Bulletin of Atmospheric Science and Technology (2023) 4:8
https://doi.org/10.1007/s42865-023-00061-8

RESEARCH

WRF data assimilation of weather stations and lightning data for a convective event in northern Italy

E. C. Maggioni[1] · T. Manzoni[2] · A. Perotto[1] · F. Spada[1] · A. Borroni[2] · M. Giurato[3] · M. Giudici[4] · F. Ferrari[4] · D. Zardi[5] · R. Salerno[2]

Received: 13 February 2023 / Accepted: 29 August 2023

Abstract
The present work shows the relevance of assimilating mesoscale observations and lightning data in the Weather Research and Forecasting (WRF) model, to simulate a strong convective event in northern Italy, poorly forecasted by available weather models even a few hours before the event itself. The data assimilation was conducted by testing the 3D-VAR and 4D-VAR assimilation algorithms implemented in the WRF data assimilation (WRFDA) suite, with different configurations and different assimilation windows. An extensive sensibility test has been operated to properly analyze the effect that the assimilation of a single station has on the model outcomes. Input data were taken from two networks of more than 1000 citizen-science meteorological stations, available in northern Italy, and from lightning flashes derived from Earth Networks Total Lightning Network, assimilated using the atmospheric water vapor as a proxy variable. Rain forecasts over an area in the north of Milan were compared to the station's measurements in the same area; POD, FAR, and CSI categorical statistics have been calculated. Results showed a positive improvement in the forecasted rain amounts with the ingestion of mesoscale weather data into 3D-VAR and 4D-VAR algorithms, more pronounced using 4D-VAR with a more frequent input data integration. A few improvements were reported by the 3D-VAR, with the lightning data assimilation, probably caused by the absence of the model's spin-up time with this configuration. An ideal simulation, which increased the water vapor of the air mass 2 h before the convective event, reported a positive enhancement of the rain amounts. The tests conducted on a single convective event are nevertheless encouraging, because they show a positive improvement of forecast with the assimilation of near-ground weather data and tropospheric water vapor 1 or 2 h before the beginning of the convection activity.

Keywords Data assimilation · Convective rainfall · WRF model · Ground meteorological stations · Lightning data · Humidity profile

✉ E. C. Maggioni
enrico.maggioni@ideamweb.com

Extended author information available on the last page of the article

Published online: 14 September 2023

1 Introduction

Data assimilation is a fundamental tool in modern numerical weather-prediction models. The purpose of data assimilation is to determine the best possible atmospheric state and its uncertainties using all available information, observations, and short-range forecasts (Fletcher 2017). Although data assimilation is widely used in global weather models from decades and several works have demonstrated the importance of data assimilation in the forecast of intense convective precipitation events (Fierro et al. 2012, Mazzarella et al. 2017, Wagner et al. 2022, Gustafsson et al. 2017), it is only a few years since the data assimilation of ground data, radar, and lightning data is routinely applied in limited-area models. The reasons behind this fact are, mainly, the high computational resources needed to assimilate observations on a high-density grid, and the lack of new and novel observation data to use in the assimilation process, i.e., observations that have not already been used in the global models.

Among the various data assimilation algorithms, 3D-VAR and 4D-VAR determine the best estimate of the atmospheric state at analysis time, through the minimization of a cost function, which reduces the gap between observations and the trajectory forecasted by the model, and are of a very common use in recent data assimilation developments.

The latest advances in the real-time availability of observation data from different sources (meteorological weather station, satellite measurements, radar data, lightning data, crowd-sourced observations) are opening new possibilities for data assimilation in limited-area weather models (Giazzi et al. 2022, Hintz et al. 2021).

Various studies about the impact of near-ground weather observations into the Weather Research and Forecasting (WRF) data assimilation system have been conducted through the years. In the work of Tong et al. (2016), synoptic stations were assimilated into the WRF model using a 3D-VAR algorithm (Barker et al. 2003, Barker et al. 2004, Chu et al. 2013) and, updating the input data every hour, an improvement of convection was reported. Wagner et al. (2022) applied a similar strategy using synoptic stations in a 3D-VAR configuration, including GNSS (Global Navigation Satellite System) data in the input information. This procedure permitted the improvement of the simulation of atmospheric water vapor content, not only in convective situations. The first study about the use of GNSS microwave signals goes back to the work by Bevis et al. (1992).

A set-up similar to Wagner's was proposed by Rohm et al. (2019), who assimilated GNSS data into the WRF model through the use of a 4D-VAR assimilation algorithm (Huang et al. 2009). The assimilation of GNSS data was also conducted in the work of Torcasio et al. (2023), with a 3Dvar implementation of the WRF model for a 1-month period, with a focus for a single heavy rain event in Italy.

Other experiences in assimilating GNSS data were conducted by Lagasio et al. (2019), and showed a positive improvement especially for precipitation and water vapor content. Water vapor plays a fundamental role in clouds and precipitation formation, so great attention is required in the best analysis estimate both at ground level and in its vertical distribution. Fersch et al. (2022) focused on the importance of the assimilation of tropospheric input data for the improvement of the water vapor content in the atmospheric column.

Various attempts have also been made to assimilate lightning data into weather forecast models. Lightning data can be assimilated through the inclusion of rain rates, used as a proxy for lightning flashes (Benjamin et al. 2004). Mansell et al. (2007) used lightning data to trigger the convection with the use of a modified Kain-Fritsch cumulus scheme. The same method was applied by Giannaros et al. (2016) for eight events

over Greece, with a positive impact on precipitation forecast. Papadopulos et al. (2005) used lightning data to force a different humidity profile derived from real humidity profiles under heavy rain occurrence and strong convection activity. Some novel ideas were introduced by Fierro et al. (2012) and Fierro et al. (2014), who used lightning data to increase the water vapor mixing ratio in a particular altitudinal range, called the mixed-phase region. The water vapor increase is proportional to the number of lightning flashes, with the use of an hyperbolic tangent function.

Different adaptations of Fierro's method were proposed by Qie et al. (2014), with the modification of graupel, snow, and ice crystal mixing ratios, and by Chen et al. (2019).

In recent years, many novel approaches have been presented for the Italian territory, and for the Mediterranean basin. Given the complex Italian orography, and the fact that Italy is surrounded by a warm sea, it is not unlike the development of heavy rainfall events and convective storms (Miglietta et al. 2021, Federico et al. 2008, Miglietta and Davolio 2022).

Different data assimilation approaches have been applied to the Italian territory. Maiello et al. (2014), Mazzarella et al. (2017), and Avolio et al. (2011) studied the assimilation of radar data into the WRF model, both with 3Dvar and 4Dvar techniques. Avolio et al. (2011) and Ferretti et al. (2005) proposed the assimilation of meteorological data from ground stations for the Italian territory, and showed a positive improvement in rain forecast.

A large amount of research has been recently conducted in the lightning data assimilation topic over the Italian territory. Federico et al. (2017) and Federico et al. (2019) show a positive impact of assimilating lightning data and radar data into the RAMS weather model, implementing the method of Fierro et al. (2012) respectively for 20 cases over Italy and for 2 heavy rain events in Italy.

Torcasio et al. (2021) underlined the importance of lightning data assimilation in the Mediterranean basin, showing positive impact in precipitation events by the assimilation of lightning strikes over land and over sea.

Also, Prat et al. (2021) provided a positive improvement of rain amount in three heavy rain events over Italy, by the assimilation of lightning data in the WRF model.

This present work provides a comprehensive analysis of a real, intense convective event which occurred in the early afternoon on July 11, 2020 in northern Italy, analyzing different data assimilation techniques and configurations. Differently from most of the works cited above, the WRF model has been used in a cold-start configuration, in order to enhance the use of the latest global model initial and boundary condition data. Little evidence is present in the literature about the use of a cold-start configuration with data assimilation for heavy rain events (Mazzarella et al. 2021).

Near-ground observations are provided from different citizen-science weather networks, whereas lightning data are provided by Earth Networks Total Lightning Network (ENTLN). In particular, lightning data are assimilated with a re-adjustment of Fierro et al.'s (2012) algorithm.

Section 2 describes the configuration of the WRF model, the input data used for the assimilation processes, and a description of the convective event studied. A specific paragraph describes the verification methodology adopted.

Section 3 describes the data assimilation configuration used to assimilate weather stations and lightning data. Additionally, the description of an assimilation sensitivity analysis has been added.

In Section 4, the obtained results are discussed. Conclusions are reported in Section 5.

2 Materials and methods

2.1 Model setup

The numerical weather forecasts were generated with the Weather Research and Forecasting model (WRF–ARW 4.2.2) developed by NCAR (National Center of Atmospheric Research). The model is run operationally by the US National Weather Service and, being open source and easily portable, it is widely used around the world for research and weather forecasts. The WRF model is a non-hydrostatic, fully compressible, primitive-equations model, and is deeply described in Skamarock et al. (2019). Simulations have been performed over a domain covering northern Italy and characterized by a 3.0-km grid step: the number of points is respectively 187 in east–west direction, and 127 in north–south direction. For all the simulations performed, the number of terrain-following vertical levels adopted was 37 with model top at 50 hPa (~20 km), and initial and boundary conditions have been provided by the Global Forecast System (GFS) (Environmental Modeling Center 2003). Boundary conditions have been provided to the WRF model every 3 h. All the simulations in the present work have been performed with a cold-start scheme. A first reason behind this setup is the attempt to use the latest GFS model data available as initial and boundary condition for the local WRF simulation: a warm-start configuration does not guarantee the use of the latest initialization data, and this aspect is of crucial importance in the particular case study analyzed in the present paper, wherein convection was not forecasted with the use of the previous global model initialization. A second reason is the attempt to explore the study of a heavy rain event with WRF model data assimilation with a configuration which is not widely used in the literature (Mazzarella et al. 2021). Regarding the model setup, convection has been parameterized using the Betts-Miller-Janjic scheme (Betts and Miller 1993). The main physics choices are the WRF single-moment, 6-class bulk microphysics scheme (WSM6) by Hong et al. (2006), the Yonsei University scheme for the boundary layer, and 5-layer thermal diffusion land surface model.

To describe the subgrid-scale processes related to radiation physics, the rapid radiation transfer model (RRTM) (Mlawer et al. 1997) and the Dudhia (Skamarock et al. 2019) parameterization schemes have been chosen respectively for long-wave and short-wave radiation.

With this configuration, the grid spacing has been chosen in the convection-allowing resolution interval, as used in several experimental NWP models, because a convection-allowing resolution is of great importance to simulate strong convective events (Prat et al. 2021, Fierro et al. 2012, Cassola et al. 2015, Ferrari et al. 2021). During the preliminary test phase, various model simulations have been performed, to better analyze the model dependence on grid resolution and convective scheme used: this analysis is reported in Section 2.3.

For consistency, all the simulations herein make use of the same physics parameterizations, numerics, and domain parameters.

2.2 Input data

The input data from meteorological weather stations comes from two major Italian citizen-science networks, namely Meteonetwork (MNW) and Centro Meteo Lombardo (CML). MNW is a non-profit organization with the task to “promote and disseminate for the benefit

of the community the knowledge of meteorological, climatological, environmental, hydrological and glaciological sciences, and their multiple expressions on the territory" (Giazzi et al. 2022).

CML is a non-profit cultural organization with the aim of studying Lombardy microclimate with real-time weather monitoring of the territory. MNW, at the time of event studied in the present paper, managed over 2000 meteorological weather stations over the Italian territory, and about 800 over the northern Italy territory. CML, at the time of event studied in the present paper, managed over 400 meteorological stations over the Lombardy region and nearby areas. Network model and technology is heterogeneous, and the most used weather stations come from Davis (Vantage Pro 2 and Vue versions), Ecowitt, Froggit, Sainlogic, Oregon Scientific, Bresser, PCE, Irox, and Lacrosse. The weather parameters used in the present study are temperature, relative humidity, surface pressure (measured at the height of 2 m), wind speed, and direction (measured at the height of 10 m).

