here, we choose a threshold of 18 ms^{-1}, as median hit rates corresponding to this value are still greater than or equal to 0.8 with a reduced false alarm rate compared with that corresponding to the 10 ms^{-1} threshold. By doing this, we are maximizing the difference

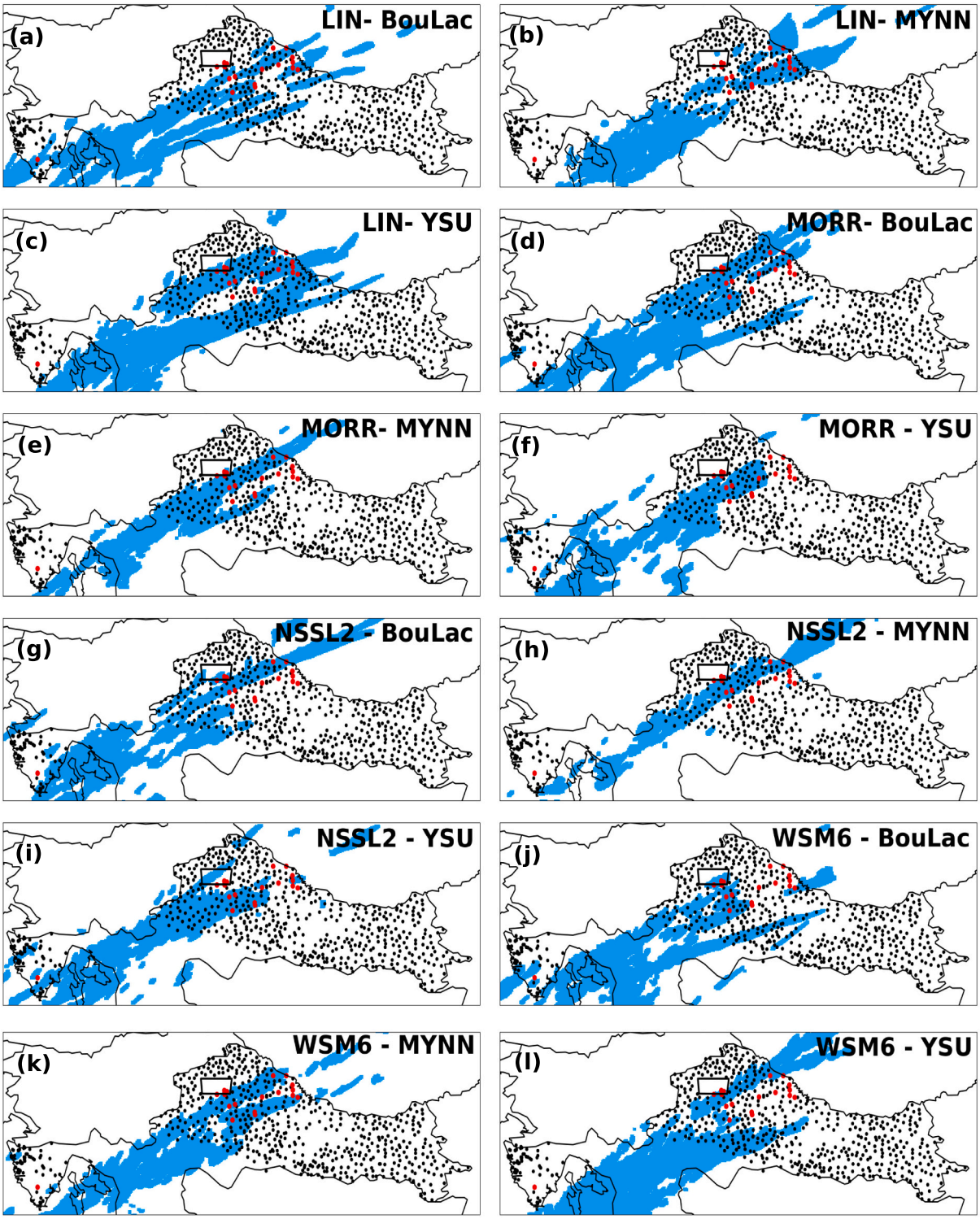

Fig. 9. Hail forecast for 16 September 2017. The area where hail was forecasted from 00 to 24 UTC on 16 September 2017 is shaded in blue. The position of hailpads is indicated with black dots. Hailpads that registered hail are marked with red circles. The position of the specially designed hailpad polygon is marked with a black rectangle. Forecasts across all ensemble members are shown: (a) LIN-BouLac, (b) LIN-MYNN, (c) LIN-YSU, (d) MORR-BouLac, (e) MORR-MYNN, (f) MORR-YSU, (g) NSSL2-BouLac, (h) NSSL2-MYNN, (i) NSSL2-YSU, (j) WSM6-BouLac, (k) WSM6-MYNN, and (l) WSM6-YSU. The best performing member is (k) WSM6-MYNN, and the worst performing member is (l) WSM6-YSU. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between hit rate and false alarm rate and that corresponds to maximizing Peirce Skill Score ($((ad-bc)/((a+c)(b+d)) = POD - FAR$) (Peirce, 1884). Looking at Fig. 8, there is a better performance in terms of Pierce Skill Score for 16 September 2017 case, but the relative change in terms of Pierce Skill Score between the 10 ms^{-1} and 18 ms^{-1} thresholds is greater for 25 June 2017 and 24 July 2017 cases. This indicates that there is a greater benefit in introducing a new threshold for 25 June 2017 and 24 July 2017 cases in terms of the relative change of Pierce Skill Score between 10 ms^{-1} and 18 ms^{-1} thresholds. However, note that this threshold is obtained based on three hail events and four threshold values and might not be generally applicable.

4.3.2. HAILCAST results

Analysis of HAILCAST results is based on the comparison between simulated hail swaths and ground observations for a period from 00 UTC to 24 UTC on the day when hail was observed. To eliminate some of the uncertainty due to temporal variability between observed and simulated convection, we accumulated a simulated dataset for a period of 24 h. First, we perform a qualitative comparison between simulated and observed hail. For this reason, Fig. 9 shows the areas where hail is

