Answer to the reviewer #1 comment of NHESS-2018-319.

First of all we acknowledge the reviewer for the fast and complete review of the paper. In the following we will give answers/actions to improve the paper. Our comments are in red.

Before starting the discussion, we note that, in reviewing the paper, we found two errors: a) the length scale of the background error matrix in the x and y direction varies between 14 and 25 km and not, as stated into the paper, between 20 and 30 km; second the lightning number for each day written into the initial manuscript are wrong. The correct numbers are 82 331 for the 9 September, 291 164 for the 10 September (170 000 is written into the manuscript) and 105 467 for the 16 September (60 000 written in the manuscript). We apologize for these errors. However, the results shown in the paper were obtained using the correct number of flashes and the correct length scales in the background error matrix.

In the first submission we stressed the improvement given by the data assimilation at the local scale on the precipitation VSF (Very Short term Forecast, 0-3h). To highlight this point, we showed the many ways in which the forecast could be improved by the assimilation of lightning, radar or both. For example, the two stages of the Serano case show that the radar (first phase 03-06 UTC on 16 September) or lightning (second phase of the event, 18-21 UTC on 16 September) were the key observation to assimilate in order to improve the precipitation VSF. Also other stages had some specific aspects that we discussed.

Our attempt, however, was not successful, given the comments of both reviewers and the results section (Section 4) underwent a substantial rewriting. In particular, in the revised version of the paper, we will delete the Section 4.1.2 (second phase of the Serano case) and Section 4.2.1 (first case of the Livorno case). The results Section 4.2.1 will be shortly commented in Section 5 (Discussion and conclusions) to highlight that there is space for improvement. Following the comments of the reviewer #2, the scores of the phases commented into the paper will be put in three tables (Tables: 4-6) for specific thresholds (1, 6, 10, 20, 30, 40 mm/3h and, for Livorno, also 50 mm/3h). This will limit the number of precipitation thresholds considered but will increase the readability of the paper.

The space gained by deleting the two sub-sections stated above will be used to extend the discussion about the methods of assimilating lightning and radar in the RAMS@ISAC and to add two short sections to the result paragraph. In particular, we will extend the section “Lightning data assimilation” to include a discussion of the useful comments raised by the Reviewer#1, we will extend the section “Radar data assimilation” to show an example of 3D-Var assimilation of reflectivity factor (this should also answer to few comments of the Reviewer #1). A draft of these revised sections is reported at the end of this answer.

Finally, we will add a section (Section 4.3) to show how the lightning and radar data assimilation works together, presenting the evolution of the total water mass averaged for all VSF of the two cases and including in this discussion the assimilation stage, as well as sensitivity tests for the nudging formulation of lightning data assimilation (Section 4.4). The latter point requested new simulations with different model settings (see Table 3 at the end of this answer).

A draft of the new Results section (Section 4) is shown at the end of this answer. This could not be the final form because minor changes are still possible.

Summary: The authors utilize a cloud-scale functional relationship between lightning and water vapor mass mixing ratio published in the literature and applied it to a homegrown 3DVAR framework at the convection-allowing scale to evaluate the analysis and short term forecast of two selected high impact weather events over Italy.

Recommendation: reject and, eventually, re-submit.

Main Comments:
While the manuscript could eventually offer some merit for this journal, I found the analysis generally very rudimentary with the authors going at length in describing in excruciable level of details individual figures/panels in a repetitive and redundant manner without distilling the content into concise arguments/hypotheses. Given its repetitive nature, the entire results section could, in fact, easily be condensed into a 2-3 pages. Most importantly, the manuscript (hereafter, m/s) lacks rigor and rationales for the set ups and methods put forth for each, respective DA approaches. Salient Major issues are itemized below.

(1) As far as the scientific content is concerned, the core ideas and notions of this lightning data assimilation (LDA) method are conceptually similar to those from many existing studies, which fundamentally aim at promoting convective development through the introduction of latent heating within a prescribed neighborhood region/column centered at observed lightning locations. Past works from Benjamin et al. (2004), Alexander et al. (1999), Chang et al. (2001), Papadopoulos et al. (2005), Pessi and Businger (2009), have used empirical relationships between lightning-rainfall rates-latent heating or lightning- reflectivity rates-latent heating [e.g., in the HRRR]. Following a similar idea, recent works such as Machand and Fuelberg (2014), Lynn et al. (2015), Lynn (2017), Fierro et al. (2012; 2014, 2015), Wang et al. (2017, 2018) proposed LDA means that essentially boost the local thermal buoyancy where lightning is observed. A very limited portion of these techniques, however, offer alternative approaches to address spurious convection (i.e., removal) – which is a far more challenging problem to tackle. For completeness and given the relatively limited advances in LDA relative to radar DA, the authors should do a better job in discussing and including all the aforementioned references in their text. I was in fact astonished to notice that the integrity of the Results section in section 4 is completely devoid of references to previous works.
In particular, since they opted to borrow an LDA method from one of these investigators, comparisons with their study should be performed more systematically throughout the m/s. For instance, the works of Federico et al. 2017b is invoked when referring to multi-day forecast statistics using the Fierro et al. method without mentioning that, such a study, was already conducted by the same author over a larger domain and using nearly three times more forecast days/cases (Fierro et al. 2015 study). Given this omission, their study (Federico et al. 2017b) inadequately state that such multi-day statistics for this LDA have never been conducted. In a similar manner, it is of relevance to underline whenever appropriate that, in this work: (i) radial velocity is not included (specify why), (ii) only cloud-to-ground lightning data are considered and (iii) spurious convection is not addressed. In the light of (i) and (ii), one on the recent studies they

cite (Fierro et al. 2016) not only assimilated level II radar data (radial velocity + reflectivity factor) but used total lightning data. This needs to be clearly stated, for completeness (Cf comment 3 below for rationales).

In the revised version of the paper we will extend the discussion of the LDA in the introduction in order to include all the above papers.
Considering the other points:

(i) We are working on the assimilation of the radial velocity but the operator is not yet implemented in the 3D-Var. Also, while the reflectivity factor measured by the radar network is operationally available, the product of radial velocity is under development. At the moment, it needs further research to solve some issues (complex orography, operations of the radars not optimal for the Doppler retrieval and others). For these reasons, the attention was on the assimilation of reflectivity factor. These motivations will be discussed in the revised version of the paper in Section 3.3 by writing:

"Radial velocity is not assimilated in the RAMS@ISAC model because the operational product of radial velocity needs research to solve issues (complex orography, operation of radars non optimal for Doppler retrieval, not homogeneous coverage of the country), and it is not available for assimilation. Also, the implementation of radial velocity data assimilation is under development in RAMS-3DVar and it is not available for testing. For these reasons, we didn't consider the assimilation of radial velocity in this work. "

Considering the point (ii) in the paper we will write that total lightning are assimilated, not only CG. For these events the fraction of IC strokes to the total number of strokes detected by LINET is about 30% (22% on 09 September, 30% on 10 September and 35% on 16 November). There are cases when the IC strokes recorded by LINET are more than 50% of the total number of stokes over Italy. In general the Section on LDA will be extended to consider this point and others; (iii) The spurious convection is not considered by the LDA but it is considered in the assimilation of radar reflectivity factor. We will specify better this point in future version of the paper, but the comment is already present in the first submission version.

(2). In term of DA methodology, I found one major drawback, which is never discussed, nor evaluated. Given that both the LDA and their "RAD" experiment make adjustments to the relative humidity (RH) field, it is expected that both techniques will overlap in their adjustments over all the (many) grid points characterized by observed lightning flash rates exceeding zero. This is because changing RH is equivalent to adjusting Qv as RH ~ Qv/Qv_saturation. A more self-consistent DA approach would adjust the pseudo- observations for the Qv or RH field in a manner that eliminates any possibility of overlap during the minimization. Toward that end, the authors should include soundings and/or horizontal cross sections of RH/Qv that shows, quantitively, how the RH field is adjusted by each respective DA approach (radar vs lightning).

Second, given that lightning is a cloud-scale observation, I cannot find any justifications for not conducting the 3DVAR analysis on the innermost, higher resolution domain. Instead, the method minimizes the cost function on the intermediate domain and, later, projects the innovations on the coarser-scale domain. This needs to be addressed.

First: we will add a complete new section (Section 4.3) to address this point. In this section we will show the evolution of the accumulated precipitation and total water mass in the atmosphere (i.e. water vapour mass+mass oh hydrometeors) as a function of time (including the spin-up period). A draft of this new section is attached at the end of this answer.

Second:

Data assimilation is not performed on domain D3 (R1) because we don't have background error statistics for this grid.
Background error statistics for the domain D2 are computed by the NMC method, which, for this paper, is based on HyMeX-SOP1 simulations. The Appendix A and B of Federico (2013) shows the detail of the application of the method, which require a number of simulations (see also Barker et al., 2004 for the general discussion). Because the application of the domain D3 is exceptional the background error matrix was not computed for this domain and no data assimilation was performed.
Of course, this limitation is only for radar reflectivity factor because lightning are assimilated by nudging. Nevertheless, we could not reproduce the rationale of the paper, i.e. compare simulations with or without data assimilation for a specific domain, assimilating lightning in the innermost domain and for this reason we assimilated flashes over the D2 only.
In the paper we will specify better the role of the domain D3 and the reason for not assimilating lightning and radar reflectivity factor over the domain D3.