Both citizen-science networks perform a quality check before the data release, in order to avoid missing or anomalous data. These checks include range tests, to exclude values out of the normal range of that particular variable, and cross-validation checks, to exclude anomalous values with respect to the nearby stations (Giazzi et al. 2022). In addition, a further check is conducted by the WRF data assimilation (WRFDA) software algorithms, since the absorption of the observations is possible only if the difference between the background model field and the observation value is smaller than a predefined threshold.

All the meteorological station data used in the present study have a sub-hourly sample rate. More than one thousand stations located in northern Italy have been considered in the assimilation process.

The other observations used in this study are the lightning data, which have been obtained by ENTLN; at the time of the event studied here, this network consisted of 1800 sensors deployed in more than 100 countries that detect wideband electric field signals emitted by both intracloud (IC) and cloud-to-ground (CG) lightning flashes (Rudlosky 2015, Zhu et al. 2022).

Each lightning detection is associated with a series of parameters used in the classification of the events: lat/lon coordinates of the event, current intensity, timestamp of the event, and type of event (CG — cloud to ground, IC — intra cloud). The medium location error of a generic lightning event is 215 m, which is a sufficient value for the purpose of this work. In the present study, every event (CG and IC) has been considered in the data assimilation process, and only the position and the timestamp parameters were taken into account.

2.3 Case study

The event analyzed in this work is a strong convective storm that interested northern Italy, during late morning and early afternoon of 11 July, 2020, caused by the interaction of cold and dry air currents at 500 hPa in the Po valley with a warm, humid air mass already present in the low levels. The synoptic map at 9 UTC (Fig. 1) depicts the 500-hPa geopotential level, along with 500 hPa temperature (shaded). The Mediterranean region was interested by a flat, high-pressure area; at the same time, a trough from Central Europe to northern Italy caused a geopotential drop and the entrance of dry and cold air in the Po valley. Figure 2 represents the relative humidity field at 500 hPa at 9

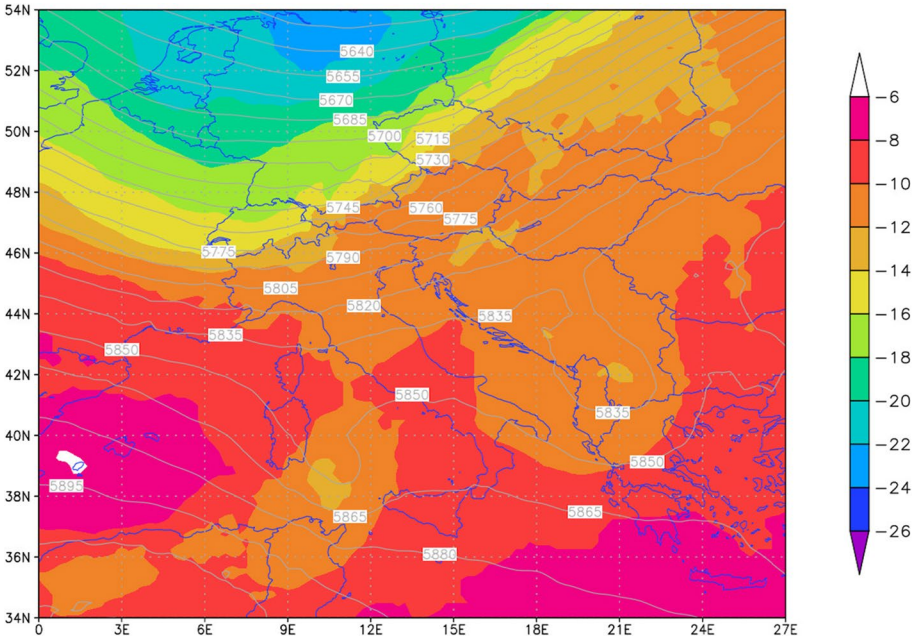

Fig. 1 500 hPa temperature (shaded — °C) — and geopotential height (contour — m) — GFS 0.25° analysis, 09UTC 11 July 2020

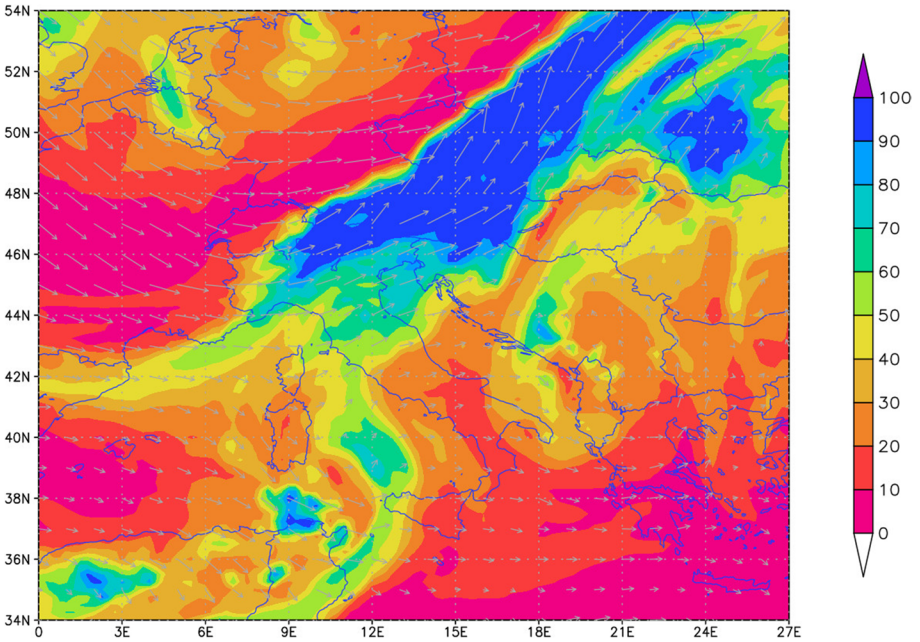

Fig. 2 500 hPa relative humidity field (%), along with wind vectors — GFS 0.25° analysis, 09UTC 11 July 2020

UTC, along with the wind vector at the same level. It can be observed the pre-existing humid air mass in the Po valley, and the dry air mass over Central Europe, crossing the Alps. Wind vectors are rotating from west to south-west direction, due to the cyclonic curvature of the trough. These synoptic conditions triggered the convection activity that developed in the late morning/early afternoon of the same day, starting from the north-eastern part of the Piedmont region, and then propagating with an east/south-east direction, crossing Lombardy and then reaching Trentino-Alto Adige and Emilia-Romagna regions in the afternoon. A sequence of radar images, depicted in Fig. 3, shows the rapid but intense passage of the thunderstorm. The present work will focus on the first phase of the thunderstorm activity, between 11 and 13 UTC, with the passage of the convection activity north of Milan, analyzing the precipitation pattern over this area using different model configurations.

In Fig. 4, it is represented a satellite map of the event, taken at 12 UTC from Eumetsat MSG4 satellite. The image is a high-resolution image in realistic colors created using different spectral channels of the SEVIRI instrument on board MSG4. It clearly shows the development of the convection activity in the north-western part of the Lombardy region. Figure 5 represents the lightning strikes registered by the ENTLN network during 11 July 2020. From the different yellow colors, it is possible to observe that the lightning strikes were concentrated between 11 and 13 UTC.

To better understand the thermo-hygrometric profile of the air mass present in the Po valley, an elaboration of the sounding of Milano Linate station at 12 UTC is depicted in Fig. 6. Along with the real relative humidity field from Linate sounding, the relative humidity profile forecasted by the WRF model for the same point and at the same time is plotted. The model, albeit initialized at 9 UTC, is not able to fully reproduce the humid air mass stationed over the area, especially in the low levels (between 1000 and 800 hPa), and in the mixed-phase region (between 700 and 500 hPa). Ground data assimilation and lightning data assimilation will act on these two key levels, modifying the thermo-hygrometric profile of the air mass.

To put in evidence the response of the numerical simulation to grid resolution and convection parameterization, the data assimilation study has been anticipated by an extensive test of different grid resolutions and convective parameterization schemes. In a 3-km grid resolution environment, both explicit convection and convective parameterization are

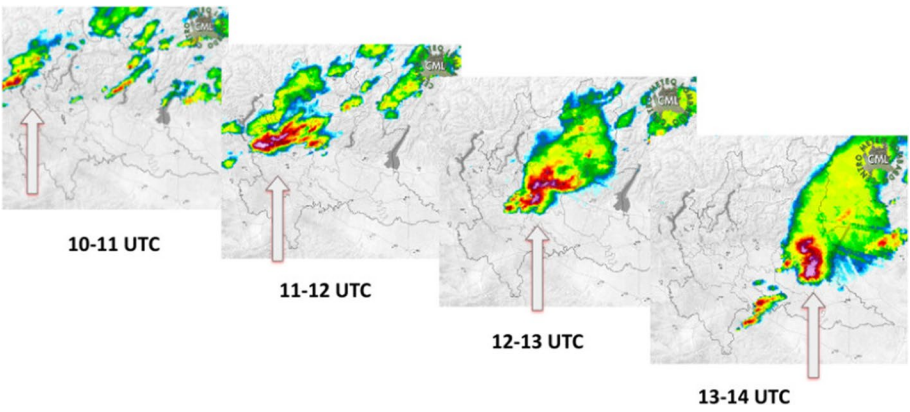

Fig. 3 Radar images of the convective event. Reprocessing from Centro Meteo Lombardo (CML) archive

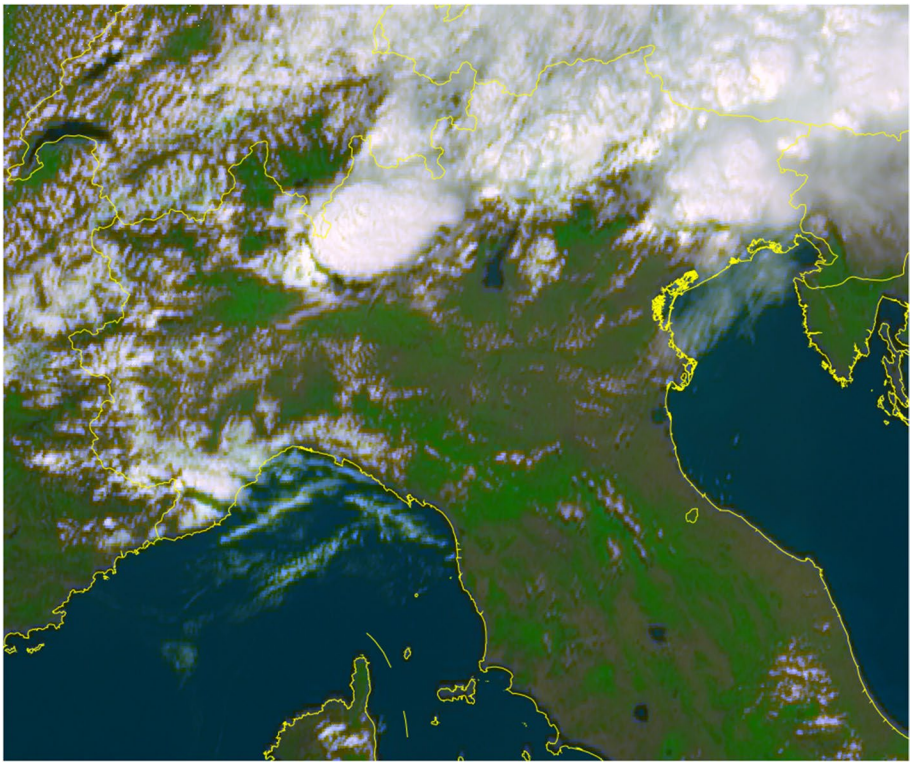

Fig. 4 High-resolution satellite image in realistic colors created using different spectral channels of the SEVIRI instrument on board MSG4 Eumetsat satellite. 12 UTC, 11 July

reasonable. A vast literature has been written on the topic (Prat et al. 2021). In the present preliminary analysis, both convective methodologies, together with various grid resolutions, have been tested, to better investigate the coupling effect of the two configuration parameters. A summary of this analysis is presented in Fig. 7, where hourly rain forecasts for 4 different configurations are plotted, both for two different grid resolutions (3 km and 1 km) and for the activation/deactivation of a convective parameterization scheme. The comparison of the different simulations shows that, in this particular case, neither the grid resolution nor the use of a convective parameterization has a significant impact on the convection development and on the rain forecast amount. No significant changes in the rain pattern or in the rain amount are noticeable, and no convection activity in the northern Milan area is present. The results of this preliminary analysis enforced the indication that the lack of convection activity has to be investigated in the initial and boundary conditions of the global model, with the use of data assimilation algorithms. Moreover, it is reasonable to employ a 3-km grid resolution in the prosecution of the present work.