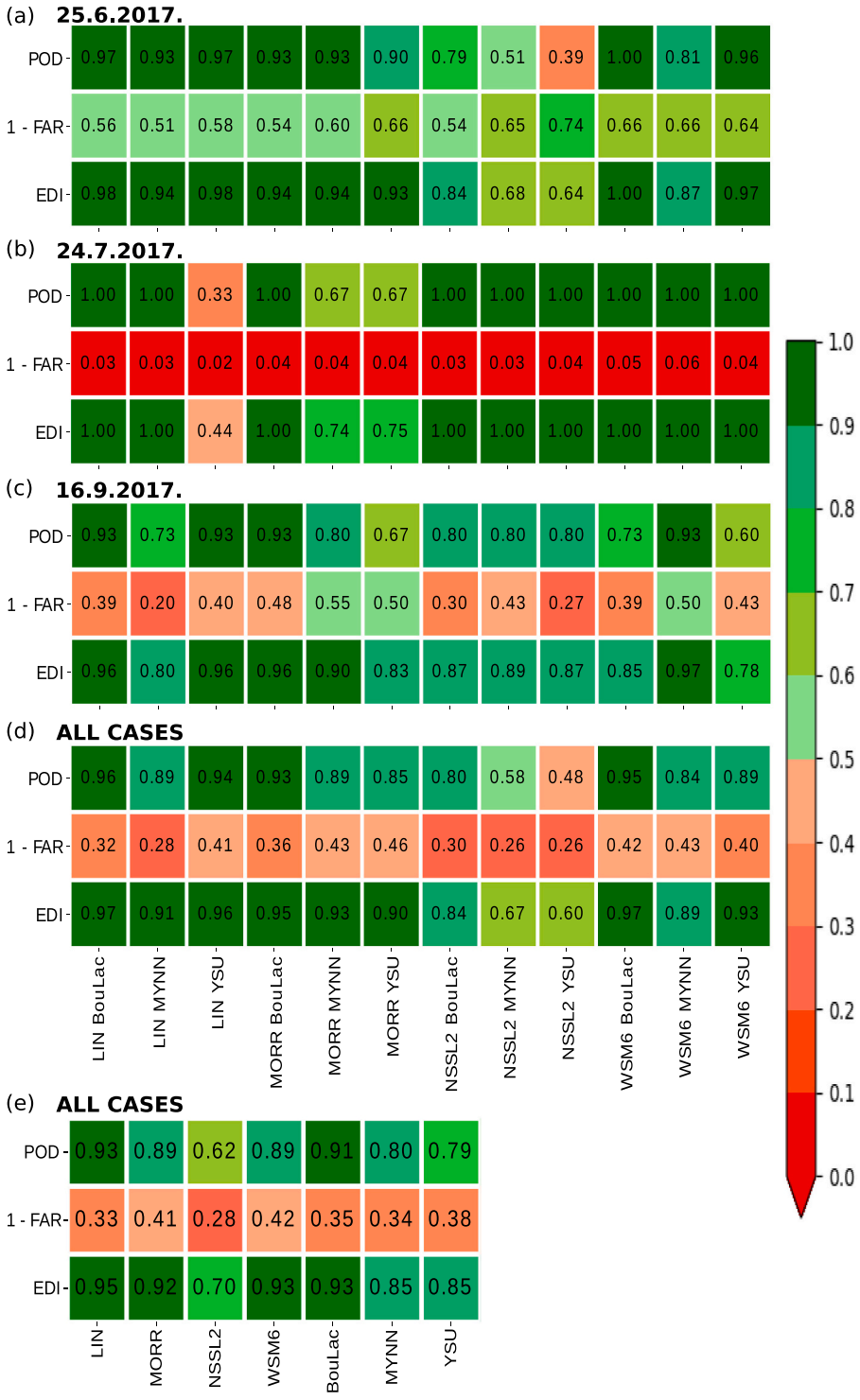

Fig. 10. Performance diagrams across all ensemble members for (a) 25 June 2017, (b) 24 July 2017, and (c) 16 September 2017, (d) all analyzed cases and (e) each parametrization scheme in all cases considered. Values of categorical skill scores probability of detection (POD), 1 - false alarm ratio (1-FAR) and extremal dependence index (EDI) are given. The colour scale denotes the best (green) and the worst (red) performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

forecasted within a 24 h period along with denoted hailpads that recorded hail for the case of 16 September 2017. Fig. 8 suggests a generally good spatial agreement between simulated and observed hail for all ensemble members. Nonetheless, all ensemble members exhibit a certain number of false alarms, i.e., hail is simulated, but the hailpad did not register hail. Furthermore, out of all combinations, the member NSSL2-MYNN shows the smallest area covered with hail.

To obtain a quantitative description of this evaluation, an upscaled neighborhood verification method is performed (as described in Section 3.4) for all ensemble members and all cases considered. The sensitivity of results is tested to the choice of allowable displacement radius. Since the average distance between the majority of hailpads is 5.5 km, three different radiuses are tested – 6, 11 and 22 km. For each of the radiuses POD, FAR, and EDI are calculated (Fig. 10 and Supplement S5, S6). Fig. 10 shows the results for allowable displacement of 11 km as it is determined that forecast has sufficient skill for that allowable displacement distance (majority of EDI values for evaluation of all cases greater than 85%) and the differences between members are better highlighted compared to the results for 22 km radius. Note that Fig. 10 indicates 1-FAR values; therefore, the perfect forecast is associated with all considered skill scores equal to 1.

For the 16 September 2017 case, POD values greater than or equal to 80% for most ensemble members confirm that the model successfully produced hail where it was observed. At the same time, all ensemble members produced a considerable number of false alarms, as indicated by FAR values greater than 45% or 1-FAR values lower than 55%. In addition to the potential tendency of the model to overestimate hail, some degree of false alarms can be attributed to the limited spatial information regarding hail occurrence provided by the hailpad network, i.e., only information about hail occurrence at the exact position of the hailpad was available. Theoretically, hail could have occurred in the vicinity of the hailpad and not have been recorded. Narrowing down towards a ranking of ensemble members, the best performance is obtained with the WSM6-MYNN member with the highest values of POD and EDI and relatively small values of FAR. In terms of FAR, the best performance shows MORR-MYNN member, but with lower POD and EDI. In contrast, the worst performing member seems to be WSM6-YSU according to POD and EDI values while LIN-MYNN member produces the most false alarms.

Furthermore, the majority of POD values associated with the 25 June 2017 hailstorm are greater than 90% (Fig. 10a), suggesting that the observed hail was successfully captured. In this case, there is less false alarms compared to the previous case; FAR values for all members are lower than 49% (or 1-FAR values greater than or equal to 51%). Overall, WSM6-BouLac member yield the best performance in terms of POD and EDI with relatively low FAR, while the NSSL2-YSU member yields the worst performance with the lowest POD and EDI values. The area affected with hail for the best- and worst-performing members is shaded in blue in Fig. 11.

Similar results are obtained for 24 July 2017 (Fig. 10b); the majority of POD values are equal to 100%, suggesting that the model successfully captures the observed hail. However, FAR values above 94% indicate that all ensemble members significantly overestimate the area affected by hail. Furthermore, POD values of 33% and 67% for LIN-YSU, MORR-MYNN and MORR-YSU members indicate that the observed hail was not captured in these configurations. For this particular case, it seems that the worst performing member is LIN-YSU since this member did not simulate observed hail and at the same time produced a considerable number of false alarms. Other members show a similar performance, and there seems to be no one member with the best performance. The area affected with hail for one of the best- and worst-performing members is shaded in blue in Fig. 11.

Additionally, the performance of the ensemble in all cases together is examined (Fig. 10d) to aggregate results and provide a more robust evaluation of each member performance during all analyzed cases. The joint analysis shows that WSM6-BouLac and LIN-YSU might be the MP-PBL combinations that yield the best results in terms of EDI and the difference between POD and FAR, while NSSL2-MYNN and NSSL2-YSU might be the worst MP-PBL choices. Although, it should be noted that the differences in performance across ensemble members are relatively small (if we exclude the members with NSSL2 microphysics scheme). Furthermore, EDI values in the joint evaluation for majority of cases are greater than 0.9 and coincide well with the trends observed in the POD values in both the joint and separate evaluations. Note that in evaluation of events performed separately, EDI values obtain the perfect value (EDI = 1) because POD obtains the perfect value as well (POD = 1) regardless of the relatively large FAR, which is avoided in assessment of the aggregated results. This behavior stems from the formulation of EDI (Ferro and Stephenson, 2011) which suffers from a small number of data used in evaluation. Finally, those remarks highlight the need of using more events for verification in order to achieve more detailed results, especially for metrics related to rare events, as well as the need of examining multiple skill scores when evaluating such extreme and rare events.

Further, a joint evaluation of each parameterization option in all cases together is performed (Fig. 10e). Joint evaluation reveals that out