We will write in section 3.1
"The third domain covers the Tuscany Region, has 4/3 km horizontal resolution (R1), and it is used for Livorno to represent with higher spatial detail the precipitation field over Tuscany and to show better the precision of the rainfall VSF using data assimilation at the local scale. The fine structures of the precipitation field are smeared out over Tuscany using only domains D1 and D2. The operational implementation of the RAMS@ISAC model uses the domains D1 and D2 and no refinement for specific areas of Italy are used because Italy is a complex orography country and grid refinements for a specific event can be done only after the occurrence of the event."

And few lines below:

"It is noted that data assimilation is performed in the domain D2 (R4) only, and the innovations are transferred to the domain D3 (R1), for the Livorno case, by the two way-nesting. The domain D3 is used for the Livorno case to refine the resolution of the precipitation field over Tuscany and to show the spatial and temporal precision of the precipitation forecast over Tuscany using data assimilation. However, its usage is exceptional because, as stated above, Italy is a complex orography country and grid refinements over specific areas are used only after the occurrence of an event. For these reasons the domain D3 is usually not used in RAMS@ISAC simulations and no statistics about the background error are available for this grid. Because lightning are assimilated by nudging, they could be easily assimilated over the domain D3. Nevertheless, to preserve the rationale of the paper, i.e. comparing simulation with or without data assimilation for specific domains, we didn't assimilate lightning for domain D3.
Of course, being lightning and radar cloud scale observations, their assimilation at higher horizontal resolution is foreseeable in future works. "

Third, the radius of influence/decorrelation length scale chosen for radar reflectivity factor (50 km) is far too large for convective scale applications and would incur unrealistically large amount of Qv mass added into the domain – which will undoubtedly yield to spin-up issues and the generation of convective-scale gravity waves that will degrade longer term (>= 3h) solutions (please provide plot of perturbation pressure in your response). In that regard, the authors should indicate and contrast the total amount of Qv mass added by RAD and LIGHT.

The 50 km length is not a distance to spread the innovation introduced by radar reflectivity factor data assimilation. It represents a search radius to compute the pseudo-profile of relative humidity used in 3D-Var. A discussion about this point will be introduced in the new section on radar reflectivity factor data assimilation.
In particular we will write:

"It is important to point out that the 50 km length-scale of the above step doesn't represent the horizontal correlation length-scale of the background error, which determines the horizontal

spreading of the innovations in the 3D-Var data assimilation (the latter length-scale is between 14 and 25 km depending on the level). The 50 km length-scale is used to set a square for computing the pseudo-profile of relative humidity (Eqn. (2)). This profile is given by a weighted average whose weights are determined by the agreement between the simulated and observed reflectivity factor. The larger the agreement the larger the weight. This distance seems appropriate because the spatial error of meteorological models in simulating meteorological features, for example fronts, can be of this order. The control simulation for the two events considered in this paper confirms this choice."

(3). In the context of forecast improvements, the Qv-based method they borrowed/adapted was scaled for total lightning data (> 50% detection efficiency of intra-cloud [IC] flashes). I was surprised to find that absolutely no information on the detection efficiency and geolocation accuracy of the lightning network used (LINET) is provided in the text [no figures either]. Given the large area covered by this study, it is thus very likely that the geolocation accuracy of this network remains very poor for low amplitude flashes and for all flashes over oceanic regions. Given the low sferics amplitudes of IC flashes, the VLF portion of the sensor will miss nearly all these flashes, while the VHF portion only is able to detect some of the IC flashes within a few tens of kilometers away from the station [e.g., Rison, MacGorman works]. Thus, it is relevant to state and underscore that LINET only detects a very small portions of the total IC flashes in the study domain (likely < 5%). Motivation for scaling the F12 method for IC flashes (in lieu of cloud-to-ground [CG] flashes), lies in the well-documented finding that, in contrast to CGs, ICs are well correlated with thunderstorm kinematic and microphysical evolution (updraft strength, updraft volume, graupel mass etc, see Wiens et al. 2005, Schultz et al. 2011 among many others). CGs, on the other hand, were found to be correlated with the descent of reflectivity cores and the onset of the demise of the storm's updraft core [MacGorman and Nielsen 1991, MacGorman et al. 1989, Rutledge and Lang's seminal works etc]. Not surprisingly, ICs were found to lag CG by an average of 15 min [see one of the recent MacGorman study]. Moreover, Boccippio et al. 2001 and Medici et al. 2017 found that in deep continental convection, IC flashes always outnumber CGs by a ratio sometime exceeding 10:1. Based on these facts, it becomes clear why the Fierro method emphasized the use of IC flashes [or total lightning] for their application. Further motivation arises from the recent successful launches of the GLM instrument aboard GEOS-16/17, which will provide continuous day/night coverage of total lightning at ~90% detection efficiency (DE) over a large domain covering the Americas (Gurka et al. 2006; Goodman et al. 2012, 2013, Rudlosky et al. 2018). Note that GLM will provide flash extent information of lightning, while the metric derived from the (limited) point flash data in this study can only provide a very rough surrogate for CG flash location density at best. Similar space-borne technology to detect lightning have been developed by China (Feng-Yun-4, yang et al. 2016) with these data being assimilated in recent works by Wang et al. (2017, 2018) – which were never referenced either. Apart from their propensity to detect total lightning at a high DE, the chief advantage of this technology lies in its ability to retrieve lightning over remote oceanic regions.

LINET has been started and used operationally since 2004. Since then, more than 100 publication have appeared that give evidence about both DE and LA. In particular, since the beginning in 2004 LINET exhibited a statistical average location accuracy of some 100 m. Because a minimum of 5 sensor reports are exploited for each stroke solution, the LA does not deteriorate within several 100 km from a sensor. Thus, the LA is excellent all over the present study region.

LINET Europe comprises more than 200 sensors and provides more extensive stroke data than any other VLF/LF system in the region.

LINET detects and records stroke signals down to currents of a few kA (CG normalization). This is the reason why LINET ranges are large enough to exploit >=5 sensors for geolocation without reducing the typical baselines of 250-300 km. The resulting DE is good enough to detect any CG. Over the Mediterranean the stroke DE diminishes due to larger baselines. However, the flash detection is less sensitive because of the stronger strokes that characterize a flash.

Like any other VLF/LF system signals are recorded whether CG or IC. Thus, the detected IC portion is certainly not lower than in any other VLF/LF system. As a consequence, total lightning is reported at least as efficient than in any other VLF/LF system, and will be beneficial for the purposes in the present paper.

IC discrimination of LINET is based on TOA analysis. The advantage is a unique discrimination when the detection geometry is within certain ranges; the disadvantage is decreasing discrimination power when the distance to the closest sensor become too large, because of too small TOA differences between CG and IC at the same 2D location. Thus, over water far from land the identified IC fraction decreases, though total lightning counts remain relevant.

We emphasize that the time evolution of IC reports in considered area (not too far from land) signify very well the change of meteorological condition, especially with respect to severe weather. Note that the relatives changes (including lightning jumps in rate and altitude) are indicative, without the need to have absolute event numbers. See for example ref. “Thunderstorm Nowcasting” in Met. Tech. Int., Sept. 2017, p.109-112.

It is true that leader steps signify discharge processes (see, e.g., well-known LMA results). However, it is well-proven that VLF/LF detects pulses from IC activity that are very similar to CG strokes; this is why CG-IC discrimination is very challenging for VLF/LF systems. We think, though, that any VHF issue is not relevant here, because there is no large-scale VHF system that covers Italy and the surrounding sea with baselines of a few 10 km.

Observations from global networks or satellites may be a point of future concern, but do not represent any focus in the present paper; also, IC discrimination is either not yet possible or poor. It may be mentioned that GLM lightning data is not yet an issue; interestingly, Eumetsat/NASA on behalf of NOAA have selected LINET to carry out the first evaluation of the new lightning data source. This has been communicated in Science Team Meetings and conferences (see GLM Cal/Val 2017 Ground Validation Field Campaign 2017).

These points will be discussed in the new section on lightning data assimilation (Section 3.2). A draft of this section is shown at the end of this answer, but it is still incomplete.

(4) The following key information pertaining to the respective DA methods are missing/never discussed:
(a) What are the background/observation errors for reflectivity/lightning? (b) What statistics are used for model error ?
Lightning are assimilated by nudging and no error is associated with them. The error matrix for model error will be clarified in the section on the radar reflectivity factor data assimilation (see the attached draft of Section 3.3).

(c) How is the adjoint for the lightning data assimilation operator derived ?
The derivation of the adjoint of lightning data assimilation was performed using two case studies of the HyMeX-SOP1 (unpublished work) as commented in Federico et al., (2017a). We will comment about this point in the new section on lightning data assimilation. Also, in future versions of this paper we will add a new section (Section 4.4) to show the sensitivity of the rainfall VSF score (POD and ETS) to the nudging formulation.