2.4 Verification methodology

In order to compare the different WRF simulations, a set of statistical indicators has been computed. The precipitation amount for the total event has been computed for a set of

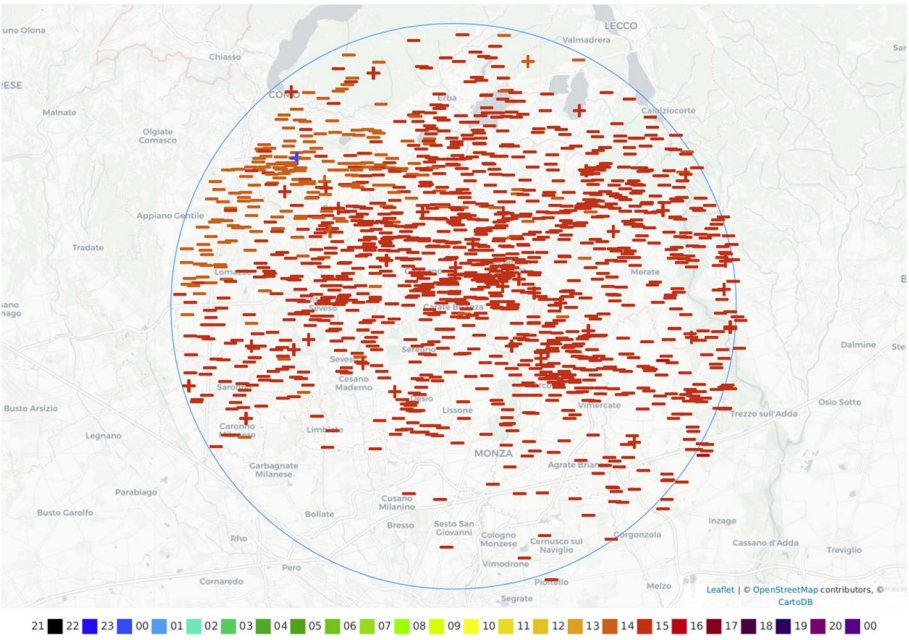

Fig. 5 Lightning strikes registered by the ENTLN network during 11 July 2020. Color table represents hour of occurrence of lightning strikes, which are concentrated between 11 and 13 UTC

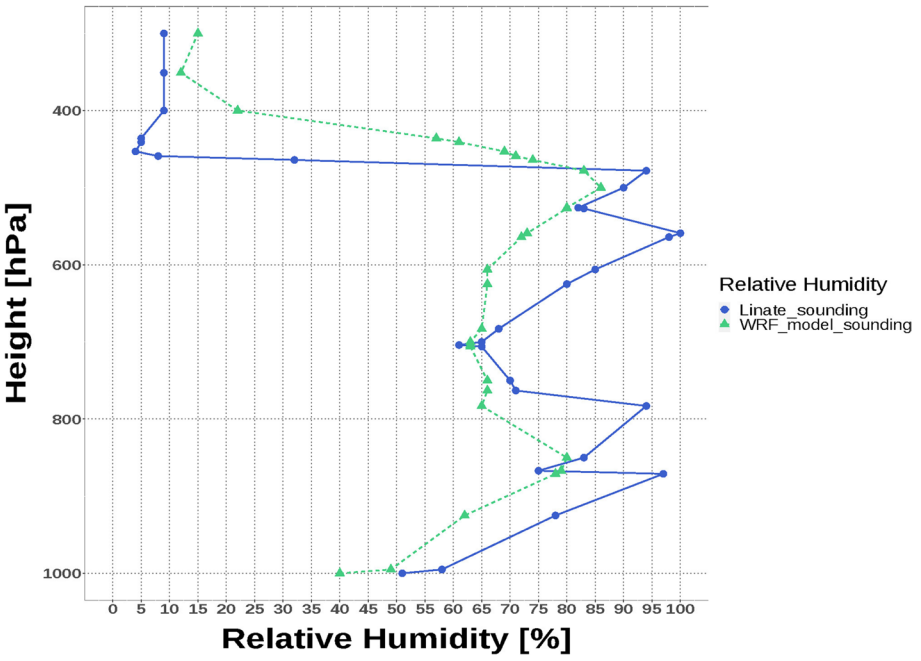

Fig. 6 Milano Linate sounding data — 12UTC, 11 July

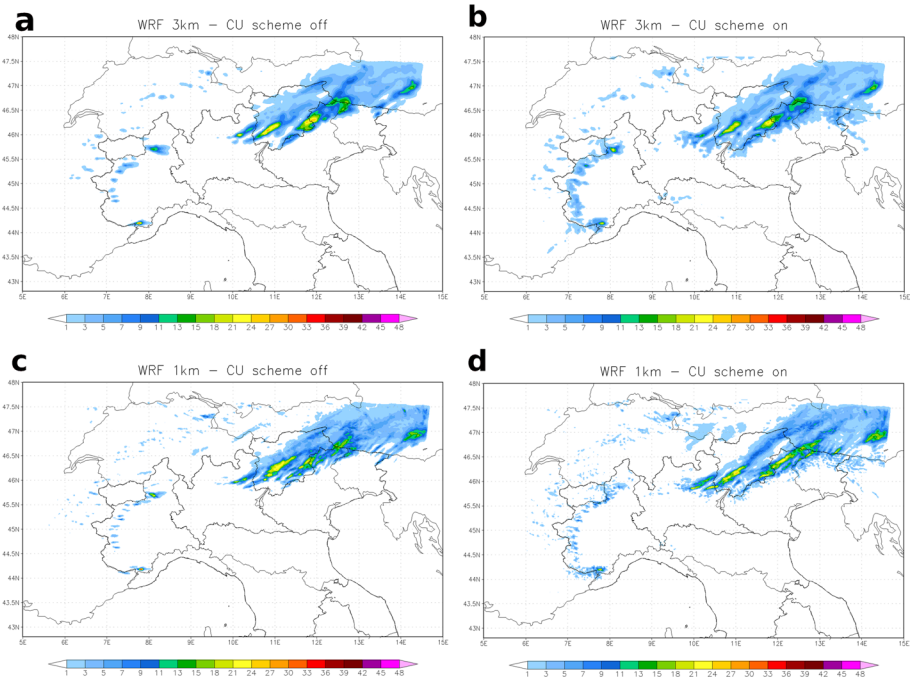

Fig. 7 One hour accumulated rain (mm/h) at 12 UTC for different grid resolution and convection configurations. **a** Three-kilometer grid resolution, explicit convection; **b** 3-km grid resolution, convective scheme activated; **c** 1-km grid resolution, explicit convection; **d** 1-km grid resolution, convective scheme activated

40 different stations in the northern Milan area, which was interested by the convective activity between 11 and 13 UTC. The study area has been chosen for the high density of weather stations, and because of the lack of rain forecasted by WRF model without data assimilation. The stations located in the northern Milan suburbs were extracted from the entire station dataset used for the data assimilation tests, which previously underwent the quality check process detailed in the input data paragraph.

The subset was used to calculate three statistical indicators widely used to verify categorical variables, such as rain occurrence: POD (probability of occurrence), FAR (false alarm ratio), and CSI (Critical Success Index) (Wilks 2011).

These indices were calculated for different rain thresholds and the event was defined as "hit" if the amount of both forecasted and measured rain was above the predetermined threshold value.

For a smaller subset of six stations located in the northern suburbs of Milan, the total precipitation values were also plotted for the simulations with different data assimilation schemes. Such stations have been chosen because of their position, since they cover a wide area interested by the passage of the thunderstorm activity.

Figure 8 shows the positions of the station on a map.

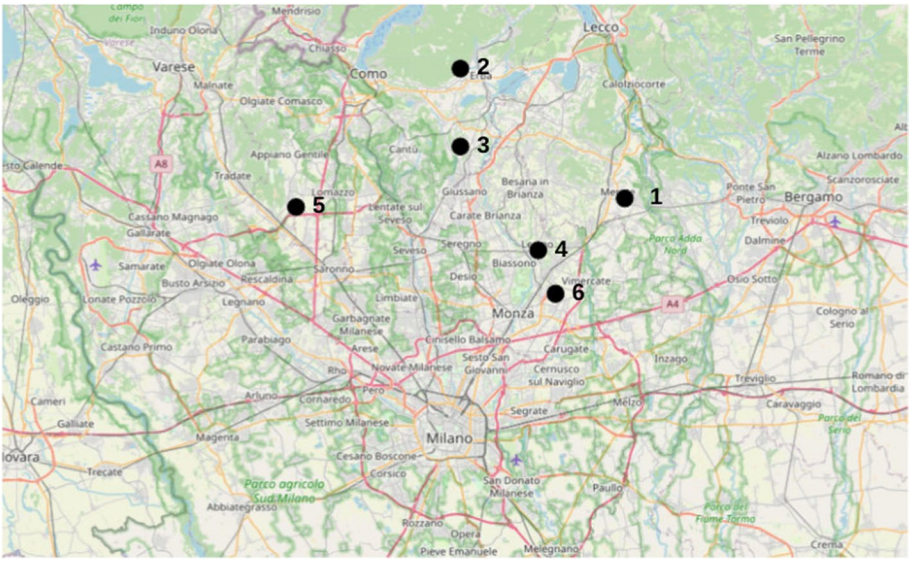

Fig. 8 Six meteorological stations in the northern Milan region used in the verification process and plotted in the verification graphs. (1) Cernusco Lombardone. (2) Erba. (3) Inverigo. (4) Lesmo. (5) Lomazzo. (6) Villasanta

3 Data assimilation methodology

3.1 Configurations of assimilation algorithms

The main focus of this study is the comparison of the output of the WRF model simulations to the observations, applying different assimilation algorithms and different input data. The simulations are conceptually separated into 4 groups:

1# Control simulation;
2#Data assimilation using meteorological weather station as input data;
3#Data assimilation using lightning data as input data;
4#Ideal data assimilation.

The first group is composed of a single simulation, named WRF-ctl, which represents the WRF model simulation without any data assimilation. The WRF model configuration has been detailed in Section 2.

The second group is composed of the WRF simulations with ground data assimilation. Different data assimilation algorithms and configurations have been applied. In particular, the assimilation schemes listed with the codes 3D-VAR, 4D-VAR, 4D-VAR_10min, and 4D-VAR_nolbc in Table 1 have been applied.

The third group is composed of a single simulation, named 3D-VAR-lgt, configured with a 3D-VAR data assimilation scheme with lightning data as the input dataset.

The fourth group is composed of a single simulation, named 3D-VAR_idealrh, configured with a 3D-VAR data assimilation scheme, with an ideal change in water vapor

Table 1 Data assimilation configurations, with algorithm, model start hour, input data, frequency of ingestion of input data, and assimilation window

Code of the assimilation scheme	Data assimilation	WRF start hour	Input data	Obs data frequency (4D-var)	Assimilation interval
WRF-ctl	-	9 UTC	-	-	-
3D-VAR	3DVAR	9 UTC	Meteo stations	-	9 UTC
4D-VAR	4DVAR	9 UTC	Meteo stations	1 h	9–10 UTC
4D-VAR_10min	4DVAR	9 UTC	Meteo stations	10 min	9–10 UTC
4D-VAR_nolbc	4DVAR	10 UTC	Meteo stations	10 min	9–10 UTC — no lbc
3D-VAR-lgt	3DVAR	11 UTC	Lightning data	-	11 UTC
3D-VAR-idealrh	3DVAR	9 UTC	Ideal	-	9 UTC

content of the air mass, and initialized from GFS global model data similarly to the other data assimilation groups.

Table 1 summarizes the main configuration details of the data assimilation tests.