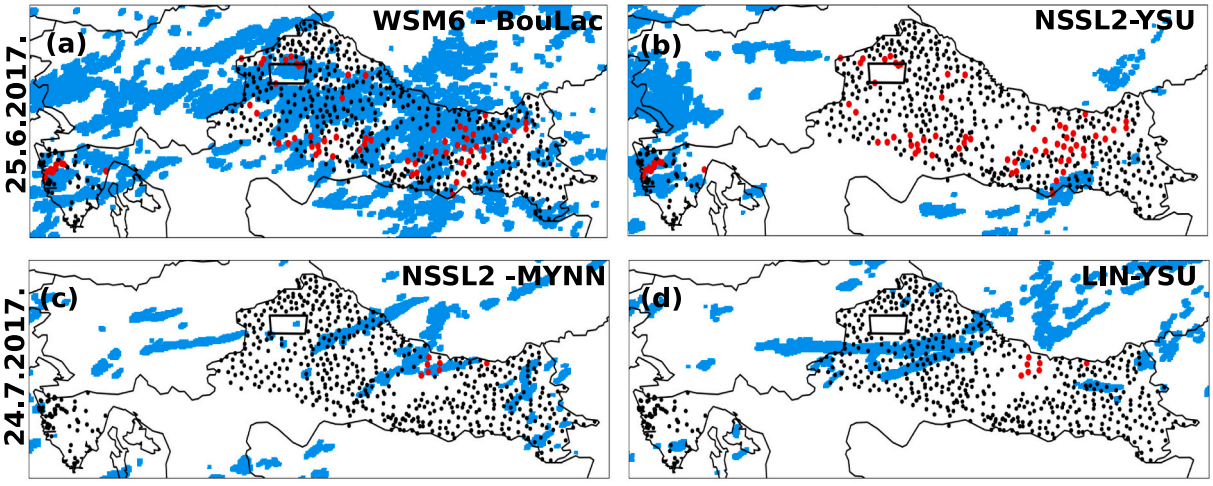

Fig. 11. Hail forecast for 25 June 2017 for the best (a) MORR-YSU and the worst (b) NSSL2-YSU performing members and 24 July 2017 for one of the best (c) NSSL2-MYNN and the worst (d) LIN-YSU performing members. The area where hail was forecasted from 00 to 24 UTC on the day hail was reported is shaded in blue. The position of hailpads is indicated with black dots. Hailpads that registered hail are marked with red circles. The position of the specially designed hailpad polygon is marked with a black rectangle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of all considered microphysics options, LIN seems to perform the best, in terms of POD and EDI values with relatively low FAR values. Not surprisingly, NSSL2 is the worst microphysics option out of all considered in terms of POD and EDI values. Looking at the performance of PBL schemes, BouLac seems to be the best option, while YSU seems to be the worst option in terms of POD and EDI values, although the difference in skill scores between YSU and MYNN are small. Interestingly, even NSSL2 scheme performs better when coupled with BouLac scheme rather than MYNN and YSU schemes (Fig. 10d). Note that YSU scheme performed the worst in LPI evaluation as well (Fig. 7d). This could be due to the fact that YSU is a nonlocal scheme (compared to the BouLac and MYNN that are local schemes) and YSU has been found to over deepen the PBL in springtime deep convective environments (Coniglio, 2012). This then results in too much dry air near the surface which leads to the underestimation of MLCAPE (e.g. Cohen et al., 2015; Milovac et al., 2016). On the other hand, BouLac is a local scheme and it is designed to better represent terrain-enhanced turbulence (Cohen et al., 2015) and given the highly complex terrain in Croatia, a better representation of terrain-enhanced turbulence could lead to better results.

Not surprisingly, analyzed skill scores obtain larger values for larger verification radiuses and smaller values for smaller verification radiuses (Figs. S1.1 and S1.2 in Supplement). When comparing the joint evaluation for all cases considered for verification radius of 6 km, WSM6-BouLac member stands out as the best performing in terms of EDI and POD with relatively small FAR compared to the other members. Interestingly enough, even at 6 km, the members with NSSL2 scheme seem to perform worse than the other members, looking at all the cases together. On the other hand, looking at the performance at 22 km of all cases together, all members obtain their largest EDI, POD and 1-FAR values. Again, members with NSSL2 scheme perform worse compared to the other members. It seems that WRF-HAILCAST forecast could be considered reliable at the scale of 22 km, although one should consider

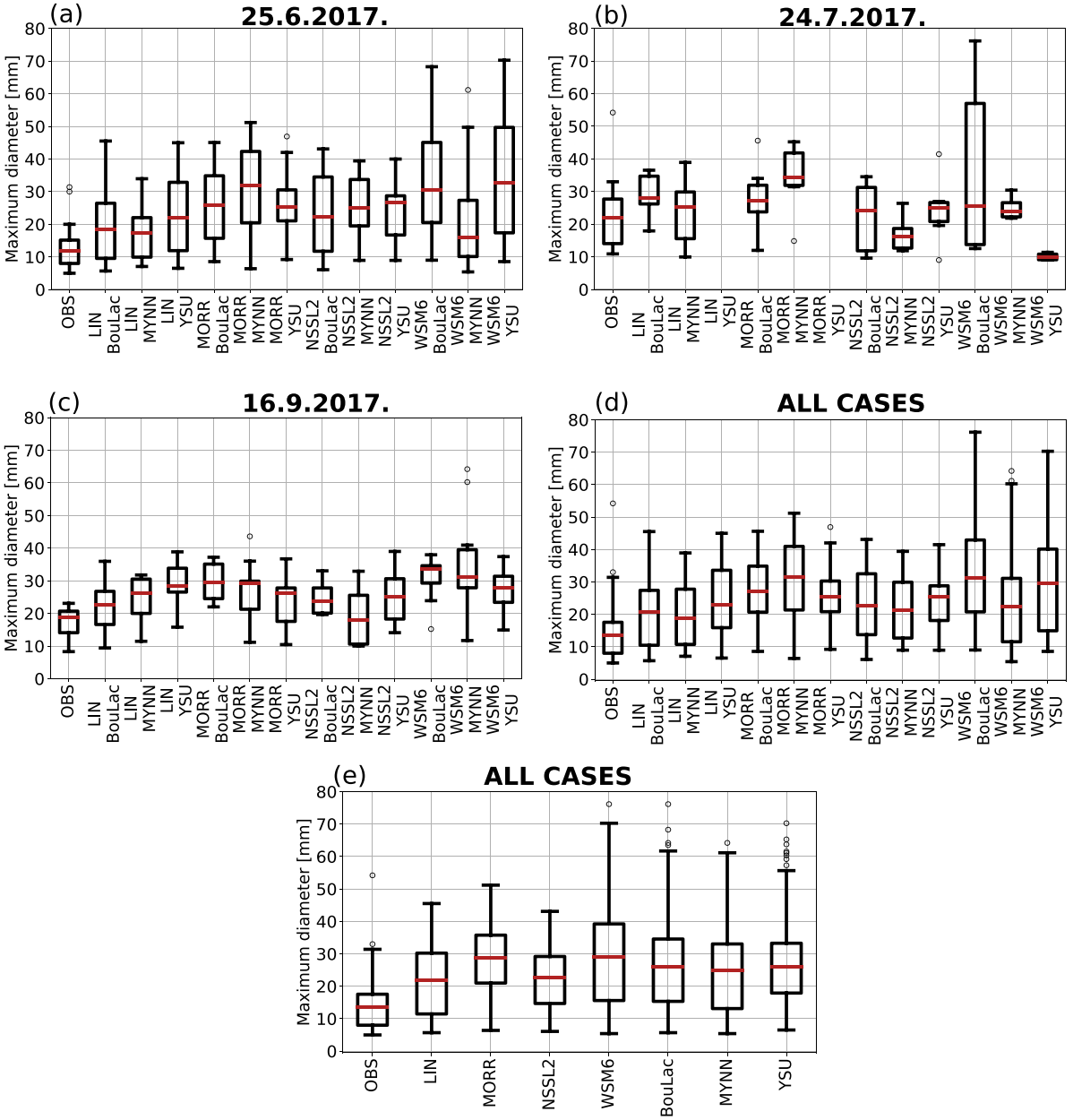

Fig. 12. Comparison between observed (OBS) and simulated maximum hailstone diameters at the point closest to the impacted hailpad for (a) 25 June 2017, (b) 24 July 2017, (c) 16 September 2017, (d) all cases considered, for all ensemble members, and (d) each parametrization scheme in all cases considered. Boxes indicate medians and lower and upper quartiles, while whiskers are within 1.5 interquartile range from the lower and the upper quartile. Outliers are marked with black dots.

that even with 22 km verification radius there still is a substantial amount of false alarms compared to the observations from hailpad network.

A point-to-point comparison between observed and simulated maximum hailstone diameters is also performed (Fig. 12). The boxplots depict maximum simulated hailstone diameters (larger than 5 mm) at the point closest to the hailpad that registered hail for each individual case, for all cases considered and for every parameterization scheme analyzed in all cases considered. All ensemble members for 25 June 2017 and 16 September 2017 show a general overestimation of observed maximum hailstone diameters. On the other hand, only the WSM6-BouLac member was able to produce hailstone diameters of 54.2 mm observed on 24 July 2017. Interestingly, WSM6-MYNN member produced the largest simulated maximum hailstone diameters for 16 September 2017 case and shows the greatest spread of simulated hailstone sizes. On the other hand, the largest hailstones with the largest spread of hailstone sizes for 25 June 2017 case are produced by WSM6-YSU and WSM6-BouLac members. Further, considering all cases together, it is apparent that WSM6 members produced the largest hailstones with the largest spread of hailstone sizes. Moreover, WSM6 and MORR schemes seem to produce larger medians of hailstone sizes compared to LIN and NSSL2 microphysics schemes. This could be connected to the analysis presented in Manzato et al. (2020) who reported that WSM6 scheme reproduced hailstone sizes more similar to those observed than the other analyzed microphysics schemes. Interestingly, looking at all cases together, the discrepancies between simulated hailstone sizes among different PBL schemes considered are less apparent, although, YSU scheme seems to produce smaller spread of hailstone sizes which could be linked with the tendency of nonlocal schemes to produce too much dry air at the surface and underestimate MLCAPE. By discussing these results, one should note that the described comparison has several limitations: (i) the probability for the hailpad to