(d) What assumptions are made for grid points with zero lightning or zero reflectivity observations ? Does the radar DA or LDA treat those as missing observations or equate those to the background values to reduce spread ?
Lighting are assimilated by nudging and this comment doesn’t apply. In the case of radar, grid points with zero reflectivity factor and zero simulated reflectivity factor are assumed missing observation, and the innovations can spread freely. Again this will be clarified in future versions of the paper (see the Section 3.3 draft).

(e) What Gaussian decorrelation length scales are assumed for each observation ? Please specify/justify/explain. How would the selection of a given length scale value, influence the results ?

The observation error matrix for radar reflectivity is diagonal (this was already stated in the first submission of the paper). We acknowledge that the sensitivity tests proposed by the reviewer are interesting, nevertheless they will be left for future studies. The importance of this point will be discussed shortly in the paper. We will write:

"The observation error matrix R in Eqn. (4) is diagonal and observations' errors are uncorrelated. This choice is partially justified by under sampling the radar reflectivity factor observation by choosing one point every five grid points in both horizontal directions of the radar observations Cartesian grid (Rohn et al., 2001) . However, correlation observations errors have significant impact on the final analysis, as shown for example in Fierro et al. (2016), and different choices of the matrix R will be considered in future studies.
The value of the elements on the diagonal of R depends on the vertical level and are 1/4 of the diagonal element of the B_z matrix at the corresponding height. By this choice, we give more credit to the observations than to the background and analyses strongly adjust the background towards observations."

(f) Is spurious convection addresses by either DA method ? Please elaborate.

Yes, in the radar reflectivity factor data assimilation, but not in the lightning data assimilation. The point is already present in the discussion paper, but it will be better clarified in future versions of the paper in the section dedicated to the radar data assimilation.

(g) Does the variational minimization set use a multi-scale approach ? If yes, what influence radii are chosen and how many cycles ?

We don't use the multi-scale approach. This will be clarified in the paper in the section dedicated to the radar data assimilation.

(5) Why did the authors not include the fractions skill score FSS as the main evaluation metric for their forecast? Several works have posited that, in contrast to ETS, FSS does not penalize displacement errors as much and, arguably, FSS offers a more accurate measure of skill on convection-allowing grids (Mittermaier et al. 2013).
Additionally, more recent studies evaluating forecast performance have been making usage of the so-called performance diagrams, which conveniently merge several key contingency table elements into one single diagram (Roebber 2009). The authors should show such diagrams to provide a more complete and succinct view of the overall forecast performance of the case they selected.

Considering POD and ETS gives the possibility to show the many facets of a forecast, and this, in our opinion, is important. These scores are also widely used in the bibliography and this make the results of this paper comparable with other papers. We, of course, acknowledge that there are other interesting measurements of the model performance, as FSS, that could be considered. Taking into consideration this comment and the comment of the reviewer #2 about the score we propose the following solution: we consider three neighborhood radii to take into account for displacement errors; nearest neighborhood (as in the first submission), 25 km and 50 km. We will put the scores in three tables (Tables 4-6 attached at the end of this answer) following a remark of the second reviewer.

(6) The case studies selected are cherry-picked given the confession that CTRL generally failed to provide reasonable forecast estimates of precipitation for both cases herein. For good measure, fairness and to better underscore the performance of the DA method, the authors should show the results for one case in which CTRL did not perform well and contrast it to one case where CTRL did preform reasonably well.

The events were selected because they were missed by several forecasts and, for this reason, they are challenging. Moreover, they had important consequences because nine people died and damage to properties was extensive. We will stress better this point in the introduction by writing:

"The forecast of severe events at the local scale still remains a challenge because of the multitude of physical processes involved over a wide range of scales (Stensrud et al., 2009). The Serano case, being localised in space, poses challenges in forecasting the exact position and timing of convection initiation; the Livorno event involves the interaction between a high impact storm with the complex orography of Italy, which is difficult to simulate at the local scale. For the above reasons the forecast of both events was challenging, as confirmed by the poor forecast of RAMS@ISAC. The difficulty to forecast timely and accurately the precipitation fields of the two cases is the main reason for choosing them as test cases for testing the lightning and radar data assimilation."

(7) The authors omit to mention that the degradation of the forecast at >= 3h is mainly due to saturation of the model solution by errors and biases within the initial / boundary conditions derived from large scale models or re-analysis datasets. This needs to be shown for both cases, especially given the unrealistically large (50 km) decorrelation length scale used for radar reflectivity factor.
Ok we will consider this point in the revised version of the paper. However, model errors plays an important role in the degradation of the forecast in addition to IC/BC. Again 50 km is not the decorrelation length scale for radar reflectivity factor.

(8) Title: Revise to include that: (i) primarily CG flashes are assimilated and (2) the model vehicle is RAMS.

We will include in the title that RAMS@ISAC is the model vehicle. As stated above, the IC flashes for the case studies considered in this paper is about 30%, which is not a small fraction of the total lightning. The discussion on the method assimilating lightning will be widened to consider this and other points.

Because these issues are collectively substantial and would require thorough rewriting of the manuscript in many places, I opted not to dwell on editorial comments for the time being. Additionally, the level of English remains, in my view, unacceptable for publication.

We will revise the English of the paper, also according to the suggestions of the reviewer 2 in the PDF file. The copy-editing service of the journal will also improve the quality of the English.

Figures:
Figures 5, 6, 8, 12a, 13a, 14a, 15a, 16a, 17a, 18a: The use of colored dots makes it very difficult to effectively compare the observations with those of the simulations: For consistency, either both sets of plots should show colored dots or shaded contours. For lightning, the authors should effectively show the gridded lightning data that were used to create the Qv or RH pseudo-observations.

Ok. It is always difficult to choose the right representation of the precipitation field when comparing model output with raingauges. We acknowledge that the solution suggested by the reviewer is a good one, however we also like our representation because: a) rainfall at the raingauges is not interpolated, avoiding in this way errors introduced by the interpolation process; b) the rainfall predicted by the model shown as a field gives the possibility to see the behavior of the model also in parts uncovered by raingauges. We propose the following solution: a) redraw the RAMS@ISAC rainfall field changing the colorbar to match exactly the raingauges colorbar; b)

adding the representation suggested by the reviewer as supplemental material to the paper (Figures S1-S3 at the end of this answer); c) recalling the supplemental material when discussing the second VSF of Livorno to highlight the wet frequency bias when assimilating radar reflectivity factor (see Figure S3 at the end of this answer).
Ok for the Figures about lightning. They will be redrawn according to the reviewer remark.

Additional comments:
General comment: What is the main rationale for using a model that is marginally known by the community (RAMS) versus a more commonly used, battle tested, publicly available model such as WRF-ARW ? The authors not only seem to re-invent the wheel here but render any potential future work dedicated to reproducibility of the results - to the least - very challenging.

RAMS@ISAC is used/maintained/developed at ISAC-CNR since several years (and it is also operational over Italy since 2000, in different versions/adaptations etc), and it is important for us to test our tool for challenging cases, as those considered in this work.
Also, we are WRF users too (see, for example, Avolio and Federico, 2018) and for the cases of this paper no specific differences were found for the performance of WRF and RAMS@ISAC models (using the same initial and dynamic boundary conditions). The performance of WRF model for the Livorno case is shown, for example, in Ricciardelli et al. (2018) and the reviewer can see that the comments given in this paper about the performance of RAMS@ISAC for the most intense phase of the Livorno case can also be applied to the WRF model (see specifically their Figures 11 and 12 for the most intense phases in Livorno). Consider also that Ricciardelli et al. used ECMWF IC/BC, which are different from that used in this paper. So, the results of this paper could be even more valuable because they are “more general” and not linked to the specific modelling tool.
We will add a reference to the above cited paper and a short discussion in Section 5 (Discussion and Conclusions).

(1) Bottom, page 2: what are “conventional data” ? Why are radial velocity data not used ? Line 70: the main advantage of using 3DVAR vs 4DVAR, EnKF or hybrid methods lies in their already low computational burden. Thus, I do not agree with this justification. Also, variables are not “perturbed”; but adjusted by VAR methods.

For radial velocity we already answered. We will change the paper according to the reviewer suggestion for the other parts of the comment.

(2) Pages 3 and 4: Please refer to Major Comments 1 and 3. Lines 105: Given that “Federico et al. (2017a) implemented the methodology of Fierro et al. (2012) ...”, how come on line 112 “We use the method of Federico et al. (2017a) to assimilate lightning...” ? Please revise accordingly.
Ok for this comment. We will add the reference to Fierro et al. (2012). The comment of line 112 come from the fact that we intended to cite the adaptation of the methodology, that is discussed in Federico et al. (2017a).

(3) Line 124: c.f. end of Major Comment 1.
Ok.