The convection activity started in the Piedmont region of north-west Italy, between 10 and 11 UTC. Therefore, the starting point of the WRF simulations was fixed, for most configurations (WRF-ctl, 3D-VAR, 4D-VAR, 4D-VAR-10 min, and 3D-VAR-idealrh), at 9 UTC. This particular starting point has been chosen since it is very close to the beginning of the convection activity, but at least 1 h before the first lightning activity, considering also the effects of the model spin-up time for cold model start.

For 3D-VAR data assimilation (3D-VAR and 3D-VAR-idealrh), observations have been assimilated into the model exactly at 9 UTC, with a 10-min observation tolerance (before and after 9 UTC), to take into account the fact that not all the observations were updated exactly at 9 UTC.

The 4D-VAR assimilation technique allows more technical configurations to be applied; some differences are detailed hereafter.

Input station data between 9 and 10 UTC are assimilated into 4D-VAR simulation, with 1-h step interval between the observations (and so, one dataset at 9 UTC, and one dataset at 10 UTC). After the data assimilation process, a new model initial state (analysis) at 9 UTC was created, and the model started from this corrected initial state. Together with the initial state, a new updated boundary condition file was created, in order to take into account the observations introduced at 9 and 10 UTC.

4D-VAR_10min is identical to 4D-VAR, but the observations are introduced into the model with a 10-min frequency (09:00 UTC, 09:10 UTC,… 10:00 UTC).

Unlike the previous 4D-VAR data assimilation configurations, 4D-VAR_nolbc does not update the lateral boundary condition file. The new analysis state is not produced at 09 UTC, but at 10 UTC, and the WRF simulation starts from this new starting point, which is closest to the convection activity, but presents also a shorter spin-up time.

3D-VAR-lgt simulation presents some different and peculiar characteristics as compared to the previous one. The 3D-VAR-lgt input dataset is dependent from the lightning activity and it cannot be as flexible as the assimilation of ground station data in terms of starting point and assimilation interval. Since the first lightning activity has been detected between 10 and 11 UTC, the lightning data has been assimilated into the model at 11 UTC, determining a starting time for the simulation at that time. To take into account the initial stage

of the thunderstorm development, every lightning strike in the 30-min interval before 11 UTC has been used in the 3D-VAR-lgt data assimilation.

This detailed description of assimilation intervals and simulation starting points permits to show one main difference between the assimilation of ground station data and lightning data: in order to have some input data, the assimilation of lightning strikes needs to be done after the initial development stages of the convection activity. In this study, 3D-VAR-lgt simulation starts in the middle of the convection activity, giving practically no spin-up time to the forecast of rain in the next 2 h.

To better understand this aspect, an ideal simulation (3D-VAR-idealrh) has been studied, in which the change in the upper-level humidity values has been included at 09 UTC. In this particular analysis, humidity values have been increased not only in the mixed-phase region, but also in the lower region, between 950 and 800 hPa, considering that the analysis of the sounding data suggested a lack of specific humidity at those levels (Fig. 6). 3D-VAR-idealrh simulation is an ideal simulation, and it cannot be replicated in a real-time operational activity for two reasons: it makes use of the lightning activity 2 h before the lightnings themselves appear; it uses a suggestion derived from the analysis of the sounding data, available after about 4 h. Instead, the other data assimilation configurations are potentially applicable in real operations, since they make use of station data before the development of convection activity.

3.2 Lightning assimilation procedure

The procedure used to assimilate the lightning data in the WRF model is a re-adjustment of the algorithm proposed by Fierro et al. (2012). Lightning flashes that occur in a predetermined time interval centered on the assimilation time in a given area are associated with the nearest model grid point. At that point, water vapor is added to the mixed phase region of that atmospheric column, via an hyperbolic tangent equation (the mixed phase region is the column interval between the 0 and − 20 °C isotherms).

$$HR = HR_0 + A\tanh(BX) \tag{1}$$

Formula 1 is a variation of the equation proposed by Fierro et al. (2012). In Fierro's formula, the water vapor increase is calculated starting from the water vapor saturation mixing ratio and from graupel mixing ratio, and then the result is converted into relative humidity before entering the data assimilation process. In the present work, the formula has been simplified with the direct use of the relative humidity, which is increased using a direct dependence with the number of lightning flashes. Additionally, Fierro's algorithm is applied only if the relative humidity field in the corresponding grid point is below a predefined threshold: Formula 1 does not take into account this aspect, and is not intended to have a general validity: the reason behind this particular choice is the attempt to easily increase water vapor into the mixed phase region for this particular case study.

HR_0 is the relative humidity extracted from the model analysis, at the assimilation time. X is the number of lightning flashes associated with the model grid point. HR is the relative humidity after the enhancement algorithm. A and B are parameters that regulate the relative humidity increase in dependence on the number of flashes. The value for parameter A has been fixed to 30, in order to set-up a maximum increase of relative humidity of 30%. The parameter B has been fixed to 0.15, so to obtain a relative humidity increase close to the maximum value for 15 flashes or more per unit time. The hyperbolic tangent function

guarantees a rapid increase in the humidity values for a small number of lightning flashes, rapidly reaching the asymptotic value.

The use of relative humidity as a dependent value of the algorithm permits to directly assimilate this variable into the WRF model.

To better visualize the actual change in the atmospheric column during the data assimilation of the lightning flashes, a histogram of the relative humidity values for a series of atmospheric levels, before and after data assimilation, is reported in Fig. 9. The point analyzed is at the core of the initial stage of the convection activity, with coordinates 46.80N, 8.40E. It is clear from the plot that the vertical space between 500 and 600 hPa has been subjected to an increase between 25 and 30% of relative humidity, reaching the maximum value of 100%. It is also clear from the plot that the other vertical levels have remained unchanged by the assimilation procedure, because only the levels in the mixed phase region are impacted.

3.3 Tuning analysis

In the WRFDA suite, weather station input data are assimilated into the model, and the assimilation features can be modified operating on several parameters.

As a first step, the calculation of the background error has been computed, to avoid the use of a generic background error (Descombes et al. 2015). The calculation of the background error has been made by comparing a series of previous WRF model forecasts, with the same geographical domain and the same parameterization, simulated using the real data 12/24 h before the simulation object of this study. In particular, the background error matrix has been calculated using cv option number 5 of the WRFDA suite, using a 36-h-long period for the matrix computation. The reason behind this value was the purpose of including a sufficient number of model cycles to produce a consistent value of the matrix, without including time intervals excessively distant from the case study event. The control simulations were compared against the correspondent analysis, in order to calculate the model error just before the data assimilation period, which is used as the background error during the data assimilation process.

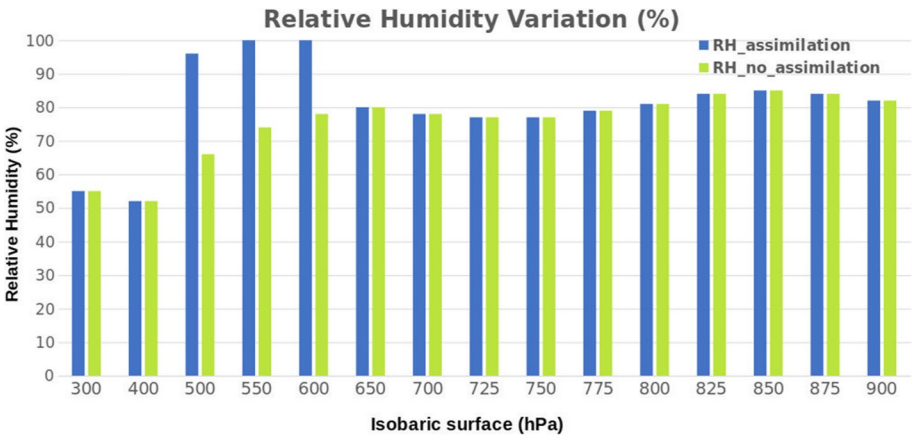

Fig. 9 Comparison between relative humidity field at different isobaric levels, before and after the lightning data assimilation, for an atmospheric column. Only the isobaric levels in the mixed phase region exhibit an increase in the relative humidity field

A second step involves the tuning of two scaling parameters in the WRFDA namelist, var_scaling and len_scaling parameters: var_scaling controls the magnitude of the perturbation introduced by the observations into the model grid, whereas len_scaling controls the spatial extension of the perturbation introduced by the observations. Various tests and simulations have been conducted, assimilating only one station at a time, in order to analyze the particular effect on the background field. The same procedure has been repeated for every input parameter that entered the assimilation process (temperature, relative humidity, wind speed, and direction). The selection of the best scaling parameter has been conducted with the aim of obtaining a spatial effect for a single observation point of about 20 km. This distance was chosen on the basis of the density of meteorological stations in northern Italy: even in the areas with the lowest station density, the minimum distance among stations is comparable or even less than that value.

A similar argument is behind the selection of the scaling parameter for the lightning data assimilation: in this case, a len_scaling which permits a perturbation of the size of a convective cell (about 20 km) has been chosen.

After a sensibility analysis, a len_scaling of 0.5 and a var_scaling of 1 were chosen for the scaling parameters of the data assimilation tests. Figure 10 represents the perturbation of the relative humidity field at 500 hPa, caused by the assimilation of a single lighting strike located at a test point of coordinates 45.4N-9.1E.

Figure 11 represents the true background perturbation caused by the assimilation of the lightning activity at 11 UTC, located in the eastern part of the Piedmont region.

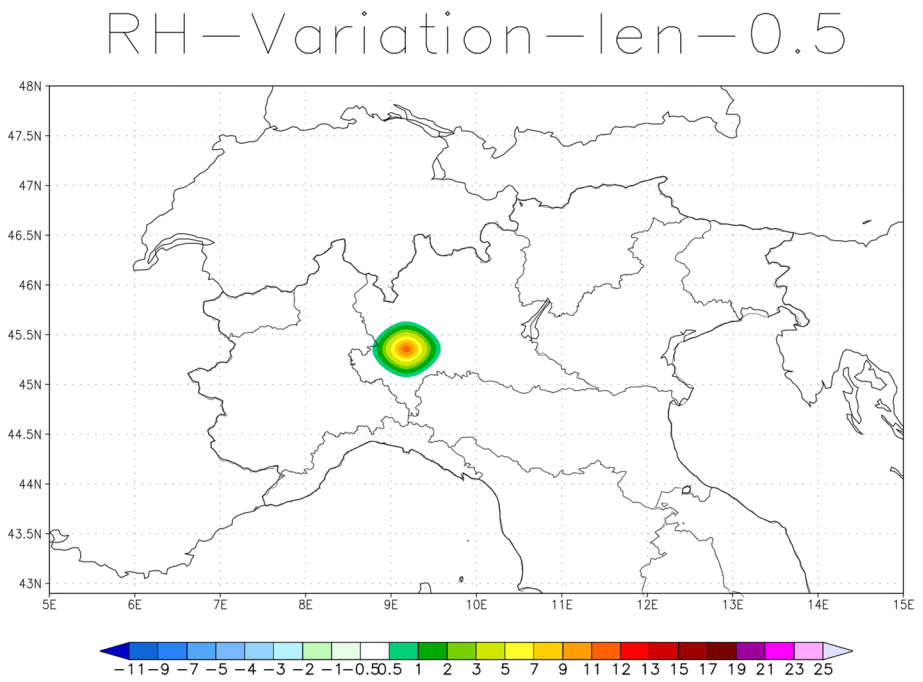

Fig. 10 Tuning analysis. Relative humidity analysis difference after the assimilation of a single lightning strike. RH field at 500 hPa (%)

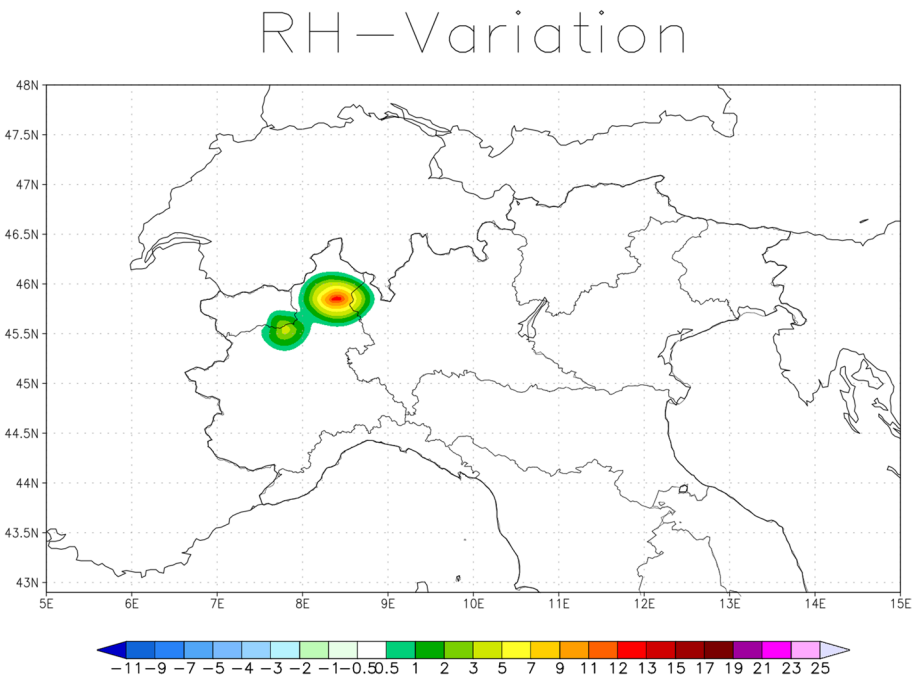

Fig. 11 Relative humidity analysis difference at 500 hPa after the assimilation of the lightning data (%) — 11 UTC

This area corresponds to the initial convective area of the thunderstorm, which is also visible from the first radar snap-shot in Fig. 3.