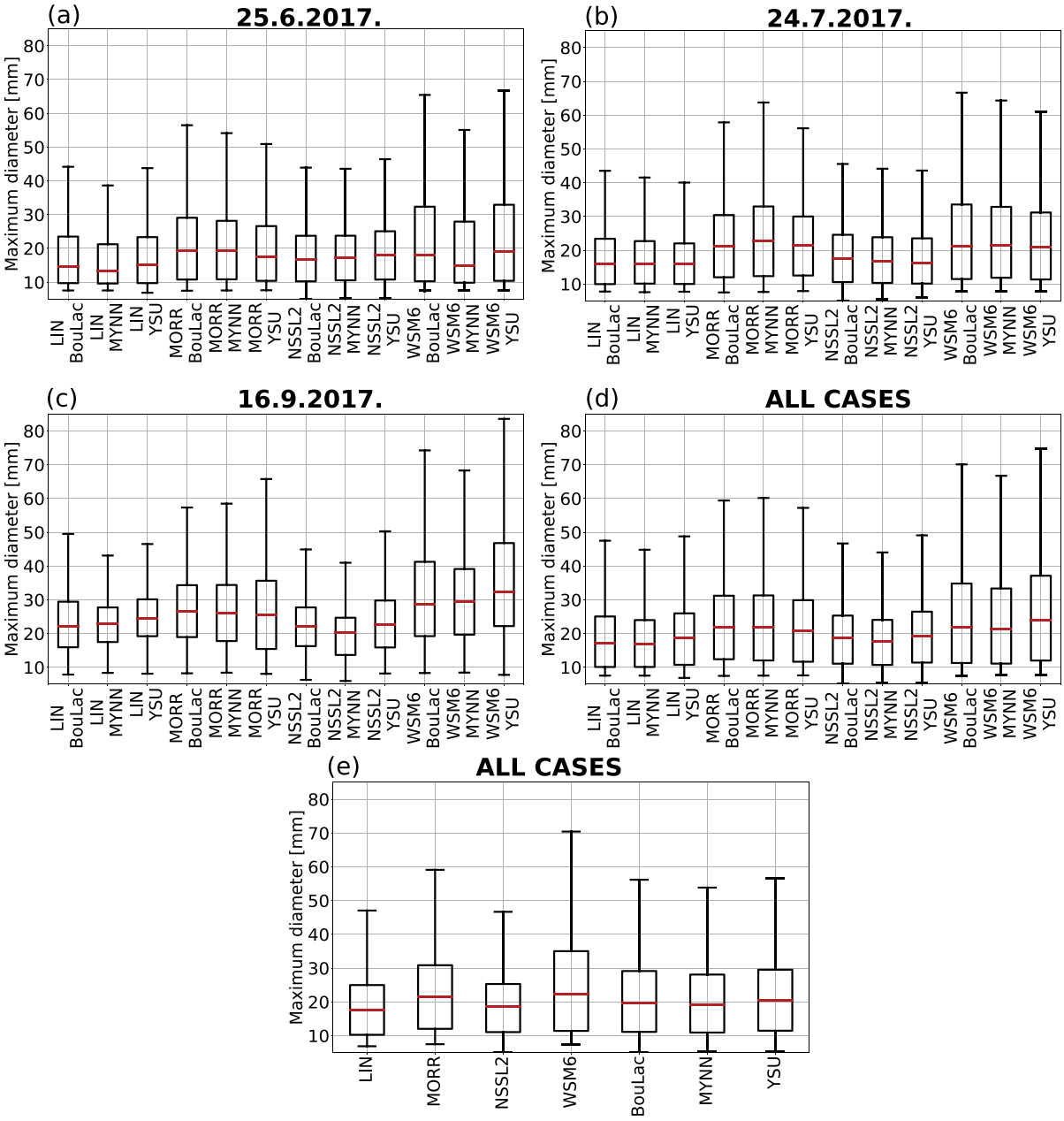

Fig. 13. Boxplots of maximum hailstone sizes across the entire domain and all ensemble members for (a) 25 June 2017, (b) 24 July 2017, (c) 16 September 2017, (d) all cases considered, and (d) each parametrization scheme in all cases considered. Boxes indicate medians and lower and upper quartiles, while whiskers are within 1.5 interquartile range from the lower and the upper quartile.

capture the largest hailstone is small given that the hailpad covers only 0.25 m^2 (Smith and Waldvogel, 1989), and (ii) such point-to-point analysis does not account for possible spatial shifts between observed and simulated hail swaths. Nonetheless, given all limitations, the obtained results seem to be in line with those by Adams-Selin et al. (2019), who reported that HAILCAST tends to underestimate hailstone sizes in the 50–74 mm category, while hailstone sizes in the 19–24 mm category are mostly overestimated (their Fig. 4e). However, given that hailstones in the 50–74 mm category are observed only in one case analyzed, a more detailed analysis of HAILCAST performance in simulating larger hailstone sizes (those in the 50–74 mm category) needs to be done to confirm observed tendencies.

4.4. Discrepancies between schemes

Simulated maximum hailstone diameters are compared among examined parameterization combinations and analyzed cases. Fig. 13 highlights boxplots of simulated maximum hailstone diameters across the entire domain. It is clear that the median hailstone sizes vary within 15 mm range across the entire ensemble. However, some tendencies could be extrapolated – maximum hailstone diameters are obtained by WSM6 microphysics members in all cases considered. On the other hand, minimum hailstone diameters are obtained by NSSL2 microphysics members. Moreover, there is an overall tendency for WSM6 and MORR members to produce larger hailstones with greater interquartile spread of maximum hailstone sizes compared with the NSSL2 and LIN members.

It is important to stress that both LPI and HAILCAST are highly dependent on updraft strengths and the graupel mixing ratios inside a thunderstorm. Fig. 14 displays comparable medians of updraft strength among ensemble members. It seems that there is a tendency for NSSL2 members to produce weaker updrafts compared to other scheme

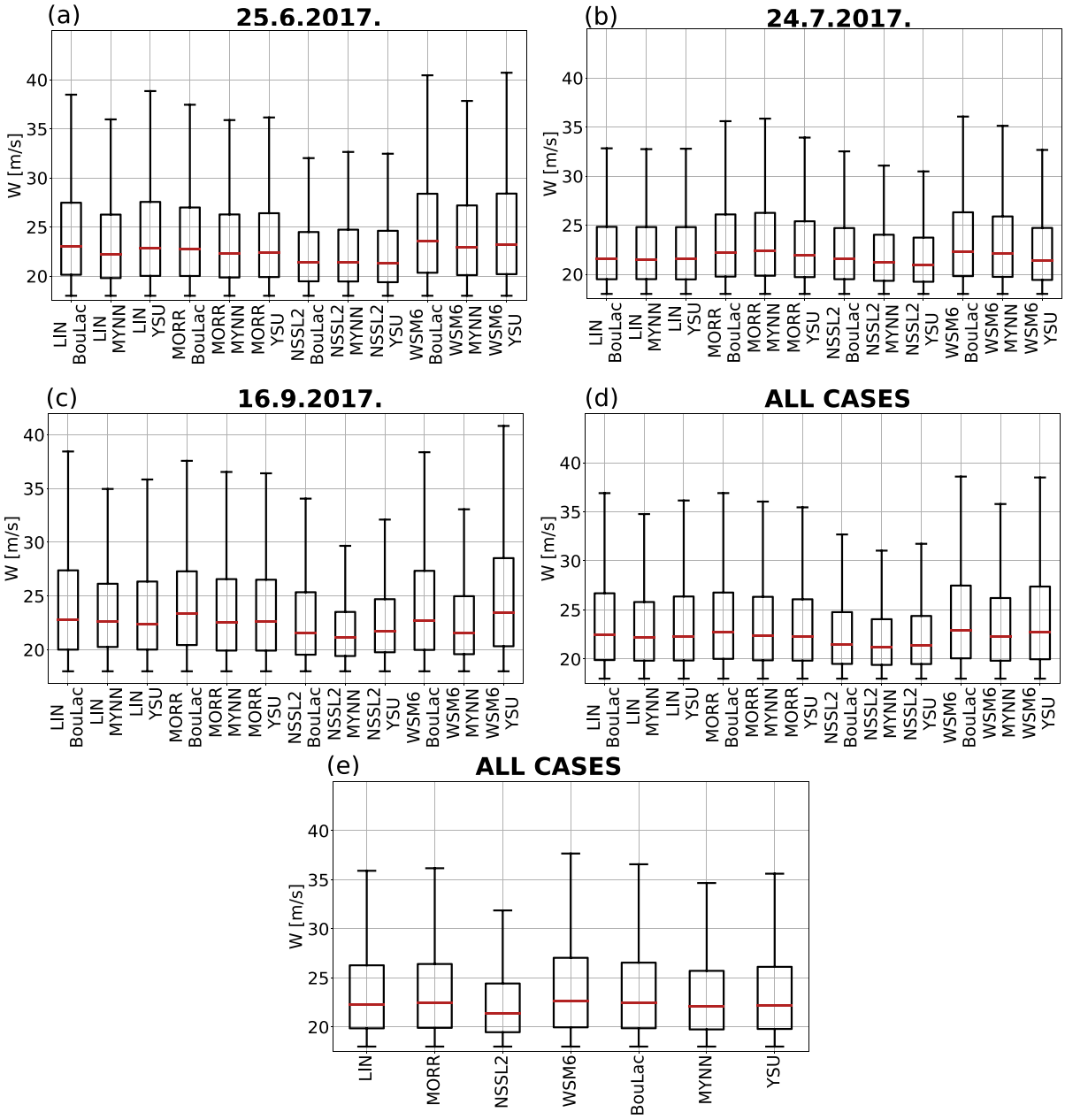

Fig. 14. Boxplots of maximum vertical velocities for the points where hail is forecasted across the entire domain and all ensemble members for (a) 25 June 2017, (b) 24 July 2017, (c) 16 September 2017, (d) all cases considered, and (d) each parametrization scheme in all cases considered. Boxes indicate medians and lower and upper quartiles, while whiskers are within 1.5 interquartile range from the lower and the upper quartile.