(4) Line 240: RAMS used diagnostic relationships (vs explicit) to forecast lightning as it does not explicitly solves for the 3D electric field. Line 243: “Fourth”
For the first comment we wrote: “Second, it predicts the occurrence of lightning following the diagnostic methodology of Dahl et al. (2012),….”

(5) Line 290: Delete equation set as these are considered basic/common knowledge.

In some papers, where we omitted the equations, we had the opposite comment. However, for this paper, to reduce length and to give more space to the important points raised by the reviewer the equations will be deleted.

(6) Section 3.2, lines 300-312: Explicitly state and indicate that equation (2) is from Fierro et al. (2012, 2015) and not from Federico et al. Line 305: Please explain the rationales behind the choices of these constants: In particular, how are the forecast metrics affected for a 20% change in A, which has been shown to exhibit the most notable influence on the forced convection?
Ok for the reference. The functional form is that of Fierro et al. (2012, 2015), but the coefficients were adapted for RAMS@ISAC as shown in Federico et al (2017a). In Federico et al. (2017a) it is clearly stated that the method is that of Fierro et al. (2012), the only difference being the adaptation to RAMS@ISAC model. Sensitivity tests to the nudging formulation will be shown in Section 4.4.

(7) Line 316-317: c.f. Major Comment 2.
Ok.

(8) End of page 11: c.f. Major Comment 2
Ok.

(9) Line 356: do the authors refer to the LFC or the LCL, (which may I add is an idea borrowed from Marchand and Fuelberg 2014 and Fierro et al. 2016). What is the top of the adjustment layer for lightning ? Please elaborate.
It is the LCL. The idea is of Caumont et al. (2010), we didn't add the reference to this point of the paper because the whole methodology is taken from Caumont et al. (2010), already cited several times. The top adjustment for lightning is -25°C. However, this is already stated in the paper (Lines 314-315 "The check and eventual substitution of the water vapor is performed every five minutes and it is made only in the charging zone (0 °C, -25°C).").

(10) Line 410 and elsewhere. This is similar to the results of Fierro et al. 2016. C.f. Major Comment 1. Please establish comparisons with previous works throughout the manuscript.
Ok.

(11) Line 669: This statement is incorrect. The DE of ground based sensors levels off very rapidly with distance from land. This is where space-borne lightning detection systems such as the GLM or Feng Yun-4 can fill the gap.
Ok, however the good coverage of the LINET network for some important areas, as between Corsica and Italian mainland (both Liguria and Tuscany) makes this point "less problematic" for the Livorno case.

(12) Lines 716-725: c.f. Major Comment 1.
Ok.

Hereafter we show the new sections on lightning data assimilation (Section 3.2) on radar data assimilation (Section 3.3) and the new results section (Section 4).

3.2 Lightning data assimilation

Lightning data are provided by LINET (LIghtning detection NETwork; Betz et al., 2009; www.nowcast.de) which has more than 500 sensors worldwide with the greatest density over Europe (more than 200 sensors). The network has a good coverage over Central Europe and

Western Mediterranean (from 10 W to 35 E and from 30 N to 60 N). The area of good coverage includes the region considered in this paper.

LINET exploits the VLF/LF electromagnetic bands and provides measurements of both IC (intra-cloud) and CG (cloud to ground) discharges. IC strokes are detected as long as lightning occurs within 120 km from the nearest sensor thanks to an optimised hardware and advanced techniques to process the data (TOA-3D, Betz et al., 2004). According to Betz et al. (2009), LINET has a location accuracy of 100 m (since 2004) for an average distance of 200 km among the sensors verified by strikes into towers of known positions.

The good performance of the LINET network and its ability to detect IC strokes is shown in Lagouvardos et al. (2009) for a storm in southern Germany, while the good performance over Italy, including both CG and IC strokes, is discussed in Petracca et al. (2014).

The lightning data assimilation scheme is that of Fierro et al. (2012; 2014) and uses the total lightning, i.e. intra-cloud plus cloud to ground flashes.

The method starts by computing the water vapour mixing ratio q_v:

$$q_v = Aq_s + Bq_s \tanh(CX)(1 - \tanh(Dq_g^{\alpha})) \quad (1)$$

Where coefficients are set to A=0.86, B=0.15, C=0.30, D=0.25, α=2.2, q_s is the saturation mixing ratio at the model atmospheric temperature, and q_g is the graupel mixing ratio (g kg^{-1}). X is the number of total flashes (IC+CG) falling in a grid box of domain D2 (R4) in the past five minutes. The mixing ratio q_v of Eq. (1) is computed only for grid points where flashes are recorded. More specifically, for each grid point we consider the number of flashes falling in a grid box centred at the grid point in the last five minutes. The mixing ratio of Eqn. (1) is compared with that predicted by the model. If the mixing ratio of Eqn. (1) is larger than the simulated one, the latter is changed with the value given by Eqn. (1), otherwise the modelled mixing ratio is left unchanged. This method can only add water vapour to the forecast.

The check and eventual substitution of the water vapor is performed every five minutes and it is made within the mixed phase layer zone (0 °C, -25°C), wherein electrification processes are the most active (Takahashi 1978, Emersic and Sounders, 2010; Fierro et al., 2015). It is also noted that some authors use the layer (0 °C, -20°C) (Fierro et al., 2012; 2015).

The scheme of Fierro et al. (2012; 2014) was adapted to RAMS@ISAC in Federico et al. (2017a). In particular, the coefficient C of Eqn. (1) was rescaled from that of Fierro et al. (2012) considering the different spatial and temporal resolution of the gridded lightning data; then the coefficient C was tuned (increased) by trials and errors considering two case studies of HyMeX-SOP1 (15 and 27 October 2012). The C constant was adapted subjectively considering two opposite requests:

increasing the hits and minimising (or not increasing substantially) the false alarms. POD and ETS scores were considered as metrics for this purpose. Then, Eqn. (1) was applied to twenty case studies of HyMeX-SOP1 giving a statistically significant (90, or 95% depending on the rainfall threshold) improvement of the RAMS@ISAC precipitation VSF (3h).

Nevertheless, an exhaustive statistics on the performance of rainfall VSF to the nudging formulation in RAMS@ISAC is missing and further studies are needed in this direction. Also, the optimal choice of the coefficients A, B, C, D and α are case dependent.

In addition to the above issues there is another important point for the application of the Fierro et al. (2012) method to RAMS@ISAC. Fierro et al (2012) applied the method using the ENTLN network, which has a detection efficiency (DE) greater than 50% for IC over Oklahoma, where the ENTLN data were used. The emphasis on IC flashes in the set-up of Fierro et al. (2012) method is given because observational and model studies have provided evidence that IC flashes are better correlated than CG flashes with various measures of intensifying convection (updraft strength, volume, graupel mass flux etc.; Carey and Rutledge 1998; MacGorman et al. 2005; Wiens et al. 2005; Fierro et al. 2006; Deierling and Petersen 2008; MacGorman et al. 2011). For this reason methods that use both IC and CG flashes performs better than those using CG, the latter being correlated with the descent of reflectivity cores and the onset of the demise of the storm' s updraft core.

A direct DE for IC strokes cannot be reliably compared with that of ENTLN, because the area is different and the technical details about IC detection remain unclear (type of signals, VLF/LF or VHF, discrimination IC-CG). An analysis for the case studies shows that IC strokes are about 30% of the total number of strokes reported. Also, the fraction of IC strokes to the total strokes depends on the position. For example, for the Serano case, the fraction of IC strokes detected by LINET over the area hit by the largest precipitation is more than 50% while over the Adriatic Sea it decreases to 10%-15%.

For all the above reasons there are limitations to the application of Eqn. (1) to RAMS@ISAC and it is appropriate to study the dependence of the rainfall VSF to the nudging formulation. This is shown in Section 4.4.

It is finally noted that despite the limitations noted above, the lightning data assimilation, as used in this paper, has a significant and positive impact of the RAMS@ISAC rainfall VSF (Federico et al., 2017a; 2017b).

3.3 Radar data assimilation

The method assimilates CAPPI of radar reflectivity factor operationally provided by the Italian Department of Civil Protection (DPC). Radar data are provided over a regular Cartesian grid with 1 km horizontal resolution and for three vertical levels (2, 3, 5 km above the sea level). The CAPPIs at 2, 3, and 5km can be considered as under-sampled vertical profiles. CAPPIs are composed starting from the 22 radars of the Italian Radar Network (Figure 13) 19 operating at the C-band (i.e., 5.6 GHz) and 3 at X-band (i.e., 9.37 GHz). Data quality control and CAPPI composition is performed by DPC. Data quality processing chain aims at identifying most of the uncertainty sources as clutter, partial beam blocking and beam broadening. The radar observations are processed according to nine steps detailed in Vulpiani et al. (2014), Petracca et al. (2018) and references therein.

Before entering the data assimilation, the Cartesian grid is reduced to 5 km by 5 km by choosing one point every five of the Cartesian grid provided by DPC in order to reduce the numerical cost of the data assimilation and to reduce the effect of correlated observation errors (Rohn et al., 2001). The radar grid (Figure 4, for example) is then a Cartesian grid with 5 km grid-spacing and three vertical levels.