4 Results and discussions

Figure 12 reports a composition of the five main simulations, representing 1-h accumulated precipitation for the domain simulation over northern Italy. To focus the attention on the northern Milan region, a circle has been drawn on the control map, and the hourly precipitation between 11 and 12 UTC has been represented. It is evident from the map that the control simulation presents a lot of convection activity in the Alpine region, but no active precipitation areas in the northern Milan region. Otherwise, 3D-VAR, 4D-VAR, and 4D-VAR_10min simulations show a presence of convection activity in the area of interest. 3D-VAR shows a more intense but geographically limited area of convection, whereas 4D-VAR and 4D-VAR_10min present a larger area interested by precipitation. 4D-VAR_nolbc shows little differences with the control simulation, apart from a slight increase of the precipitation in the eastern part of the northern Milan area. It is worthwhile to mention that, apart from the 4D-VAR_nolbc simulation, all the data assimilation cases show an increase of the convection activity and the related precipitation, not only in the area of interest, but also in other areas, e.g., in the alpine region, where convection activity was already present in the control simulation, notably the eastern part of Lombardy, Trentino-Alto Adige, and Veneto regions. Although the data assimilation maps show the presence of

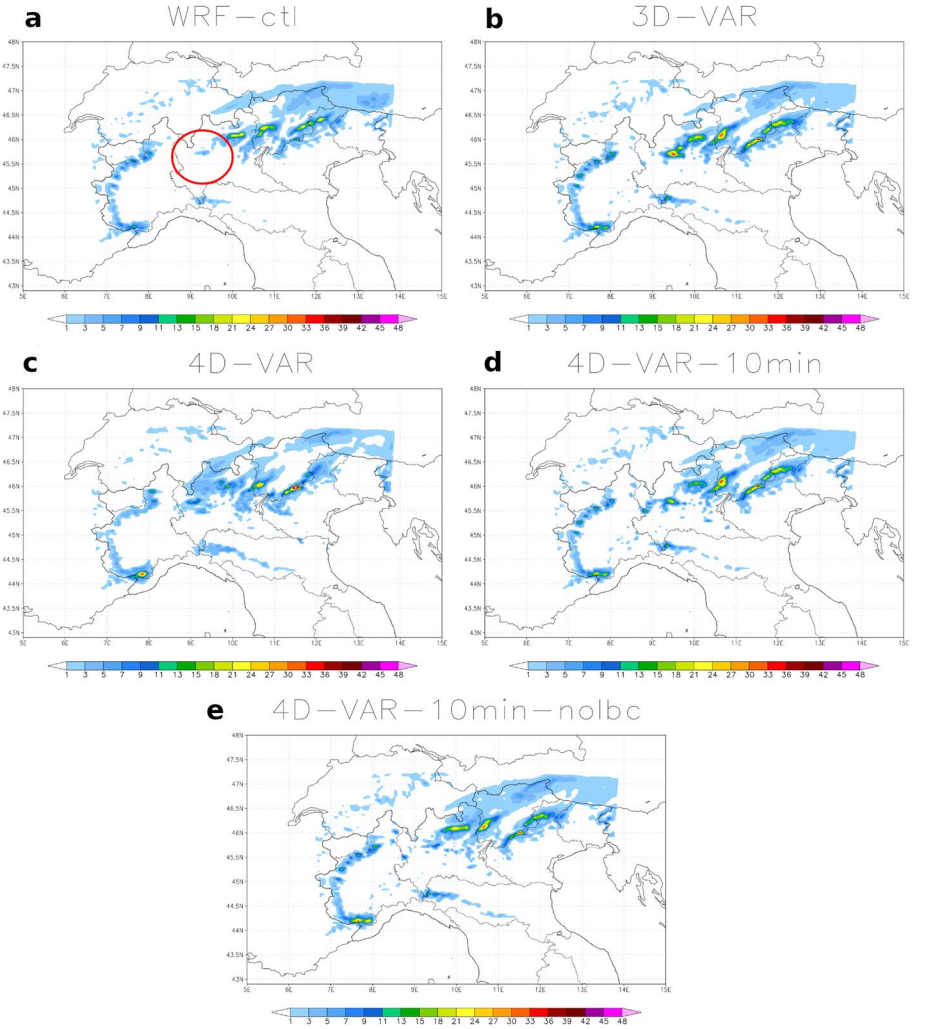

Fig. 12 One hour accumulated rain (mm/h) at 12 UTC from different data assimilation configurations. **a** WRF model without data assimilation. **b** 3D-var assimilation with synoptic stations as input data. **c** 4D-var assimilation with 1-h assimilation window. **d** 4D-var assimilation with 10-min input data frequency. **e** 4D-var assimilation without lateral boundary condition update. Rain accumulated from 11 to 12 UTC

convection activity in the northern Milan region, the intensity is not comparable to the real radar representation of the event.

A more quantitative comparison among simulations has been presented in Fig. 13. Total precipitation of the convective event is shown for the six meteorological stations under investigation, along with the five WRF simulations analyzed. First, the comparison between real station amount (observed) and WRF-ctl shows the lack of convection activity forecasted by the control simulation, except for a small amount in the Inverigo station. 3D-VAR, 4D-VAR, and 4D-VAR_10min exhibit an increased amount of precipitation with respect to the WRF-ctl. This behavior is present for every station analyzed, and, generally,

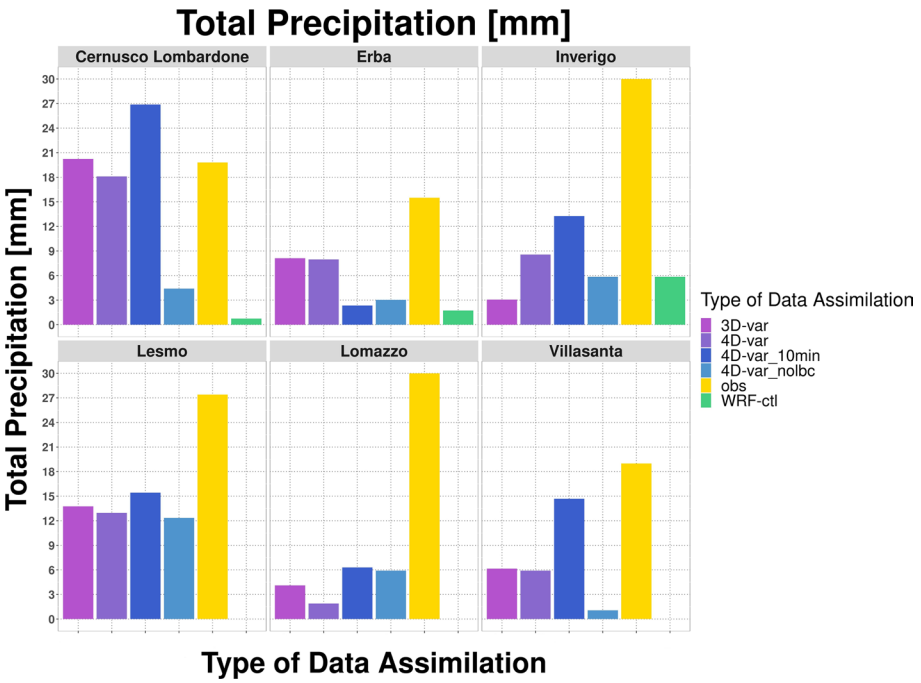

Fig. 13 Total precipitation amounts of the event for the 6 stations under investigation. In addition to the observation data (obs) and the WRF model simulation without data assimilation (WRF-ctl), 4 different data assimilation configurations are plotted: 3D-var, 4D-var, 4D-var_10min, and 4D-var_nolbc. Observation values are between 15 and 30 mm for the total event

4D-VAR_10min shows a greater amount of precipitation than 3D-VAR and 4D-VAR. The total precipitation values of these simulations do not reach the corresponding values at the ground stations, except for Cernusco Lombardone, where the observed and forecasted values are very similar. More pronounced differences are present in the stations at the edge of the area of interest, e.g., Lomazzo in the west, and Lesmo in the south-east. 4D-VAR_nolbc shows an intermediate behavior, given that only for some stations the precipitation is comparable with the other data assimilation simulations. Overall, WRF simulations with data assimilation show an increased ability in the convection representation, with respect to the control simulations, but they do not reach the total amount observed at the ground stations.

Figure 14 represents a plot of 3 different categorical indices (POD, FAR, and CSI), calculated for the 40 stations included in the area of interest, and for the entire duration of the convective event. The indices have been calculated for different rain thresholds, from 1 to 33 mm (maximum rain amount), with intervals every 4 mm, in order to analyze the ability of the forecast of different rain amounts. The probability of detection plot presents high values in the lower range of precipitation amount (less than 5 mm), with a generalized decrease of POD values as the total precipitation increases. For precipitation amounts greater than 15 mm, POD values are below 0.1 for every simulation. This is coherent with the previous analysis: no simulation was able to forecast the true amount of rain during the entire event, which explains the poor POD values for large rain amounts. Among the different WRF simulations, WRF-ctl exhibits the lowest performance even for the smallest rain thresholds, with a maximum value of 0.2, and rapidly decreasing toward zero. WRF