combinations. Additionally, there is also an indication that WSM6 members produce slightly stronger updrafts; however, this difference between members is not as pronounced as it is when comparing hailstone diameters. Nonetheless, in Fig. 15, a clear distinction in simulated graupel mixing ratios among members is displayed. NSSL2 members produce the smallest graupel mixing ratios, presumably due to weaker updrafts. The stronger the updrafts are, the higher the altitude at which particles can be transported and consequently the higher the number of ice particles that could collide with each other in the presence of supercooled liquid water, finally resulting in charge separation and hailstone growth processes. This could explain why NSSL2 members produce less intense lightning activity and hail intensity. These members produce weaker updrafts, resulting in fewer graupel particles, which leads to fewer collisions between species inside a thundercloud and thus weaker lightning and hail intensity. Interestingly, LIN members exhibit the greatest medians of graupel mixing ratios, but with larger spread compared with WSM6 and MORR members and greater vertical velocities compared to MORR members, but at the same time tend to produce smaller hailstones. Further, the discrepancies between the performance of different PBL schemes are not as pronounced indicating that simulated hailstone sizes, updrafts and graupel mixing ratios are more sensitive to the choice of microphysics than PBL scheme.

These results could be linked to those reported by Lagasio et al. (2017), who analyzed LPI performance in a multi-microphysical ensemble simulation of a back-building mesoscale convective system over Genoa, Italy. They found that the NSSL2 scheme with the predicted cloud condensation nuclei (CCN) concentration (note that we use NSSL2 scheme with steady background CCN) produced a weaker vertical velocity field and lower graupel mixing ratios with respect to the WSM6 scheme. This led to a less pronounced LPI and less intense lightning

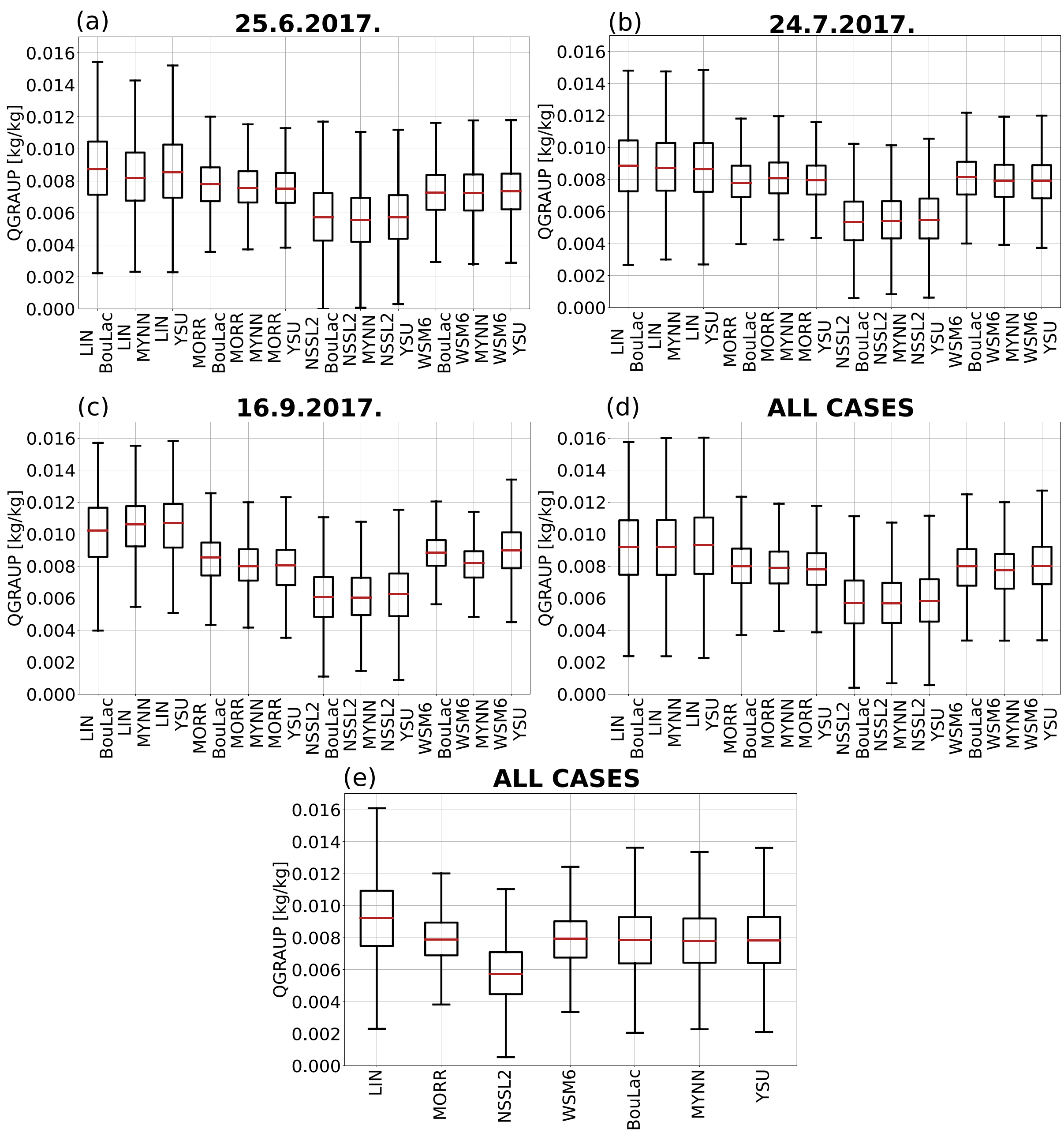

Fig. 15. Boxplots of maximum graupel mixing ratios for the points where hail was forecasted across the entire domain and all ensemble members for (a) 25 June 2017, (b) 24 July 2017, (c) 16 September 2017, (d) all cases considered, and (d) each parametrization scheme in all cases considered. Boxes indicate medians and lower and upper quartiles, while whiskers are within 1.5 interquartile range from the lower and the upper quartile.

activity obtained by the NSSL2 scheme with predicted CCN compared with those obtained by the WSM6 scheme. Moreover, the strongest updrafts and highest graupel mixing ratios combined with the highest LPI and most intense lightning activity were produced by the member with the WSM6 parameterization scheme.

The presented results showed that the simulated area affected with hail is sensitive both to the choice of microphysics and PBL scheme. Namely, microphysics scheme NSSL2 produces smaller areas affected with hail, thus resulting in smaller POD and EDI values compared to the other microphysics options. However, the performance between other microphysics options (LIN, MORR and WSM6) is similar. Similarly, PBL scheme BouLac seems to perform better in terms of simulating the area affected with hail properly, while the performance between other PBL schemes, namely, MYNN and YSU is similar, although YSU performs slightly worse. This could be due to the fact that nonlocal PBL schemes (such as YSU) produce drier and deeper PBL resulting in MLCAPE underestimation. On the other hand, BouLac scheme is designed to perform better in cases with terrain-induced turbulence (Bougeault and Lacarrere, 1989). Since our simulations are performed in a highly complex terrain, this could be the reason that BouLac performs better compared to the other PBL schemes.

Further, simulated hailstone sizes, updrafts and graupel mixing ratios are more sensitive to the choice of microphysics scheme. Namely, microphysics schemes WSM6 and MORR produce larger median hailstone sizes compared to the LIN and NSSL2 microphysics schemes, with WSM6 producing larger hailstones compared to MORR. On the other hand, NSSL2 microphysics scheme produces the smallest hailstones, the weakest updrafts and the smallest graupel mixing ratios.