The methodology to assimilate radar reflectivity factor is that of Caumont et al. (2010), named 1D+3DVar, which is a two-step process: first, using a Bayesian approach inspired to GPROF (Olson et al., 1996; Kummerow et al., 2001), 1D pseudo-profiles of model variables are computed, then those pseudo-profiles are assimilated by 3DVar. Both steps are discussed below.

The first step computes a pseudo-profile of relative humidity weighting the model profiles of relative humidity around the radar profile (Bayesian approach). The pseudo-profile is computed by:

$$\mathbf{z_o^p} = \frac{\sum_i \mathbf{RH_i} W_i}{\sum_j W_j} \tag{2}$$

Where RH_i is the RAMS@ISAC vertical profile of relative humidity at a grid point inside a square of 50*50 km^2 centred at the radar vertical profile, W_i is the weight of each profile and $\mathbf{z_o^p}$ is the relative humidity pseudo-profile. The summation is taken over all the grid points inside a square of 50*50 km^2 around the observed profile and the denominator is a normalisation factor. The weights are determined by the agreement between the simulated and observed reflectivity factor:

$$W_i = \exp\left\{-\frac{1}{2}\left[\mathbf{z_o} - h_z(x_i)\right]^T \mathbf{R_z^{-1}}\left[\mathbf{z_o} - h_z(x_i)\right]\right\} \tag{3}$$

Where h_z is the forward observation operator, transforming the background column $\mathbf{x_i}$ into the observed reflectivity factor. The forward observation operator is specific for the WSM6 microphysics scheme and is available in WRF release 3.8. It assumes Marshall-Palmer hydrometeors size-distribution, Rayleigh scattering, and depends on the mixing ratios of rain, graupel and snow.

The matrix $\mathbf{R_z}$ in Eqn. (3) is diagonal and its value is $n\sigma^2$, where σ is 1 dBz and n is the number of available observations in the vertical profile (from 1 to 3). In this way, we give more weight to vertical profiles containing more data.

It is important to point out that the 50 km length-scale of the above step doesn't represent the horizontal correlation length-scale of the background error, which determines the horizontal spreading of the innovations in the 3D-Var data assimilation (the latter length-scale is between 14 and 25 km depending on the level). The 50 km length-scale is used to set a square for computing the pseudo-profile of relative humidity (Eqn. (2)). This profile is given by a weighted average whose weights are determined by the agreement between the simulated and observed reflectivity factor. The larger the agreement the larger the weight. This distance seems appropriate because the spatial error of meteorological models in simulating meteorological features, for example fronts, can be of this order. The control simulation for the two events considered in this paper confirms this choice.

The method is not able to force convection when the model has no rain, snow and graupel in a square around (50*50 km^2) a specific radar profile with reflectivity factor greater than zero. In this case, the pseudo-profile of relative humidity is assumed saturated above the lifting condensation level and with no data below to force convection into the model.

It is also noted that the method is able to reduce spurious convection when the reflectivity factor is simulated but not observed, because the pseudo-profile of relative humidity gives more weight to the drier relative humidity profiles simulated by RAMS@ISAC inside the 50*50 km^2 square centred at the radar profile examined. Of course, the ability to reduce spurious convection depends on the availability of drier model profiles around the specific radar profile (see the example below). Finally, if the observed profile is dry and the profile simulated by RAMS@ISAC is dry too, the pseudo-profile is not computed.

In summary, pseudo-profiles are computed for each profile of the radar grid whenever reflectivity is observed or simulated.

The pseudo-profiles computed with the procedure introduced above, are then used as observations in the RAMS-3DVar data assimilation (Federico, 2013), minimising the cost-function:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(\mathbf{z}_o^p - h(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{z}_o^p - h(\mathbf{x})) \quad (4)$$

Where **x** is the state vector giving the analysis when *J* is minimized, $\mathbf{x}_b$ is the background, **B** and **R** are the background and observations error matrices, $\mathbf{z}_o^p$ is the pseudo vertical profile computed by Eqn. (2) and *h* is the forward observation operator transforming the state vector (RAMS@ISAC water vapour mixing ratio) into observations. The cost function in RAMS-3DVar is implemented in incremental form (Courtier et al., 1994) and its minimization is performed by the conjugate-gradient method (Press et al., 1992). No multi-scale approach is used.

The background error matrix is divided in three components along the three spatial directions (*x, y, z*). The $\mathbf{B_x}$ and $\mathbf{B_y}$ matrices account for the spatial correlation of the background error. The correlations are Gaussian with length-scales between 14 and 25 km, depending on the vertical level. These distances are computed using the NMC method (Barker et al., 2012) applied to the HyMeX-SOP1 (Hydrological cycle in the Mediterranean Experiment – First Special Observing Period occurred in the period 6 September-6 November 2012; Ducroq et al., 2014) period. It is again stressed that the spread of the innovations along the horizontal spatial directions in the 3D-Var analysis is determined by the length scales of $\mathbf{B_x}$ and $\mathbf{B_y}$ matrices and varies between 14 and 25 km. The Bz matrix contains the error for the water vapour mixing ratio, which is the control variable used in RAMS-3DVar. This error is about 2 g/kg at the surface and decreases with height. In particular it is larger than 0.5 g/kg below 4 km and less than 0.2 g/kg above 5 km. The vertical decorrelation of the background error depends on the level and can be roughly estimated in 500-2000 m. The observation error matrix R in Eqn. (4) is diagonal and observations' errors are uncorrelated. This choice is partially justified by under sampling the radar reflectivity factor observation by choosing one point every five grid points in both horizontal directions of the radar observations Cartesian grid. However, correlation observations errors have significant impact on the final analysis, as shown for example in Fierro et al. (2016), and different choices of the matrix **R** will be considered in future studies.

The value of the elements on the diagonal of R depends on the vertical level and are 1/4 of the diagonal element of the B_z matrix at the corresponding height. By this choice, we give more credit to the observations than to the background and analyses strongly adjust the background toward observations. We could choose to give more credit to the background compared to the observations, nevertheless the poor performance of the control forecast for the selected cases justifies this choice. The background error matrix is computed using the NMC method (Parrish and

Derber, 1992; Barker et al. 2004) applied to the HyMeX-SOP1 (Hydrological cycle in the Mediterranean Experiment – First Special Observing Period occurred in the period 6 September-6 November 2012; Ducroq et al., 2014). This choice is motivated by the fact that HyMeX-SOP1 contains several heavy precipitation events over Italy and the background error matrix is representative of the convective environment of the cases considered in this paper. In particular, 10 out of 20 declared IOP (Intense Observing Period) of HyMeX-SOP1 occurred in Italy (Ferretti et al., 2014). On the contrary, the period of September 2017, before and after the events selected in this study was characterized by fair and stable weather conditions over Italy and the background error matrix for September 2017 is less representative of the convective environments that characterise the events of this paper. It is also important to highlight that the dependence of the results on the choice of the background error matrix is mainly determined by the choice of the horizontal and vertical length scales of the background error correlation because the observation error matrix (**R**) is ¼ of the background error at the same level to give more credit to the observations than to the background at this level (comparison at the levels above an below that of Figure 14a shows that the method was able to dry the model west of Sardinia).

Because it is the first time that the assimilation of radar reflectivity factor in RAMS@ISAC model is shown it is useful to discuss an example of analysis. We select the analysis for the Livorno case at 06 UTC. The observed CAPPI at 3km above sea level is shown in Figure 10b. The corresponding CAPPI simulated by the background is shown in Figure 14a. In general, the comparison between simulated and observed reflectivity factor shows the difficulty of the model to represent convection properly. In particular, the model is able to represent the convection over Northern Italy but it has poor performance over Sardinia, south of Sicily and over Central Italy. The difference between the analysis and background relative humidity after and before the analysis is shown in Figure 14b (absolute values less than 1% are suppressed in the figure for clarity). Both positive (convection enhancing) and negative (convection suppressing) adjustments can be found. Over Central Italy, Sardinia and South of Sicily relative humidity is increased because the model doesn't simulate the observed reflectivity (Figure 10b). Over northern Italy the model is partially dried for two different reasons: over northwest of Italy because RAMS@ISAC simulates unobserved reflectivity, over north and northeast of Italy because the model simulates larger values of reflectivity factor compared to the observations. The RAMS-3DVar is able to dry the relative humidity field north of Corsica island, where the RAMS@ISAC predicts unobserved reflectivity, while RAMS-3DVar didn't suppress the unobserved convection west of Sardinia because the pseudo profiles computed over this area weren't appreciably drier than the

background. Cross correlations among variables are neglected in this study and the applications of the RAMS-3DVar affects the water vapour mixing ratio only.

Because the lightning data assimilation perturbs the water vapour mixing ratio, it follows that the data assimilation presented in this study changes only this parameter.