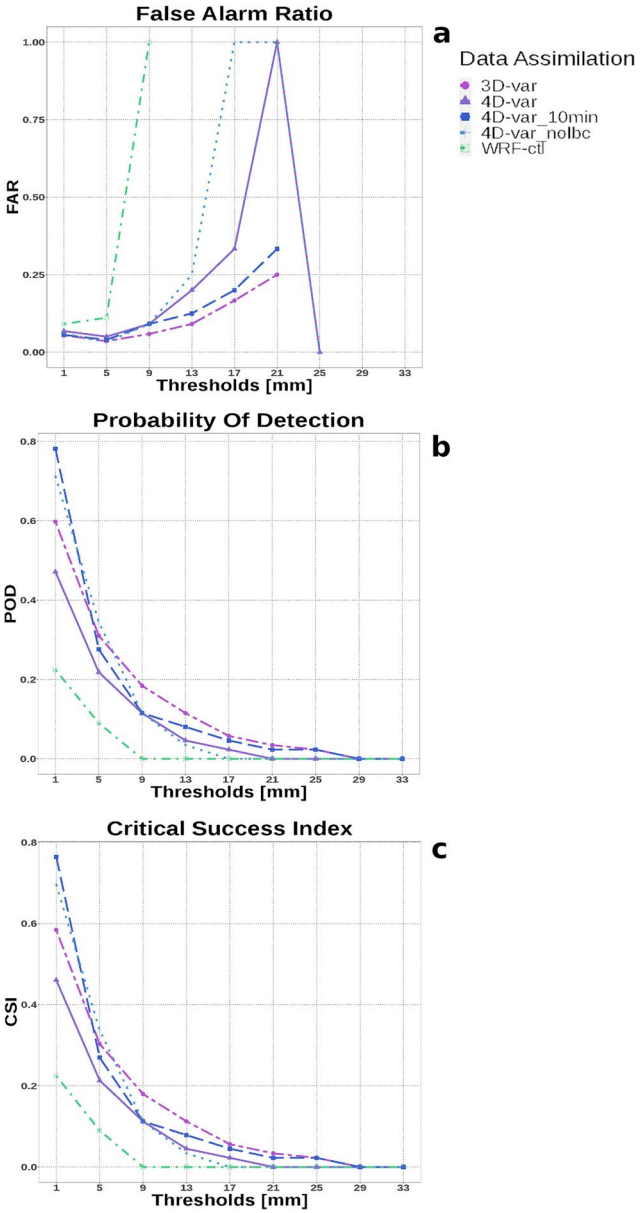

Fig. 14 Categorical indices calculated using 40 meteorological stations in the northern Milan area, with the rain amounts of the total event. Data are taken from WRF model without data assimilation (WRF-ctl), WRF model 3D-var using ground station data as input data (3D-VAR), WRF model 4D-var using 1-h assimilation window (4D-var), WRF model 4D-var using 10-min obs frequency (4D-var-10 min), and from WRF model 4D-var without lateral boundary condition update (4D-var_nolbc); **a** false alarm ratio, **b** probability of detection, **c** critical success index

simulations with data assimilation overall exhibit better performances as compared with WRF-ctl. 4D-VAR-10 min and 4D-VAR-nolbc show a wide area of precipitation, so that POD values are relatively great at the lowest rain thresholds: in fact, if a rainy event is

spread over a wide area, the rain event is correctly forecasted at a great number of stations if the chosen rain threshold is low; on the contrary, the rain spread results in only a few stations correctly forecasted if the rain threshold is high. 3D-VAR and 4D-VAR have smaller POD values at low rain threshold, but greater values at medium rain threshold (especially 3D-VAR). Overall, 3D-VAR and 4D-VAR-10 min exhibit the better performance.

The FAR plot shows that, for the majority of rain thresholds, values are very low, less than 0.3. This can be easily explained with the rain underestimation shown in the total precipitation plot: it is more difficult to give a false alarm when the rain values are generally underestimated. Some points in the plot present the unusual very high value of 1: this can also be explained with a slight displacement of the rainy area that determines the presence of a simulated peak at the margin of the real precipitation area, resulting in a false alarm. WRF simulations with data assimilation do not present different trends among them.

The Threat Score Index (CSI) presents a behavior comparable to the POD index. The reason is that the CSI is a composition of POD and FAR indices, and, since the FAR index presents very low values, CSI is very close to the POD index. The POD plot interpretation is also valid for the CSI plot interpretation.

In Fig. 15, a map comparison between 1 h accumulated rain is presented. On the left, there is the map of 1-h precipitation between 11 and 12 UTC for 3D-VAR simulation, on the right, the same map but for 3D-VAR-lgt simulation. It is clear from the precipitation map that the 3D-VAR-lgt simulation is not able to correctly reproduce the convection, not only in the area of interest, but also diffusely on the Alpine region. Precipitation amount is lower, and in some areas precipitation is not represented. Moreover, Fig. 16 shows a plot of the amount of precipitation for the whole event and for the same locations shown in Fig. 13. The comparison clearly demonstrates that 3D-var-lgt is not able to recreate the convection activity in the northern Milan area, presenting only weak precipitations in a few stations. On the contrary, 3D-VAR outperforms 3D-VAR-lgt both with regard to the localization of the events and with regard to the amount of the precipitation. For this particular case, 3D-VAR-lgt shows only a little improvement on the WRF-ctl simulation.

A similar conclusion is inferred from Fig. 17, which represents the CSI for 3D-VAR and 3D-VAR-lgt simulations. 3D-VAR outperforms 3D-VAR-lgt for every rain threshold, especially in the highest range of precipitation, since 3D-VAR-lgt maximum value is 10 mm.

A possible explanation of the poor 3D-VAR-lgt behavior could be the very short spin-up period for this simulation: since the convection activity started at about 10 UTC in

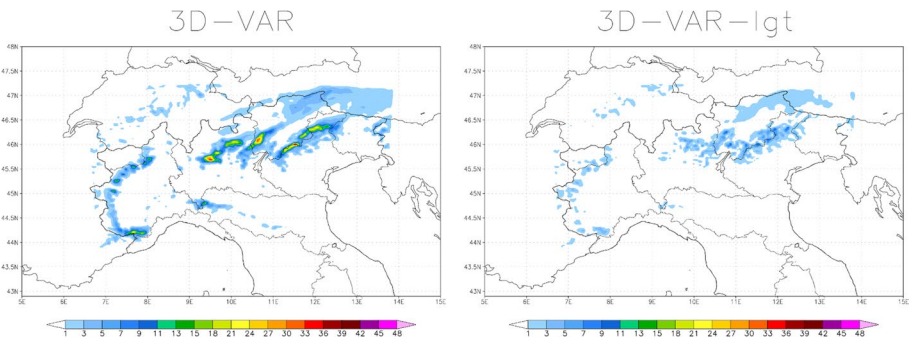

Fig. 15 One hour accumulated rain between 11 and 12 UTC. On the left: WRF model with 3D-var data assimilation, using synoptic meteorological stations as input data (3D-VAR). On the right: WRF model with 3D-var data assimilation, using lightning data as input data (3D-VAR-lgt)

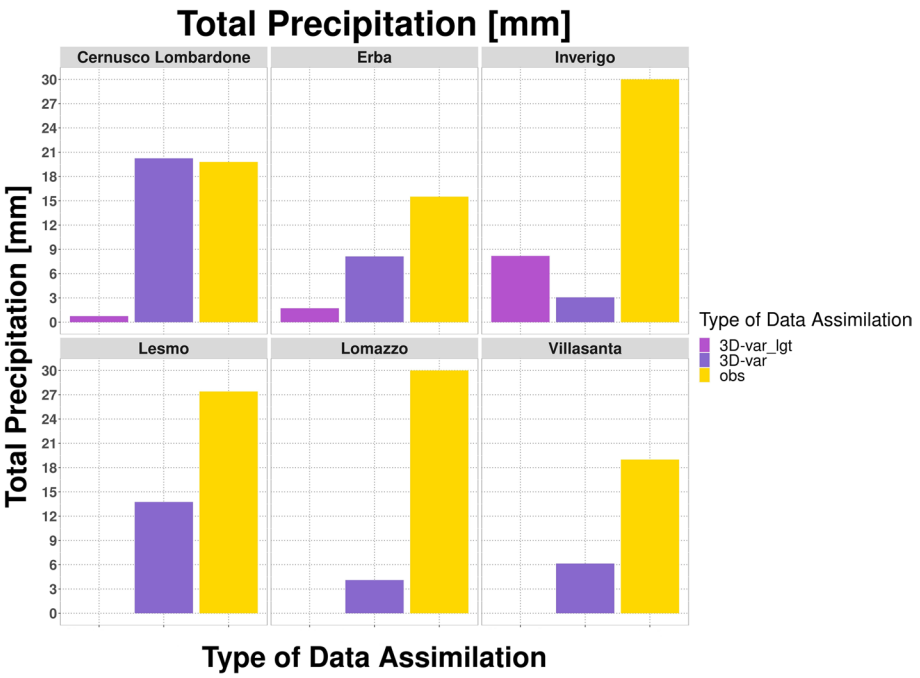

Fig. 16 Total precipitation amounts of the event for the 6 stations under investigation. In addition to the observation data (obs), 2 different data assimilation configurations are plotted: WRF model 3D-var data assimilation with synoptic station as input data (3D-var); WRF model 3D-var data assimilation with lightning data as input data (3D-var_lgt)

the eastern part of the Piedmont region, the first lightning activity developed between 10 and 11 UTC. This lightning activity has been assimilated at 11 UTC, and the 3D-VAR-lgt simulation started at the same time. Since most of the precipitation amount was generated in the region of interest between 11 and 13 UTC, it is evident that the spin-up time was almost negligible, and the simulation started with the convection activity already in place. This feature is mostly responsible for the poor performance of that model. The short duration of the thunderstorm activity is another important factor, since lightning data assimilation has a better performance for longer convective events, in which the data assimilation of the first lightning strikes has a sufficient time to produce a successful result in the later stages of the thunderstorm activity (Federico et al. 2021). Moreover, the humidity changes in the mixed-phase region could not have been sufficiently high and widespread to start and maintain a convection activity as in reality. A possible explanation of this behavior could be the fact that during the summer, the 0 °C and − 20 °C isotherms are very high, and so the increase in water vapor happens at very high altitude, limiting the possibility of maintaining a robust convection process. This is also underlined by the comparison with 3D-VAR-idealrh, which does not suffer from this issue.

As already pointed out, the present work aims at studying data assimilation from a cold-start point of view, prioritizing the use of the latest GFS model emission compared with a more complex system of model restart through the warm-start process, using more elaborated sets of forecast-data assimilation cycles (Federico et al. 2019, Prat et al. 2021). And so it is important to notice that, even in this particular case study, 3D-VAR-lgt achieves a

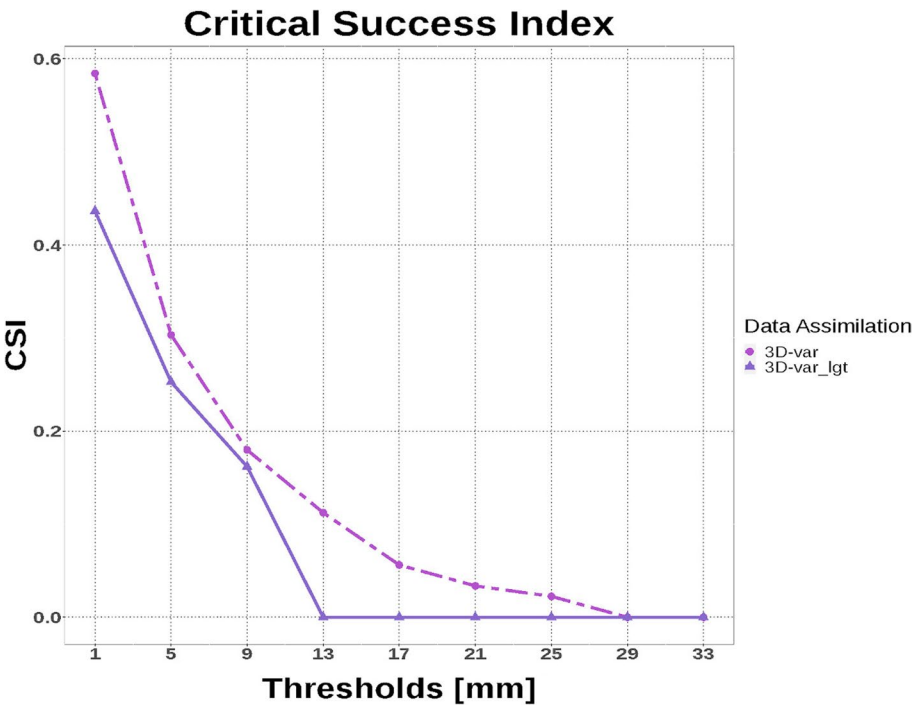

Fig. 17 CSI calculated using 40 meteorological stations in the northern Milan area, with the rain amounts of the total event. Data are taken from WRF model data assimilation using ground station data as input data (3D-VAR) and from WRF model data assimilation using lightning data as input data (3D-VAR_lgt)

better performance than WRF-ctl, although not comparable to the other data assimilation simulations which present a more pronounced spin-up time.

To better understand these aspects, spin-up time and relative humidity increase only in the mixed-phase region, a precipitation map between 11 and 12 UTC for 3D-VAR-idealrh is proposed in Fig. 18.