5. Conclusions

Although relatively frequent in Croatia, hail still remains a difficult phenomenon to forecast. For this reason, three selected hailstorms that occurred in Croatia are simulated with the WRF model at a 1 km grid spacing using HAILCAST and LPI alongside with a multiphysics ensemble of 12 members with combinations of four different microphysics and three different PBL parameterization schemes. The main goal of this work is to assess the forecast ability and sensitivity to parameterization choice of HAILCAST and LPI in predicting hailstorms in Croatia. A detailed evaluation of hail and lightning results against hail observations from the hailpad network and lightning observations provided by the LINET network is performed.

Along these lines, the first part of the paper focuses on an assessment of model performance to simulate general surface conditions associated with hailstorms in Croatia. Model results are evaluated against hourly measurements of 2-m temperature, 2-m relative humidity, 2-m equivalent potential temperature and hourly maximums of 10-m wind speed from the Croatian network of automated meteorological stations. Using RMSE decomposition, it is shown that all ensemble members yield very similar performance across all hail cases analyzed. Moreover, the greatest contribution to RMSE in most members stems from dispersion or phase errors. Overall, surface measurements are reproduced in the complementary range of other numerical studies and ECMWF IFS by all ensemble members.

The second part of this paper focuses on a comparison between LPI and observed lightning activity. Here, a conversion of LPI to the number of lightning flashes is performed, and the results of converted LPI are assessed via the SAL method against the observed number of lightning flashes. All ensemble members show good performance in simulating daily lightning activity associated with all three hailstorms analyzed, except those with the NSSL2 microphysics scheme. These members systematically underestimate the number of observed lightning flashes. On the other hand, in most cases, the most intense lightning activity is produced by the members with the WSM6 microphysics scheme. The temporal characteristics of observed lightning are reproduced only for the case with nocturnal convection; in other two cases with afternoon convection, lightning activity is simulated earlier than observed.

The third part of the paper assesses the HAILCAST results against hail observations from the hailpad network. First, tuning of HAILCAST is performed and a new threshold for HAILCAST activation of 18 ms^{-1} for vertical velocity is introduced to eliminate some of the false alarms connected to activating HAILCAST in too many grid points. Second, to eliminate some of the temporal uncertainty between simulated and observed convection, only daily accumulated fields of hail forecasting are analyzed. The results show good agreement between the observed and modelled hail occurrences. However, there are many false alarms present, indicating the model's tendency to generally overestimate the area affected by hail. Notably, some of the false alarms could be attributed to the limited spatial information regarding hail occurrence that hailpad networks provide. Moreover, comparing simulated and observed maximum hailstone diameters, it seems that HAILCAST, on average, tends to overestimate observed hailstone diameters. Finally, it must be noted that evaluation is performed for three hailstorms only, and to achieve statistically more robust conclusions, a considerably larger number of hailstorms with larger spread of hailstone sizes needs to be included in the evaluation.

Furthermore, the simulated maximum hailstone diameters are compared across the ensemble members. Maximum hailstones are simulated by the members with the WSM6 microphysics scheme, while the minimum hailstones are obtained by the members with the NSSL2 microphysics scheme. Following this, the vertical velocity maximums and graupel mixing ratio maximums, representing the variables related to lightning and hail occurrence, reveal that the main difference between members with the NSSL2 microphysics scheme and other members is the representation of convective updrafts resulting in significantly different graupel mixing ratios. Particularly, the NSSL2 microphysics scheme members produce weaker updrafts in analyzed case studies, lower graupel mixing ratios and consequently lower potential for lightning activity and smaller hailstones. Interestingly, LIN members exhibit the greatest medians of graupel mixing ratios, but with larger spread compared with WSM6 and MORR members and greater vertical velocities compared to MORR members, but at the same time tend to produce smaller hailstones.

Therefore, this analysis indicates that the members that better reproduced observed lighting activity and hail occurrence also produce a better distribution of updrafts and mixing ratios, comparable to those occurring during hailstorms in Croatia. This highlights the fact that there is no much impact on the results solely from HAILCAST and LPI, rather most of the skill depends on the WRF's skill to simulate convection properly, if not also on the initialization fields from ECMWF analysis. For that reason, an ensemble approach should be adopted in operational use for prediction of lightning and hail. Moreover, even though hailpads are probably the only source of hail measurements at the ground, it would be extremely useful to assess the simulation results using other data sources as well, such as radar products, insurance claims or crowd-sourced hail reports. That would complement the hailpad dataset and would provide a spatially continuous information on hail occurrence. Following this, to get statistically more robust conclusions on HAILCAST and LPI performance, a considerably larger number of hailstorms needs to be examined. Further, in operational setting, data assimilation should be implemented to further improve predictability of such extreme events. However, given all limitations, the results presented here are promising and show that both HAILCAST and LPI could be valuable tools for real-time forecasting and climatological assessment of hail and lightning occurrence in current and possibly changing climates.

Data Availability

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). ECMWF model analysis fields used as initial and boundary

conditions can be obtained through the following link (http://www.ecmwf.int/). Hail and surface measurements are available through inquiries of the Croatian Meteorological and Hydrological Service; for more information, please contact the following email address (uslu ge@cirus.dhz.hr).

Declaration of Competing Interest

The authors declare no financial and personal relationships with other people or organizations that could inappropriately influence the work reported in this paper.

Acknowledgments

This research is enabled by 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; 29 July 2021). This research was supported by the ECMWF (http://www.ecmwf.int/) data and the WRF model (freely available at www.wrf-model.org/index.php). The simulations were run using HPC Bura (https://cnrm.uniri.hr/bura/). We thank Adams-Selin, R.D. for providing us with the HAILCAST code coupled within WRF and DHMZ for providing us with hail and surface measurements. Special thanks to members of Climate and Water Cycle group from ETH for providing us with LPI code and many valuable and useful discussions. The anonymous reviewers are gratefully acknowledged for their in-depth review and valuable suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2022.106143.

References

Adams-Selin, R.D., Ziegler, C.L., 2016. Forecasting hail using a one-dimensional hail growth model within WRF. Mon. Weather Rev. 144, 4919–4939. https://doi.org/10.1175/MWR-D-16-0027.1.

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 Forecast. 34, 61–79. https://doi.org/10.1175/WAF-D-18-0024.1.

Allen, D.J., Pickering, K.E., 2002. Evaluation of lightning flash rate parameterizations for use in a global chemical transport model. J. Geophys. Res. Atmos. 107 https://doi.org/10.1029/2002JD002066. ACH 15-1-ACH 15-21.

Altaratz, O., 2005. Simulation of the electrification of winter thunderclouds using the three-dimensional Regional Atmospheric Modeling System (RAMS) model: single cloud simulations. J. Geophys. Res. 110, D20205. https://doi.org/10.1029/2004JD005616.

Ban, N., Schmidli, J., Schär, C., 2015. Heavy precipitation in a changing climate: does short-term summer precipitation increase faster? Geophys. Res. Lett. 42, 1165–1172. https://doi.org/10.1002/2014GL062588.

Barthe, C., Pinty, J.-P., 2007. Simulation of a supercellular storm using a three-dimensional mesoscale model with an explicit lightning flash scheme. J. Geophys. Res. 112, D06210. https://doi.org/10.1029/2006JD007484.

Barthe, C., Chong, M., Pinty, J.P., Bovalo, C., Escobar, J., 2012. CELLS v1.0: Updated and parallelized version of an electrical scheme to simulate multiple electrified clouds and flashes over large domains. Geosci. Model Dev. 5, 167–184. https://doi.org/10.5194/gmd-5-167-2012.

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.P., Beljaars, A., Bormann, N., 2014. Representing equilibrium and nonequilibrium convection in large-scale models. J. Atmos. Sci. 71, 734–753. https://doi.org/10.1175/JAS-D-13-0163.1.

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. Atmos. Res. 91, 564–573. https://doi.org/10.1016/j.atmosres.2008.06.012.

Bougeault, P., Lacarrere, P., 1989. Parameterization of orography-induced turbulence in a mesobeta-scale model. Mon. Weather Rev. 117, 1872–1890. https://doi.org/10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2.

Brimelow, J.C., Reuter, G.W., Poolman, E.R., 2002. Modeling maximum hail size in Alberta thunderstorms. Weather Forecast. 17, 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. Clim. Dyn. 2021 (1), 1–15. https://doi.org/10.1007/S00382-021-05791-Z.