4. Results

In this section, we discuss the most intense phase of the Serano case, 03-06 UTC on 16 September, and two VSF forecasts, 00-03 UTC and 06-09 UTC on 10 September, for the Livorno case. The two VSF for Livorno correspond to the most intense phase of the storm in Livorno and to a very intense phase over Lazio region, Central Italy. The aim of the section is to show the notable improvement given to the very short term forecast by the lightning and radar reflectivity factor data assimilation. In the discussion paper two additional VSF are discussed, one for Serano and one for Livorno; also, the discussion paper shows the behaviour of the scores for a number of rainfall thresholds larger than those shown in this section.

In particular, we consider four types of VSF: a) CTRL, without radar reflectivity factor and lightning data assimilation; b) LIGHT, assimilating lightning but not radar reflectivity factor; c) RAD, assimilating radar reflectivity factor but not lightning; d) RADLI, assimilating both lightning and radar reflectivity factor. A_76 and SAT show the sensitivity of the results to the nudging formulation. Table 3 shows the types of simulations considered in this paper.

4.1 Serano: 03-06 UTC 16 September 2017

In this period, an intense and localised storm hit the central Italy, while light precipitation occurred over northern Italy (Figure 15a). Considering the storm over central Italy, 10 raingauges observed more than 30 mm/3h, 6 more than 40 mm/3h, 3 more than 50 mm/3h and 1 more than 60 mm/3h, the maximum observed value being 63 mm/3h.

The CTRL forecast, Figure 15b, misses the storm over central Italy and considerably underestimates the precipitation over Northern Italy, giving unsatisfactory results.

The assimilation of the radar reflectivity factor improves the forecast, as shown by Figure 15c. In particular, RAD forecast shows localized precipitation (30-35 mm/3h) close to the area were the most abundant precipitation was observed. However, the maximum precipitation is underestimated. Also, the RAD forecast better represents the precipitation over Northern Italy compared to CTRL.

The precipitation forecast of LIGHT, Figure 15d, shows some improvements compared to CTRL because the precipitation over central Italy has a maximum of 25-30 mm/3h, close to the area where the maximum precipitation was observed. LIGHT, however, has a worse performance compared to RAD because it misses the small precipitation amount over northern Italy. Also, similarly to RAD, LIGHT underestimates the maximum precipitation.

RADLI forecast, Figure 15e, shows the best performance. The precipitation over central Italy is well represented because the maximum rainfall (40-45 mm/3h) is in reasonable agreement with observations, and also because the area with intense precipitation (> 25 mm/3h) is elongated in the SW-NE direction in agreement with raingauge measurements, giving a much better idea of the real storm intensity compared to RAD and LIGHT, as well as CTRL. The precipitation over northern Italy is well represented by RADLI.

Table 4 shows the ETS and POD scores for selected rainfall thresholds for different neighbourhood radii. Different radii are considered to account for the well-known double penalty error (Mass et al., 2002; Mittermaier et al., 2013) caused by displacement errors of the detailed precipitation forecast in convection allowing grids.CTRL was unable to predict rainfall larger than 6 mm/3h. The comparison between RAD and LIGHT shows that assimilating radar reflectivity factor performs better than assimilating lightning. This behaviour, however, is not general and sometimes the assimilation of lightning has a better performance than assimilating radar reflectivity factor (see section 4.2.1).

RADLI forecast has the best performance among all model configurations. In particular, it is the only forecast having positive scores for thresholds larger than 30 mm/3h.

In conclusion, for this VSF, the assimilation of lightning and radar reflectivity factor acted synergistically to improve the precipitation VSF and the simulation assimilating both data performs considerably better than simulations assimilating either lightning or radar reflectivity factor.

4.2 Livorno

The Livorno case lasted for several hours starting at 18 UTC on 9 September 2017 and ending more than a day later. The most intense phase in Livorno and its surroundings was observed during the night between 9 and 10 September. In the following, we will show two representative VSF (3h), including the most intense phase in Livorno.

4.2.1 Livorno: 00-03 UTC 10 September 2017

This period represents the most intense phase of the storm in Livorno. In particular, the raingauge close to the label A (Figure 16a) reported 151 mm/3h (Collesalvetti), while the one close to the label B measured 82 mm/3h. Among the 518 raingauges reporting valid data, 75 observed more than 10 mm/3h, 31 more than 20 mm/3h, 17 more than 30 mm/3h, 9 more than 40 mm/3h, and 6 more than 50 mm/3h.

The CTRL precipitation forecast is shown in Figure 16b. The forecast is poor because it misses the precipitation swath from the coast towards NE. A precipitation swath is forecasted about 50 km to the North of the real occurrence, but it is less wide compared to the observations.

The forecast of RAD, Figure 16c, shows that the assimilation of radar reflectivity factor gives a clear improvement to the forecast. The largest precipitation in the coastal part of the swath (we searched the maximum value in the area with longitudes between 10.20E and 10.70E and latitudes between 43.10N and 43.60N) is 94 mm/3h. Another local maximum is in the southern part of the domain (label B). The maximum location is well represented, but the forecast value is 55 mm/3h while the observed maximum is 82 mm/3h.

An improvement, compared to both CTRL and RAD, is given by the assimilation of lightning (Figure 16d). Also for this simulation there is a precipitation swath from coastal Tuscany to the Apennines, but the shape of the swath better resembles that observed. The maximum value close to Livorno, i.e. in the coastal part of the swath, is 158 mm/3h.

The LIGHT simulation also shows the local maximum in the southern part of the domain (about 50 mm/3h), but the amount is underestimated.

Figure 16e shows the rainfall forecast by RADLI. The precipitation swath from coastal Tuscany towards NE is more apparent compared to LIGHT and RAD. The maximum rainfall accumulated close to Livorno is 186 mm/3h. Also, the second precipitation maximum in the southern part of the domain reaches 70 mm/3h in good agreement with observations (82 mm/3h). RADLI is the only run giving a satisfactory precipitation field over the south-eastern Emilia Romagna (north-eastern part of the domain), on the lee of the Apennines.

It is also noted that the main precipitation swath forecasted by RADLI is too broad in the direction crossing the swath compared to the observations. This is confirmed by the FBIAS of RADLI (not shown), which is more than 3 for thresholds larger than 42 mm/3h.

The analysis of the scores (Table 5) confirms the results outlined above. CTRL has the lowest performance and the improvement given by the data assimilation to the VSF is apparent for POD and ETS for all thresholds and neighbourhood radii considered. For this specific VSF, lightning data assimilation gives a better improvement to rainfall forecast compared to RAD. RADLI has the best

performance, especially for 25 km and 50 km neighbourhood radii, nevertheless it over forecast the precipitation field (Figure 16). Because ETS penalizes false alarms, the value of this score for RADLI is sometimes lower than that for LIGHT.*4.2.3 Livorno: 06-09 UTC 10 September 2017*

In this period, the most intense phase of the precipitation occurred over central Italy, over the coastal part of Lazio (Figure 17a). More in detail, among the 2695 raingauges reporting valid data over the domain of Figure 17a, 307 reported more than 10 mm/3h, 132 more than 20 mm/3h, 86 more than 30 mm/3h, 66 more than 40 mm/3h, 49 more than 50 mm/3h and 35 more than 60 mm/3h. Among the 35 raingauges measuring more than 60 mm/3h, 33 were over Lazio, showing the heavy rainfall occurred over the Region.

Some precipitation persisted over Tuscany but the rainfall is much lower compared to previous 6h (the rainfall over Tuscany between 03 and 06 UTC was very intense, not shown). Other notable precipitation areas are over the NE of Italy (moderate to low amounts), over Central Alps (moderate values) and over the whole Sardinia (small amounts).

Figure 17b shows the rainfall simulated by CTRL. The forecast is unsatisfactory, mainly for the following two reasons: a) heavy precipitation is simulated over Tuscany (> 75 mm/3h), also close to the Livorno area; b) very small precipitation is forecasted over central Italy. The rainfall over NE Italy is well represented in space, but overestimated.

Considering the evolution of CTRL rainfall forecast for the two VSF of Livorno, we conclude that CTRL was able to predict abundant rain over Livorno, but this was delayed compared to the real event.

The rainfall simulated by RAD (Figure 17c) clearly improves the forecast compared to CTRL. First, the precipitation over Lazio is very well predicted and the rainfall values are higher than 40 mm/3h (up to 65 mm/3h), so the RAD forecast well represents the main precipitation spot over Italy for this VSF. Second, the precipitation over Tuscany is less than for CTRL, showing the ability of radar reflectivity factor data assimilation to dry the model when it predicts rain that is not observed. Third, the precipitation over central Alps is represented, even if located about 30 km to the East. It is noted, however, that the area of intense rainfall (>60 mm/3h) is overestimated by RAD, showing a wet forecast. This is confirmed by the wet frequency bias of the RAD simulation, which is greater than 3 between 14 and 44 mm/3h.

LIGHT forecast, Figure 17d, shows a worse performance compared to RAD for this time period. The precipitation forecast is mainly over Tuscany, where it is overestimated, with a small precipitation spot over Lazio.