3D-VAR-idealrh clearly outperforms every other data assimilation configuration in the forecast of the convection activity in the northern Milan area. The shape of the area interested by the heavy rain event is very similar to the radar image of Fig. 3 between 11 and 12 UTC, and the precipitation amount is clearly much larger than the previous simulations, reaching values above 20 mm in 1 h. Apart from the eastern part of the Piedmont region, and the western part of the Lombardy region, the other areas have remained almost unchanged by the data assimilation process: that is reasonable, since the relative humidity change has been operated only in the eastern part of the Piedmont region, replicating the humidity increase in the areas interested by 3D-VAR-lgt.

Figure 19 reports the total precipitation amount of the whole event, with a comparison between 3D-VAR-idealrh and the control simulation without data assimilation. 3D-VAR-idealrh shows precipitation amounts very similar to the observed values in every point of the validation set, differently from the other data assimilation simulations, which provided far less total precipitation values. In some cases (Cernusco Lombardone, Lesmo, and Villasanta), 3D-VAR-idealrh precipitation amounts even exceed the observation values. This

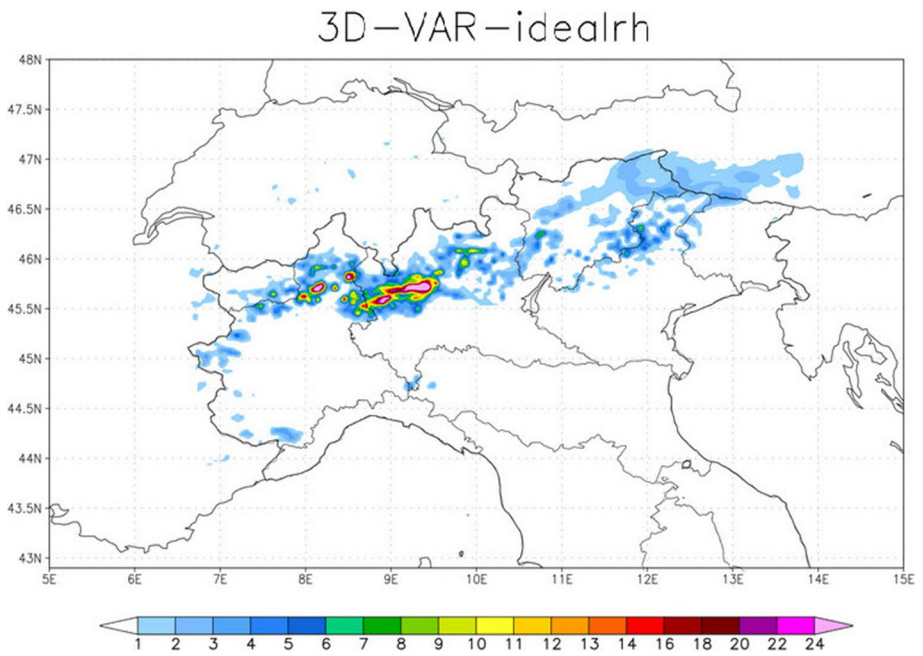

Fig. 18 One hour accumulated rain between 11 and 12 UTC. WRF model with 3D-var data assimilation, using atmospheric column ideal data as input data (3D-VAR-idealrh). Data assimilation produced a forcing of the relative humidity field over the entire column before the beginning of the convection

could be due to the heavy water vapor increase in the atmospheric column, not only in the mixed-phase region, but also in the lower atmospheric levels; this consideration is inferred from the Linate sounding data, and it has to be considered a limit case of a hypothetical water vapor ingestion in the atmospheric column.

5 Conclusions

A strong convective event located in northern Italy has been accurately studied and simulated through the use of the WRF model, at a convective-permitting resolution. Various data assimilation algorithms have been tested, and two different data sources have been used (weather stations and lightning data). Promising results have been obtained through the use of the 3D-VAR algorithm with weather station data. This could be due to the increase in water vapor amount near the ground just before the start of the convection. The precipitation area is better defined, and there is a generalized raise in precipitation amount in the northern Milan area. A slightly better result is obtained using the 4D-VAR algorithm, and a further enhancement of precipitation amount is found. This might be due to the large number of weather station data ingested into the data assimilation with the 4D-VAR algorithm, together with the propagation of analysis increment over the whole temporal assimilation window, and not only in a fixed instant in time, like in the 3D-VAR algorithm. This is especially obtained with the model configuration which permits the update of the lateral

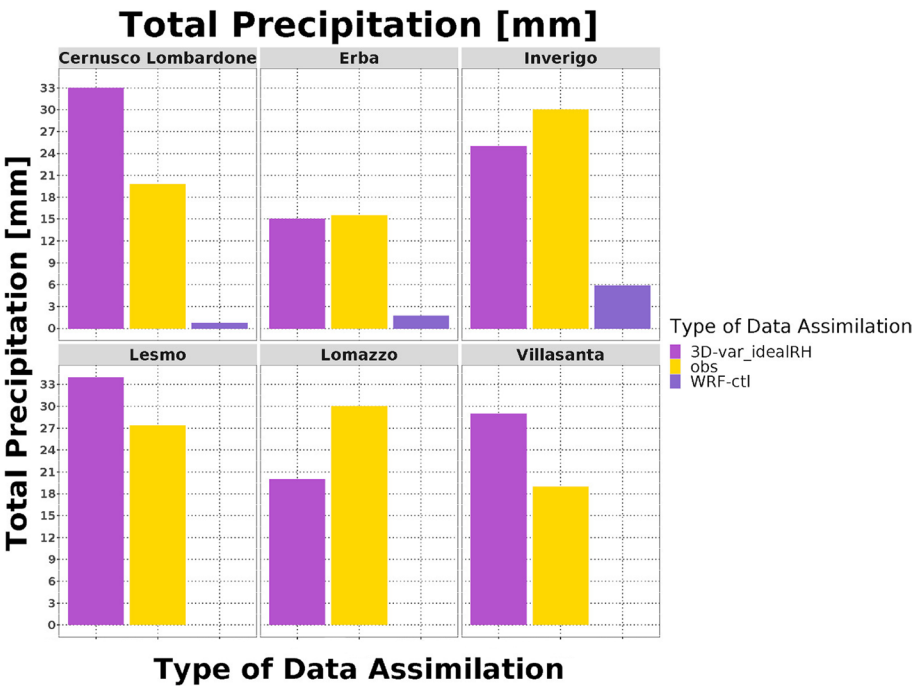

Fig. 19 Total precipitation amounts of the event for the 6 stations under investigation. In addition to the observation data (obs), 2 different WRF configurations are plotted: WRF model without data assimilation (WRF-ctl); WRF model 3D-var data assimilation using atmospheric column ideal data as input data (3D-var_idealrh)

boundary condition during the assimilation window, and draws attention to the improvement of the forecast using weather station data.

Lightning data, assimilated into the model adding water vapor in the mixed-phase region, did not produce significant results, generating only a little increase of precipitation compared to the WRF-ctl simulation. In fact, in a real operational environment, the availability of lightning data is subordinated to the convection activity already in place, and so the assimilation of this type of data is not really able to increase the precipitation effects in the subsequent temporal intervals. This fact confirms the hypothesis that lightning data, as well as rain observations, could be more valuable in extended and stationary convection systems, in which data assimilation could improve the central and final stages of the convective development. It should be emphasized that this research is based on WRF model cold-start simulation, which has the property to increase the problems related to spin-up features. This is true especially for lightning data assimilation, because the spin-up time is even shorter. Differences in the behavior of data assimilation have been found comparing our research with some recent works proposed in the literature (Federico et al. 2019, Mazzarella et al. 2021), especially in the lower improvement of rain forecast with lightning data assimilation. Again, we think that this could be caused by the lack of warm-start configuration, with more forecast/data assimilation cycles before the real forecast, and by the particular case study event, in which no evidence of the starting of convection activity was present using only the global model initial condition. Data from Linate sounding

also suggested that the thermo-hygrometric forecast in the Po valley suffered from a lack of water vapor not only in the mixed-phase region, but also in the lower region between 950 and 800 hPa, which may play an important role in the thunderstorm development. To support this indication, an ideal simulation, with the increase of water vapor in these two key levels, was also performed. This simulation showed an increased ability in the forecast of precipitation localization and amount, producing a result very similar to the radar images of the considered event. The modification of the tropospheric water vapor profile, provided with a sufficient advance (1–2 h) with respect to the beginning of the convection activity, has been proved to be, in this case study, an important element which drastically improved the forecast (Román-Cascón et al. 2016, Wang et al. 2012). This work also suggests the possibility to explore other input data sources, such as lidar sounders or geostationary satellite atmospheric sounders, which may provide valuable information about the atmospheric thermo-hygrometric profile sufficiently in advance of the start of the convection. More research has to be conducted in the exploration of the warm-start configuration, to realize the importance of this key aspect, especially for lightning data assimilation.

Acknowledgements The authors would like to thank "Associazione Meteonetwork OdV" and "Associazione culturale Centro Meteo Lombardo" for providing the near-ground weather data, and Earth Networks, Inc., for providing the lightning data. A special thanks to Centro Meteo Expert, for providing the HPC infrastructure used for the model simulations. Francesco Ferrari & Mauro Giudici contributed to this research within the "project ENDAS - Enhancement of data assimilation and data driven modeling to improve the meteorological predictions at different space and time scales", financed through the PON "Ricerca e Innovazione" 2014-2020, Action IV.6 "Research contracts on Green themes", and in the framework of the project "Geosciences for society: resources and their evolution" supported by the Italian Ministry of University and Research (MUR) through the funds 'Dipartimenti di Eccellenza 2023/2027'.

Author contribution All authors contributed to the study design and to the simulation planning. E.C.M., M.G., and T.M. conducted the WRF simulations and data assimilations. D.Z., M.G., R.S, A.P., F.S., F.F., and A.B. contributed to the data analysis. The first draft of the manuscript was written by E.C.M, R.S., and M.G., and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The authors have no rights to provide experimental data.

Declarations

Conflict of interest The authors declare no competing interests.

References

Avolio E, Federico S, Sempreviva A, Calidonna C, Leo L, Bellecci C (2011) Preliminary meteorological results of a four-dimensional data assimilation technique in southern Italy. Atmos Clim Sci 1(3):134–141. https://doi.org/10.4236/acs.2011.13015

Barker D, Huang W, Guo Y, Bourgeois A (2003) A Three-dimensional Variational (3DVAR) Data Assimilation System for Use With MM5 (No. NCAR/TN-453+STR). University Corporation for Atmospheric Research, 73 pp. https://doi.org/10.5065/D6CF9N1J

Barker D, Huang W, Guo Y, Bourgeois A, Xiao A (2004) A three-dimensional variational data assimilation system for MM5: Implementation and initial results. Mon Weather Rev 132:897–914. https://doi.org/10.1175/1520-0493(2004)132

Benjamin SG et al (2004) An hourly assimilation–forecast cycle: The RUC. Mon Wea Rev 132:495–518. https://doi.org/10.1175/1520-0493(2004)132%3c0495:AHACTR%3e2.0.CO;2

Betts AK, Miller MJ (1993) The Betts-Miller scheme. Chapter 9 in "The Representation of Cumulus Convection in Numerical Models of the Atmosphere". (Eds. K.A. Emanuel and D.J. Raymond.). Amer Meteor Soc Meteor Mon 24(46):107–121

Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH (1992) GPS Meteorology: Remote sensing of atmospheric water vapor using the global positioning system. J Geophys Res 97:15787. https://doi.org/10.1029/92JD01517

Cassola F, Ferrari F, Mazzino A (2015) Numerical simulations of mediterranean heavy precipitation events with the WRF model: A verification exercise using different approaches. Atmos Res 164–165:3–18. https://doi.org/10.1016/j.atmosres.2015.05.010

Chen Z, Qie X, Liu D, Xiong Y (2019) Lightning data assimilation with comprehensively nudging water contents at cloud-resolving scale using WRF model. Atmos Res 221. https://doi.org/10.1016/j.atmosres.2019.02.001

Chu K, Xiao Q, Liu C (2013) Experiments of the WRF three/four-dimensional variational (3/4DVAR) data assimilation in the forecasting of Antarctic cyclones. Meteorol Atmos Phys 120:145–156. https://doi.org/10.1007/s00703-013-0243-y