Brown, T.M., Pogorzelski, W.H., Giammanco, I.M., 2015. Evaluating hail damage using property insurance claims data. Weather. Clim. Soc. 7, 197–210. https://doi.org/10.1175/WCAS-D-15-0011.1.

Bryan, G.H., Morrison, H., 2012. Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics. Mon. Weather Rev. 140, 202–225. https://doi.org/10.1175/MWR-D-11-00046.1.

Changnon, S.A., 2009. Increasing major hail losses in the U.S. Clim. Chang. 96, 161–166. https://doi.org/10.1007/s10584-009-9597-z.

Chen, S.-H., Sun, W.-Y., 2002. A one-dimensional time dependent cloud model. J. Meteorol. Soc. Japan 80, 99–118.

Clark, P., Roberts, N., Lean, H., Ballard, S.P., Charlton-Perez, C., 2016. Convection-permitting models: a step-change in rainfall forecasting. Meteorol. Appl. 23, 165–181. https://doi.org/10.1002/MET.1538.

Cohen, A.E., Cavallo, S.M., Coniglio, M.C., Brooks, H.E., 2015. A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. Weather Forecast. 30, 591–612. https://doi.org/10.1175/WAF-D-14-00105.1.

Coniglio, M.C., 2012. Verification of RUC 0–1-h forecasts and SPC mesoscale analyses using VORTEX2 soundings. Weather Forecast. 27, 667–683. https://doi.org/10.1175/WAF-D-11-00096.1.

Czernecki, B., Taszarek, M., Marosz, M., Półrolniczak, M., Kolendowicz, L., Wyszogrodzki, A., Szturc, J., 2019. Application of machine learning to large hail prediction - the importance of radar reflectivity, lightning occurrence and convective parameters derived from ERA5. Atmos. Res. 227, 249–262. https://doi.org/10.1016/j.atmosres.2019.05.010.

Dalezios, N.R., Loukas, A., Bampzelis, D., 2002. Universal kriging of hail impact energy in Greece. Phys. Chem. Earth 27, 1039–1043. https://doi.org/10.1016/S1474-7065(02)00137-7.

Davis, C.A., Brown, B., Bullock, R., 2006. Object-based verification of precipitation forecasts. Part I: application to convective rain systems. Mon. Weather Rev. 134, 1785–1795. https://doi.org/10.1175/MWR3146.1.

Dudhia, J., 1989. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. J. Atmos. Sci. 46, 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. Meteorol. Appl. 15, 51–64. https://doi.org/10.1002/met.25.

Farnell, C., Rigo, T., Pineda, N., 2017. Lightning jump as a nowcast predictor: application to severe weather events in Catalonia. Atmos. Res. 183, 130–141. https://doi.org/10.1016/j.atmosres.2016.08.021.

Farnell, C., Rigo, T., Pineda, N., 2018. Exploring radar and lightning variables associated with the Lightning Jump. Can we predict the size of the hail? Atmos. Res. 202, 175–186. https://doi.org/10.1016/j.atmosres.2017.11.019.

Fawbush, E.J., Miller, R.C., 1953. A method for forecasting hailstone size at the earth's surface. Bull. Am. Meteorol. Soc. 34, 235–244. https://doi.org/10.1175/1520-0477-34.6.235.

Ferro, C.A.T., Stephenson, D.B., 2011. Extremal dependence indices: improved verification measures for deterministic forecasts of rare binary events. Weather Forecast. 26, 699–713. https://doi.org/10.1175/WAF-D-10-05030.1.

Fierro, A.O., Mansell, E.R., Macgorman, D.R., Ziegler, C.L., 2013. The implementation of an explicit charging and discharge lightning scheme within the wrf-arw model: benchmark simulations of a continental squall line, a tropical cyclone, and a winter storm. Mon. Weather Rev. 141, 2390–2415. https://doi.org/10.1175/MWR-D-12-00278.1.

Finney, D.L., Doherty, R.M., Wild, O., Huntrieser, H., Pumphrey, H.C., Blyth, A.M., 2014. Using cloud ice flux to parametrise large-scale lightning. Atmos. Chem. Phys. 14, 12665–12682. https://doi.org/10.5194/acp-14-12665-2014.

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. Atmos. Res. 138, 13–29. https://doi.org/10.1016/j.atmosres.2013.10.007.

Franc, B., Filipović-Grčić, B., Milardić, V., 2016. Lightning overvoltage performance of 110 kV air-insulated substation. Electr. Power Syst. Res. 138, 78–84. https://doi.org/10.1016/J.EPSR.2015.12.002.

Gagne, D.J., McGovern, A., Haupt, S.E., Sobash, R.A., Williams, J.K., Xue, M., 2017. Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. Weather Forecast. 32, 1819–1840. https://doi.org/10.1175/WAF-D-17-0010.1.

Helsdon, J.H., Wu, G., Farley, R.D., 1992. An intracloud lightning parameterization scheme for a storm electrification model. J. Geophys. Res. 97, 5865. https://doi.org/10.1029/92JD00077.

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

Hong, S.Y., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Weather Rev. 134, 2318–2341. https://doi.org/10.1175/MWR3199.1.

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. J. Geophys. Res. Atmos. 117 https://doi.org/10.1029/2012JD017432.

Horvath, K., Šepić, J., Telišman Prtenjak, M., 2019. Atmospheric forcing conducive for the adriatic 25 June 2014 meteotsunami event. In: Meteorology and Climatology of