The precipitation forecast of RADLI, Figure 17e, represents very well the precipitation over Lazio, and the rainfall amount is better predicted compared to RAD. The precipitation over Sardinia is well represented by RADLI as well as the precipitation over Central Alps, giving the best results among all forecasts.

The analysis of the scores confirms the above results (Table 6). CTRL has a poor performance as shown by the POD and ETS values, close to zero, for all thresholds above 30 mm/3h and for all neighbourhood radii. The simulations assimilating radar reflectivity factor performs better than LIGHT, the difference being larger for higher rainfall thresholds and for smaller neighbourhood radii.

It is also notable the good performance of RADLI forecast for the nearest neighbourhood radii (ETS=0.43, POD=0.92) for the 50 mm/3h threshold.

4.3 Evolution of total water

Because lightning data assimilation and radar reflectivity factor data assimilation both adjust the water vapour mixing ratio (q_v), it is interesting to evaluate the contribution of each data source to the q_v adjustment including in that evaluation the assimilation phase (0-6 h).

For the 3D-Var approach the impact of the contribution of data assimilation on q_v can be done using maps similar to Figure 14b. For example, Fierro et al. (2016), using a 3D-Var approach to assimilate lightning, used the layer averaged q_v between 3 and 10 km to quantify the water vapour added to the WRF model by lightning data assimilation. However, because in this paper lightning are assimilated by nudging, this kind of representation is not practicable because it is difficult to separate the contribution of the nudging from other processes in the evolution of q_v.

Fierro et al. (2015) used the total water substance mass (forecasted accumulated precipitation + total hydrometeors and water vapour mass) to quantify the impact of lightning data assimilation by nudging. In this paper, a similar approach is used. More specifically, we consider the forecasted accumulated precipitation and the total hydrometeors and water vapour mass in the atmosphere averaged over the grid columns. Also, we averaged all VSFs for Serano and Livorno. The evolution of the forecasted accumulated precipitation is shown in Figure 18a, while the evolution of the total hydrometeors and water vapour mass in the atmosphere is shown in Figure 18b.

Considering the Figures 18a and 18b it is apparent that flashes add less water vapour compared to radar reflectivity factor data assimilation and, of course, RADLI has the largest impact. In particular, the total water mass added to the background is 2.5%, 5.7% and 7.4% for LIGHT, RAD and RADLI, respectively. Importantly, the total water substance mass added by RADLI to the

background is less than the sum of the total water substance mass added by RAD and LIGHT. This happens because 3D-Var adds water to the background limiting the impact of nudging during the simulation. For example, in an already saturated atmosphere the nudging of Eqn. (2) doesn't have any impact.

Accumulated precipitation accounts for the largest part of the water vapour added to the simulation, similarly to Fierro et al. (2015). At the end the assimilation phase (6h), the evolution of the total water vapour and hydrometeors mass in the atmosphere converges towards the background as boundary conditions propagates into the domain.

4.4 Sensitivity to nudging formulation

As stated in Section 3.2, there are limitations when applying the nudging method of Fierro et al. (2012) to RAMS@ISAC. Also, the optimal setting of the coefficients of Eqn. (2) depends on the case study. For these reasons, it is interesting to evaluate the sensitivity of the results to changes in nudging formulation. For this purpose, we show the variability of ETS and POD scores to changes in the A and B coefficients of Eqn. (1). The scores are computed considering all the VSF for the two case studies for different configurations: A_76 has the coefficients A=0.76 and B=0.25; LIGHT has A=0.86 and B=0.15 (default setting), SAT has A=1.01 and B=0; RADLI has A=0.86 and B=0.15 (default setting); CTRL, and RAD are as defined in Table 3.

The scores are computed for the second RAMS@ISAC domains and are shown for the nearest neighbourhood. ETS score (Figure 19a) shows that all configurations assimilating either lightning or radar reflectivity factor alone or a combination of lightning and radar reflectivity factor improves the forecast for all thresholds. RADLI has the best ETS for rainfall intensity larger than 32 mm/3h in line with the results of the three VSF discussed above.

For rainfall lower than 32 mm/3h, the simulations assimilating lightning perform better, because they have less false alarms compared to those assimilating radar reflectivity factor (not shown). From the comparison of LIGHT and SAT with A_76, it is apparent that the latter has the worst score. The comparison between LIGHT and SAT shows mixed results: SAT performs better up to 38 mm/3h, while LIGHT is better for higher thresholds. This result is confirmed by POD, Figure 19b, which shows that SAT performs better up to 32 mm/3h, while LIGHT is better for higher thresholds. A visual inspection of the model output reveals that SAT can generate spurious convection in some areas while missing convection in other areas that are correctly forecast by LIGHT or even A_76, i.e. adding less water vapour to the model because of the different trajectories in the phase space followed by the model using different settings.

Lynn et al. (2015) implemented a method suggested by Fierro et al. (2012) to suppress spurious convection in WRF model. The method compares the lightning forecast during the assimilation period with lightning observations to filter out spurious convection. The application of the methodology on 10 July 2013 improved the forecast of the squall line from Texas to Iowa, which was the focus of the forecast on that day; however, the application of the method to 19 and 21 March 2012 over the US gave mixed results, improving the forecast in the first 6h and worsening it after 6h.

The implementation of this method in the RAMS@ISAC could be used to suppress spurious convection in simulations assimilating lightning, especially SAT.

It is finally noted that RAD and RADLI have high POD values for all thresholds, nevertheless their ETS is below that of LIGHT and SAT up to 32 mm/3 h (RADLI) and 42 mm/3h (RAD). This behaviour is caused by the larger number of false alarms given by assimilating radar reflectivity factor compared to those assimilating lightning. This result shows again that the RAD and RADLI configurations have a wet frequency bias. In particular, the frequency bias of RAD and RADLI configuration is about 3 between 20 and 40 mm/3h.

Table 3: Types of simulations performed.

Experiment	Description	Data assimilated	Model variable impacted
CTRL	Control run	None	None
RAD	RADAR data assimilation	Reflectivity factor CAPPI (RAMS-3DVar)	Water vapour mixing ratio
LIGHT	Lightning data assimilation (A=0.85; B=0.16 in Eqn (2))	Lightning density (nudging)	Water vapour mixing ratio
RADLI	RADAR + lightning data assimilation (A=0.86; B=0.15 in Eqn (2))	Reflectivity factor CAPPI (RAMS-3DVar) + Lightning density (nudging)	Water vapour mixing ratio
A_76	Lightning data assimilation (A=0.76; B=0.25 in Eqn (2))	Lightning density (nudging)	Water vapour mixing ratio

SAT	Lightning data assimilation (A=1.01; B=0. in Eqn (2))	Lightning density (nudging)	Water vapour mixing ratio

Table 4: ETS and POD scores for three different neighbourhood radii. Scores are computed over the domain D2.

Threshold (mm/3 h)	ETS nearest neighboorhood (CTRL, RAD, LIGHT, RADLI)	POD nearest neighbourhood (CTRL, RAD, LIGHT, RADLI)	ETS 25 km (CTRL, RAD, LIGHT, RADLI)	POD 25 km (CTRL, RAD, LIGHT, RADLI)	ETS 50 km (CTRL, RAD, LIGHT, RADLI)	POD 50 km (CTRL, RAD, LIGHT, RADLI)
1	(0.42,0.36,0.44, 0.33)	(0.57,0.87,0.60, 0.81)	(0.68,0.73,0.68, 0.73)	(0.77,0.93,0.75, 0.89)	(0.79,0.89,0.82, 0.87)	(0.84,0.92,0.84, 0.90)
6	(0.06,0.10,0.14, 0.13)	(0.0,0.5,0.20,0. 72)	(0.11,0.44,0.72, 0.41)	(0.11,0.86,0.72, 0.83)	(0.19,0.86,0.86, 0.92)	(0.19,0.86,0.86, 0.92)
10	(0.,0.05,0.,0.15)	(0.,0.26,0.,0.79)	(0.,0.66,0.58,0. 74)	(0.0,0.84,0.58,0 .89)	(0.,0.95,0.74,0. 90)	(0.,0.95,0.74,0. 90)
20	(0.,0.,0.,0.41)	(0.,0.,0.,0.8)	(0.0,0.41,0.33,0 .87)	(0.,0.47,0.3,0.9)	(0.,0.73,0.80,1. 0)	(0.,0.73,0.80,1. 0)
30	(0.,0.,0.,0.31)	(0.,0.,0.,0.5)	(0.,0.,0.,0.90)	(0.,0.,0.,0.9)	(0.,0.,0.,1.0)	(0.,0.,0.,1.0)
40	(0.,0.,0.,0.)	(0.,0.,0.,0.)	(0.,0.,0.,0.33)	(0.,0.,0.,0.33)	(0.,0.,0.,0.50)	(0.,0.,0.,0.50)

Table 5: ETS and POD scores for three different neighbourhood radii. Scores are computed over the domain D3.