Descombes G, Auligné T, Vandenberghe F, Barker DM, Barré J (2015) Generalized background error covariance matrix model (GEN_BE v2.0). Geosci Model Dev 8:669–696. https://doi.org/10.5194/gmd-8-669-2015

Environmental Modeling Center (2003) The GFS Atmospheric Model. NCEP Office Note 442, Global Climate and Weather Modeling Branch, EMC, Camp Springs, Maryland. https://repository.library.noaa.gov/view/noaa/11406/noaa_11406_DS1.pdf

Federico S, Avolio E, Bellecci C, Lavagnini A, Colacino M, Walko RL (2008) Numerical analysis of an intense rainstorm occurred in southern Italy. Nat Hazards Earth Syst Sci 8:19–35. https://doi.org/10.5194/nhess-8-19-2008

Federico S, Petracca M, Panegrossi G, Dietrich S (2017) Improvement of RAMS precipitation forecast at the short-range through lightning data assimilation. Nat Hazards Earth Syst Sci 17:61–76. https://doi.org/10.5194/nhess-17-61-2017

Federico S, Torcasio RC, Avolio E, Caumont O, Montopoli M, Baldini L, Vulpiani G, Dietrich S (2019) The impact of lightning and radar reflectivity factor data assimilation on the very short-term rainfall forecasts of RAMS@ISAC: application to two case studies in Italy. Nat Hazards Earth Syst Sci 19:1839–1864. https://doi.org/10.5194/nhess-19-1839-2019

Federico S, Torcasio RC, Puca S, Vulpiani G, Prat AC, Dietrich S, Avolio E (2021) Impact of radar reflectivity and lightning data assimilation on the rainfall forecast and predictability of a summer convective thunderstorm in southern Italy. Atmosphere 12(8):958. https://doi.org/10.3390/atmos12080958

Ferrari F, Cassola F, Tuju P, Mazzino A (2021) RANS and LES face to face for forecasting extreme precipitation events in the Liguria region (Northwestern Italy). Atmos Res 259(105):654. https://doi.org/10.1016/j.atmosres.2021.105654

Ferretti R, Faccani C, Cimini D, Marzano FS, Memmo A, Cucurull L, Pacione R (2005) Simulations of deep convection in the Mediterranean area using 3DVAR of conventional and non-conventional data. Adv Geosci 2:65–71. https://doi.org/10.5194/adgeo-2-65-2005

Fersch B, Wagner A, Kamm B, Shehaj E, Schenk A, Yuan P, Geiger A, Moeller G, Heck B, Hinz S, Kutterer H, Kunstmann H (2022) Tropospheric water vapor: a comprehensive high-resolution data collection for the transnational Upper Rhine Graben region. Earth Syst Sci Data 14:5287–5307. https://doi.org/10.5194/essd-14-5287-2022

Fierro AO, Mansell ER, Ziegler CL, MacGorman DR (2012) Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. Mon Weather Rev 140(8):2609–2627. https://doi.org/10.1175/MWR-D-11-00299.1

Fierro AO, Gao J, Ziegler CL, Mansell ER, MacGorman DR, Dembek SR (2014) Evaluation of a cloud-scale lightning data assimilation technique and a 3DVAR method for the analysis and short-term forecast of the 29 June 2012 Derecho Event. Mon Weather Rev 142(1):183–202. https://doi.org/10.1175/MWR-D-13-00142.1

Fletcher SJ (2017) Data assimilation for the geosciences from theory to application. Elsevier. https://doi.org/10.1016/B978-0-12-804444-5.09996-7

Giannaros TM, Kotroni V, Lagouvardos K (2016) WRF-LTNGDA: a lightning data assimilation technique implemented in the WRF model for improving precipitation forecasts. Environ Model Softw 76:54–68. https://doi.org/10.1016/j.envsoft.2015.11.017

Giazzi M, Peressutti G, Cerri L, Fumi M, Riva IF, Chini A, Ferrari G, Cioni G, Franch G, Tartari G, Galbiati F, Condemi V, Ceppi A (2022) Meteonetwork: an open crowdsourced weather data system. Atmosphere 2022(13):928. https://doi.org/10.3390/atmos13060928

Gustafsson N, Janjic T, Schraff C, Leuenberger D, Weissmann M, Reich H, Brousseau P, Montmerle T, Wattrelot E, Bučánek A, Mile M, Hamdi R, Lindskog M, Barkmeijer J, Dahlbom M, Macpherson B, Ballard S, Inverarity G, Carley J, Fujita T (2017) Survey of data assimilation methods for

convective-scale numerical weather prediction at operational centers: assimilation methods convective-scale NWP. Q J R Meteorol Soc 144. https://doi.org/10.1002/qj.3179

Hintz K, McNicholas C, Randriamampianina R, Williams H, Macpherson B, Mittermaier M, Onvlee J, Szintai B (2021) Crowd-sourced observations for short-range numerical weather prediction: report from EWGLAM/SRNWP Meeting 2019. Atmos Sci Lett 22. https://doi.org/10.1002/asl.1031

Hong SY, Kim J, Lim J, Dudhia J (2006) The WRF single moment microphysics scheme (WSM). J Korean Meteorol Soc 42:129–151

Huang XY, Xiao Q, Barker DM, Zhang X, Michalakes J, Huang W, Henderson T, Bray J, Chen Y, Ma Z, Dudhia J, Guo Y, Zhang X, Won DJ, Lin HC, Kuo YH (2009) Four-dimensional variational data assimilation for WRF: formulation and preliminary results. Mon Weather Rev 137(1):299–314. https://doi.org/10.1175/2008MWR2577.1

Lagasio M, Parodi A, Pulvirenti L, Meroni AN, Boni G, Pierdicca N, Marzano FS, Luini L, Venuti G, Realini E, Gatti A, Tagliaferro G, Barindelli S, Monti Guarnieri A, Goga K, Terzo O, Rucci A, Passera E, Kranzlmueller D (2019) Rommen B (2019) A synergistic use of a high-resolution numerical weather prediction model and high-resolution earth observation products to improve precipitation forecast. Remote Sensing 11(20):2387. https://doi.org/10.3390/rs11202387

Maiello I, Ferretti R, Gentile S, Montopoli M, Picciotti E, Marzano F, Faccani C (2014) Impact of radar data assimilation for the simulation of a heavy rainfall case in central Italy using WRF-3DVAR. Atmos Meas Tech 7:2919–2935. https://doi.org/10.5194/amt-7-2919-2014

Mansell ER, Ziegler CL, MacGorman DR (2007) A lightning data assimilation technique for mesoscale forecast models. Mon Wea Rev 135:1732–1748. https://doi.org/10.1175/MWR3387.1

Mazzarella V, Maiello I, Capozzi V, Budillon G, Ferretti R (2017) Comparison between 3D-Var and 4D-Var data assimilation methods for the simulation of a heavy rainfall case in central Italy. Adv Sci Res 14:271–278. https://doi.org/10.5194/asr-14-271-2017

Mazzarella V, Ferretti R, Picciotti E, Marzano FS (2021) Investigating 3D and 4D variational rapid-update-cycling assimilation of weather radar reflectivity for a heavy rain event in central Italy. Nat Hazards Earth Syst Sci 21:2849–2865. https://doi.org/10.5194/nhess-21-2849-2021

Miglietta M, Carnevale D, Levizzani V, Rotunno R (2021) Role of moist and dry air advection in the development of Mediterranean tropical-like cyclones (medicanes). Q J R Meteorol Soc 147:876–899. https://doi.org/10.1002/qj.3951

Miglietta M, Davolio S (2022) Dynamical forcings in heavy precipitation events over Italy: lessons from the HyMeX SOP1 campaign. Hydrol Earth Syst Sci 26:627–646. https://doi.org/10.5194/hess-26-627-2022

Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA (1997) Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. J Geophys Res 102D, 16 663–16 682. https://doi.org/10.1029/97JD00237

Papadopoulos A, Chronis TG, Anagnostou EN (2005) Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale model. Mon Weather Rev 133:1961–1977. https://doi.org/10.1175/MWR2957.1

Prat AC, Federico S, Torcasio RC, Fierro AO, Dietrich S (2021) Lightning data assimilation in the WRF-ARW model for short-term rainfall forecasts of three severe storm cases in Italy. Atmos Res 247:105246. https://doi.org/10.1016/j.atmosres.2020.105246. (ISSN 0169-8095)

Qie X, Zhu R, Yuan T, Wu X, Li W, Liu D (2014) Application of total-lightning data assimilation in a mesoscale convective system based on the WRF model. Atmos Res 145–146:255–266. https://doi.org/10.1016/j.atmosres.2014.04.012. (ISSN 0169-8095)

Román-Cascón C, Steeneveld G, Yagüe C, Sastre M, Arrillaga J, Maqueda G (2016) Forecasting radiation fog at climatologically contrasting sites: evaluation of statistical methods and WRF. Quart J Roy Meteor Soc 142:1048–1063. https://doi.org/10.1002/qj.2708

Rohm W, Guzikowski J, Wilgan K, Kryza M (2019) 4DVAR assimilation of GNSS zenith path delays and precipitable water into a numerical weather prediction model WRF. Atmos Meas Tech 12:345–361. https://doi.org/10.5194/amt-12-345-2019

Rudlosky SD (2015) Evaluating ENTLN performance relative to TRMM/LIS. J Operational Meteor 3(2):11–20. https://doi.org/10.15191/nwajom.2015.0302

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Liu Z, Berner J, Wang W, Powers JG, Duda MG, Barker DM, Huang XY (2019) A description of the advanced research WRF version 4; No. NCAR/TN-556+STR, NCAR Technical Note; National Center for Atmospheric Research: Boulder, CO, USA, 145p. https://doi.org/10.5065/1dfh-6p97

Tong W, Li G, Sun J, Tang X, Zhang Y (2016) Design strategies of an hourly update 3DVAR data assimilation system for improved convective forecasting. Weather Forecast 31(5):1673–1695. https://doi.org/10.1175/WAF-D-16-0041.1

Torcasio RC, Federico S, Comellas Prat A, Panegrossi G, D'Adderio LP (2021) Dietrich S (2021) Impact of lightning data assimilation on the short-term precipitation forecast over the central Mediterranean Sea. Remote Sensing 13(4):682. https://doi.org/10.3390/rs13040682

Torcasio RC, Mascitelli A, Realini E, Barindelli S, Tagliaferro G, Puca S, Dietrich S, Federico S (2023) The impact of GNSS Zenith Total Delay data assimilation on the short-term precipitable water vapor and precipitation forecast over Italy using the WRF model. Nat Hazards Earth Syst Sci Discuss. [preprint], https://doi.org/10.5194/nhess-2023-18. in review

Wagner A, Fersch B, Yuan P, Rummler T, Kunstmann H (2022) Assimilation of GNSS and synoptic data in a convection permitting limited area model: improvement of simulated tropospheric water vapor content. Front Earth Sci 10:869504. https://doi.org/10.3389/feart.2022.869504

Wang C, Wilson D, Haack T, Clark P, Lean H, Marshall R (2012) Effects of initial and boundary conditions of mesoscale models on simulated atmospheric refractivity. J Appl Meteor Climatol 51:115–132. https://doi.org/10.1175/JAMC-D-11-012.1

Wilks DS (2011) Statistical methods in the atmospheric sciences. Academic press

Zhu Y, Stock M, Lapierre J, DiGangi E (2022) Upgrades of the Earth Networks Total Lightning Network in 2021. Remote Sens 2022:14. https://doi.org/10.3390/rs14092209

Authors and Affiliations

E. C. Maggioni[1] · T. Manzoni[2] · A. Perotto[1] · F. Spada[1] · A. Borroni[2] · M. Giurato[3] · M. Giudici[4] · F. Ferrari[4] · D. Zardi[5] · R. Salerno[2]

[1] Ideam Srl, Cinisello Balsamo, Milan, Italy

[2] Meteo Expert, Cinisello Balsamo, Milan, Italy

[3] CMCC Foundation - Euro-Mediterranean Center On Climate Change, Bologna, Italy

[4] Department of Earth Sciences, Università Degli Studi Di Milano, Milan, Italy

[5] Department of Civil, Environmental and Mechanical Engineering, University of Trento, 38123 Trento, Italy