the Mediterranean and Black Seas. Birkhäuser, Cham, pp. 97–117. https://doi.org/10.1007/978-3-030-11958-4_7.
Jelić, D., Megyeri, O.A., Malečić, B., Belušić Vozila, A., Strelec Mahović, N., Telišman Prtenjak, M., 2020. Hail climatology along the northeastern adriatic. J. Geophys. Res. Atmos. 125 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. Atmos 12, 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 Forecast. 24, 1592–1609. https://doi.org/10.1175/2009WAF2222230.1.
Jirak, I.L., Coniglio, M., Clark, A.J., Correia, J., Knopfmeier, k.H., Melick, C.J., Weiss, S. J., Kain, J.S., Xue, M., Kong, F., Thomas, K.W., Brewster, K., Wang, Y., Jung, Y., Willington, S., 2014. An overview of the 2014 NOAA hazardous weather testbed spring forecasting experiment.
Johns, R.H., Doswell, C.A., 1992. Severe local storms forecasting. Weather Forecast. 7, 588–612. https://doi.org/10.1175/1520-0434(1992)007<0588:slsf>2.0.co;2.
Jurković, P.M., Mahović, N.S., Počakal, D., 2015. Lightning, overshooting top and hail characteristics for strong convective storms in Central Europe. Atmos. Res. 161–162, 153–168. https://doi.org/10.1016/j.atmosres.2015.03.020.
Kain, J.S., Kain, J., 2004. The Kain - Fritsch convective parameterization: an update. J. Appl. Meteorol. 43, 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.
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.
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 Forecast. 100804092600065 https://doi.org/10.1175/2008waf2007106.1.
Kehler-Poljak, G., Prtenjak, M.T., Kvakić, M., Šariri, K., Večenaj, Ž., 2017. Interaction of sea breeze and deep convection over the northeastern adriatic coast: an analysis of sensitivity experiments using a high-resolution mesoscale model. Pure Appl. Geophys 17411 (174), 4197–4224. https://doi.org/10.1007/S00024-017-1607-X.
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. Q. J. R. Meteorol. Soc. 144, 231–250. https://doi.org/10.1002/qj.3197.
Lagasio, M., Parodi, A., Procopio, R., Rachidi, F., Fiori, E., 2017. Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: the Genoa 2014 event. J. Geophys. Res. 122, 4238–4257. https://doi.org/10.1002/2016JD026115.
Lin, Y.L., Farley, R.D., Orville, H.D., 1983. Bulk parameterization of the snow field in a cloud model. J. Clim. Appl. Meteorol. 22, 1065–1092. https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.
Lopez, P., 2016. A lightning parameterization for the ECMWF integrated forecasting system. Mon. Weather Rev. 144, 3057–3075. https://doi.org/10.1175/MWR-D-16-0026.1.
Lopez, P., 2018. Promising results for lightning predictions | ECMWF. ECMWF Newsl. 155.
Lynn, B.H., Yair, Y.Y., 2008. Lightning Power Index: A new tool for predicting the lightning density and the potential for extreme rainfall. Geophys. Res. Abstr. 10 (EGU2008-A-01571).
MacGorman, D.R., Straka, J.M., Ziegler, C.L., 2001. A lightning parameterization for numerical cloud models. J. Appl. Meteorol. 40, 459–478. https://doi.org/10.1175/1520-0450(2001)040<0459:ALPFNC>2.0.CO;2.
Mansell, E.R., 2005. Charge structure and lightning sensitivity in a simulated multicell thunderstorm. J. Geophys. Res. 110, D12101. https://doi.org/10.1029/2004JD005287.
Mansell, E.R., Ziegler, C.L., Bruning, E.C., 2010. Simulated electrification of a small thunderstorm with two-moment bulk microphysics. J. Atmos. Sci. 67, 171–194. https://doi.org/10.1175/2009JAS2965.1.
Manzato, A., 2008. A verification of numerical model forecasts for sounding-derived indices above Udine, northeast Italy. Weather Forecast. 23, 477–495. https://doi.org/10.1175/2007WAF2007018.1.
Manzato, A., 2013. Hail in Northeast Italy: a neural network ensemble forecast using sounding-derived indices. Weather Forecast. 28, 3–28. https://doi.org/10.1175/WAF-D-12-00034.1.
Manzato, A., Pucillo, A., Cicogna, A., 2018. Improving ECMWF-based 6-hours maximum rain using instability indices and neural networks. Atmos. Res. 217, 184–197. https://doi.org/10.1016/j.atmosres.2018.10.020.
Manzato, A., Riva, V., Tiesi, A., Marcello Miglietta, M., 2020. Observational analysis and simulations of a severe hailstorm in northeastern Italy. Q. J. R. Meteorol. Soc. 146, 3587–3611. https://doi.org/10.1002/qj.3886.
Marzban, C., Witt, A., 2001. A Bayesian neural network for severe-hail size prediction. Weather Forecast. 16, 600–610. https://doi.org/10.1175/1520-0434(2001)016<0600:ABNNFS>2.0.CO;2.
Mikuš, P., Telišman Prtenjak, M., Strelec Mahović, N., 2012. Analysis of the convective activity and its synoptic background over Croatia. Atmos. Res. 104–105, 139–153. https://doi.org/10.1016/J.ATMOSRES.2011.09.016.
Milovac, J., Warrach-Sagi, K., Behrendt, A., Späth, F., Ingwersen, J., Wulfmeyer, V., 2016. Investigation of PBL schemes combining the WRF model simulations with scanning water vapor differential absorption lidar measurements. J. Geophys. Res. Atmos. 121, 624–649. https://doi.org/10.1002/2015JD023927.
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. J. Geophys. Res. Atmos. 102, 16663–16682. https://doi.org/10.1029/97jd00237.
Moore, J.T., Pino, J.P., 1990. An interactive method for estimating maximum hailstone size from forecast soundings. Weather Forecast. 5, 508–525. https://doi.org/10.1175/1520-0434(1990)005<0508:aimfem>2.0.co;2.
Morrison, H., Thompson, G., Tatarskii, V., 2009. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment schemes. Mon. Weather Rev. 137, 991–1007. https://doi.org/10.1175/2008MWR2556.1.
Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Weather Rev. 116, 2417–2424. https://doi.org/10.1175/1520-0493(1988)116.
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 Meteorol. 119, 397–407. https://doi.org/10.1007/s10546-005-9030-8.
Nisi, L., Ambrosetti, P., Clementi, L., 2014. Nowcasting severe convection in the Alpine region: the COALITION approach. Q. J. R. Meteorol. Soc. 140, 1684–1699. https://doi.org/10.1002/qj.2249.
Orlić, M., Belušić, D., Janeković, I., Pasarić, M., 2010. Fresh evidence relating the great Adriatic surge of 21 June 1978 to mesoscale atmospheric forcing. J. Geophys. Res. Ocean. 115, 6011. https://doi.org/10.1029/2009JC005777.
Peirce, C.S., 1884. The numerical measure of the success of predictions. Science (80-.) 4, 453–454. https://doi.org/10.1126/SCIENCE.NS-4.93.453-A.
Pinty, J.P., Barthe, C., 2008. Ensemble simulation of the lightning flash variability in a 3D cloud model with parameterization of cloud electrification and lightning flashes. Mon. Weather Rev. 136, 380–387. https://doi.org/10.1175/2007MWR2186.1.
Počakal, D., 2011. Hailpad data analysis for the continental part of Croatia. Meteorol. Zeitschrift 20, 441–447. https://doi.org/10.1127/0941-2948/2011/0263.
Počakal, D., Večenaj, Ž., Štalec, J., 2009. Hail characteristics of different regions in continental part of Croatia based on influence of orography. Atmos. Res. 93, 516–525. https://doi.org/10.1016/j.atmosres.2008.10.017.
Počakal, D., Večenaj, Ž., Mikuš Jurković, P., Grisogono, B., 2018. Analysis of orographic influence on hail parameters in NW Croatia. Int. J. Climatol. 38, 5646–5658. https://doi.org/10.1002/joc.5769.
Poljak, G., Prtenjak, M.T., Kvakić, M., Strelec Mahović, N., Babić, K., Poljak, G., Prtenjak, M.T., Kvakić, M., Strelec Mahović, N., Babić, K., 2014. Wind patterns associated with the development of daytime thunderstorms over Istria. AnGeo 32, 401–420. https://doi.org/10.5194/ANGEO-32-401-2014.
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, p. 113.
Prein, A.F., Rasmussen, R.M., Wang, D., Giangrande, S.E., 2021. Sensitivity of organized convective storms to model grid spacing in current and future climates. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 379 https://doi.org/10.1098/rsta.2019.0546.
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. Mon. Weather Rev. 147, 3901–3916. https://doi.org/10.1175/MWR-D-19-0204.1.
Raupach, T.H., Martius, O., Allen, J.T., Kunz, M., Lasher-Trapp, S., Mohr, S., Rasmussen, K.L., Trapp, R.J., Zhang, Q., 2021. The effects of climate change on hailstorms. Nat. Rev. Earth Environ. 2, 213–226. https://doi.org/10.1038/s43017-020-00133-9.
Renko, T., Ivušić, S., Telišman Prtenjak, M., Šoljan, V., Horvat, I., 2018. Waterspout forecasting method over the eastern adriatic using a high-resolution numerical weather model. Pure Appl. Geophys. 175, 3759–3778. https://doi.org/10.1007/s00024-018-1833-x.
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 (80-.) 346, 851–854. https://doi.org/10.1126/science.1259100.
Saunders, C., 2008. Charge separation mechanisms in clouds. Space Sci. Rev. 137, 335–353. https://doi.org/10.1007/s11214-008-9345-0.
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. Nat. Hazards Earth Syst. Sci. 5, 613–620. https://doi.org/10.5194/nhess-5-613-2005.
Šepić, J., Vilibić, I., Belušić, D., 2009. Source of the 2007 Ist meteotsunami (Adriatic Sea). J. Geophys. Res. 114, 3016. https://doi.org/10.1029/2008JC005092.
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.-Y., Wang, W., Powers, J.G., 2008a. A Description of the Advanced Research WRF Version 3.
Skamarock, William C., Klemp, J.B., Skamarock, W.C., Klemp, J.B., 2008b. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. JCoPh 227, 3465–3485. https://doi.org/10.1016/J.JCP.2007.01.037.
Smith, P.L., Waldvogel, A., 1989. On determinations of maximum hailstone sizes from hallpad observations. J. Appl. Meteorol. 28, 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. Atmos. Res. 237 https://doi.org/10.1016/j.atmosres.2020.104862.
Takacs, L.L., 1985. A two-step scheme for the advection equation with minimized dissipation and dispersion errors - NASA Technical Reports Server (NTRS). Mon. Weather Rev. 113.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192. https://doi.org/10.1029/2000JD900719.

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. Atmos. Res. 209, 76–94. https://doi.org/10.1016/j.atmosres.2018.03.007.

Tsenova, B.D., Mitzeva, R.P., 2009. New parameterization of non-inductive charge transfer based on previous laboratory experiments. Atmos. Res. 91, 79–86. https://doi.org/10.1016/j.atmosres.2008.07.001.

Weisman, M.L., Skamarock, W.C., Klemp, J.B., 1997. The resolution dependence of explicitly modeled convective systems. Mon. Weather Rev. 125, 527–548. https://doi.org/10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.

Wernli, H., Paulat, M., Hagen, M., Frei, C., 2008. SAL - a novel quality measure for the verification of quantitative precipitation forecasts. Mon. Weather Rev. 136, 4470–4487. https://doi.org/10.1175/2008MWR2415.1.

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. J. Geophys. Res. Atmos. 115 https://doi.org/10.1029/2008JD010868.