Threshold (mm/3 h)	ETS nearest neighboorhood (CTRL, RAD, LIGHT, RADLI)	POD nearest neighbourhood (CTRL, RAD, LIGHT, RADLI)	ETS 25 km (CTRL, RAD, LIGHT, RADLI)	POD 25 km (CTRL, RAD, LIGHT, RADLI)	ETS 50 km (CTRL, RAD, LIGHT, RADLI)	POD 50 km (CTRL, RAD, LIGHT, RADLI)

1	(0.43,0.64,0.70, 0.56)	(0.67,0.86,0.98, 0.99)	(0.68,0.80,0.82, 0.71)	(0.83,0.92,0.98, 0.99)	(0.68,0.80,0.82, 0.71)	(0.83,0.92,0.98, 0.99)
6	(0.1,0.31,0.60,0 .49)	(0.24,0.58,0.89, 0.95)	(0.49,0.70,0.91, 0.96)	(0.55,0.76,0.96, 0.97)	(0.49,0.70,0.91, 0.96)	(0.55,0.76,0.96, 0.97)
10	(0.11,0.33,0.56, 0.54)	(0.19,0.56,0.75, 0.80)	(0.48,0.76,0.91, 0.97)	(0.52,0.79,0.92, 0.97)	(0.48,0.76,0.91, 0.97)	(0.52,0.79,0.92, 0.97)
20	(0.02,0.30,0.52, 0.59)	(0.03,0.39,0.74, 0.81)	(0.18,0.73,0.97, 0.93)	(0.19,0.74,0.97, 0.97)	(0.18,0.73,0.96, 0.93)	(0.19,0.74,0.97, 0.97)
30	(0.,0.27,0.51,0. 47)	(0.,0.29,0.76,0. 76)	(0.,0.64,0.94,1.)	(0.,0.65,1.,1.)	(0.,0.64,0.94,1.)	(0.,0.65,1.,1.)
40	(0.,0.44,0.27,0. 27)	(0.,0.44,0.56,0. 67)	(0.,0.89,1.,1.)	(0.,0.89,1.,1.)	(0.,0.89,1.,1.)	(0.,0.89,1.,1.)
50	(0.,0.33,0.66,0. 50)	(0.,0.33,0.67,0. 67)	(0.,0.67,1.,1.)	(0.,0.67,1.,1.)	(0.,0.66,1.,1.)	(0.,0.67,1.,1.)

Table 6 ETS and POD scores for three different neighbourhood radii. Scores are computed over the domain D2.

Threshold (mm/3 h)	ETS nearest neighboorhood (CTRL, RAD, LIGHT, RADLI)	POD nearest neighbourhood (CTRL, RAD, LIGHT, RADLI)	ETS 25 km (CTRL, RAD, LIGHT, RADLI)	POD 25 km (CTRL, RAD, LIGHT, RADLI)	ETS 50 km (CTRL, RAD, LIGHT, RADLI)	POD 50 km (CTRL, RAD, LIGHT, RADLI)
1	(0.41,0.63,0.61, 0.65)	(0.66,0.89,0.89, 0.93)	(0.79,0.83,0.82, 0.83)	(0.89,0.95,0.95, 0.96)	(0.88,0.92,0.93, 0.94)	(0.93,0.97,0.98, 0.98)
6	(0.2,0.4,0.39,0. 47)	(0.43,0.82,0.77, 0.88)	(0.45,0.63,0.71, 0.76)	(0.63,0.90,0.95, 0.96)	(0.72,0.86,0.88, 0.92)	(0.82,0.96,0.97, 0.96)
10	(0.,0.24,0.18,0. 28)	(0.14,0.78,0.55, 0.80)	(0.14,0.47,0.58, 0.62)	(0.24,0.86,0.82, 0.93)	(0.32,0.91,0.96, 0.95)	(0.35,0.95,0.97, 0.97)
20	(- 0.03,0.18,0.13, 0.22)	(0.01,0.81,0.30, 0.80)	(0.09,0.46,0.57, 0.61)	(0.11,0.86,0.59, 0.90)	(0.15,0.84,0.91, 0.96)	(0.15,0.90,0.92, 0.97)
30	(- 0.02,0.22,0.13, 0.28)	(0.,0.90,0.23,0. 88)	(0.01,0.79,0.46, 0.80)	(0.01,0.93,0.47, 0.94)	(0.02,0.95,0.93, 0.99)	(0.02,0.95,0.93, 0.99)
40	(- 0.1,0.24,0.08,0. 36)	(0.,0.83,0.12,0. 89)	(0.01,0.83,0.37, 0.83)	(0.02,0.97,0.38, 0.97)	(0.1,0.97,0.95,0 .98)	(0.02,0.98,0.95, 0.98)
50	(-	(0.,0.67,0.,0.92)	(0.,0.90,0.,0.90)	(0.,0.94,0.,0.96)	(0.,0.96,0.,0.96)	(0.,0.96,0.,0.96)

	0.01,0.27,0.,0.43)					

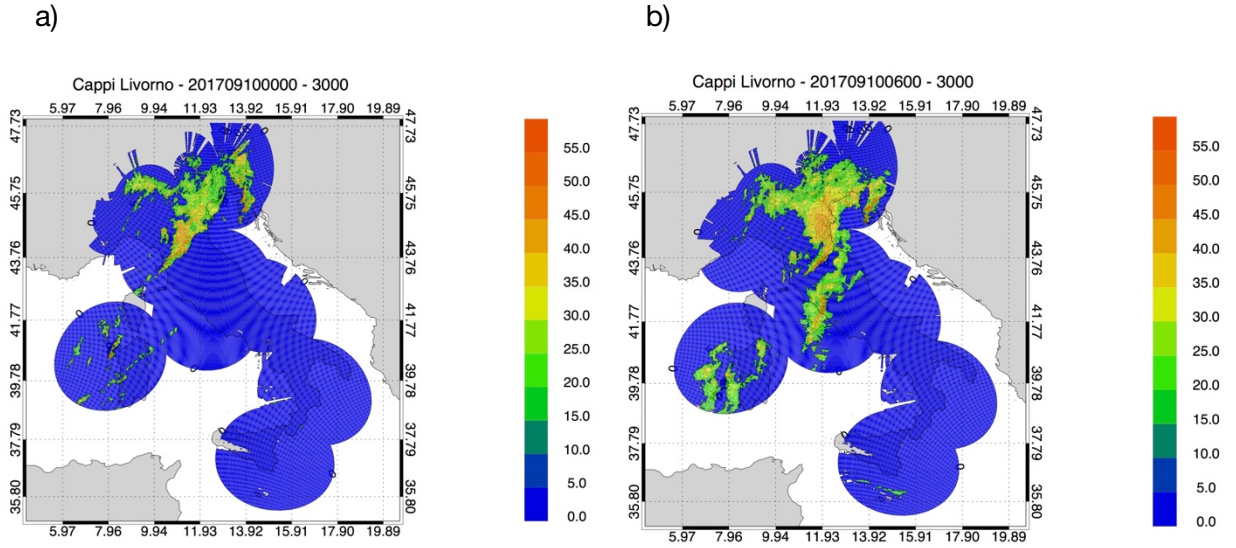


Figure 10: a) National radar mosaic at 3 km above the sea level observed at 00 UTC on 10 September 2017; b) as in a) at 06 UTC. (Not modified but referred in the new Section 3.3).

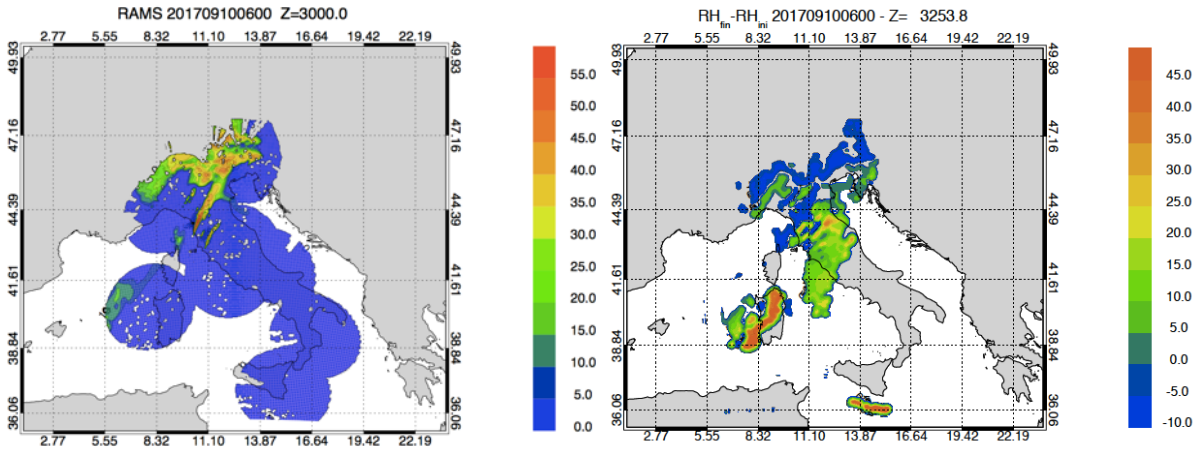
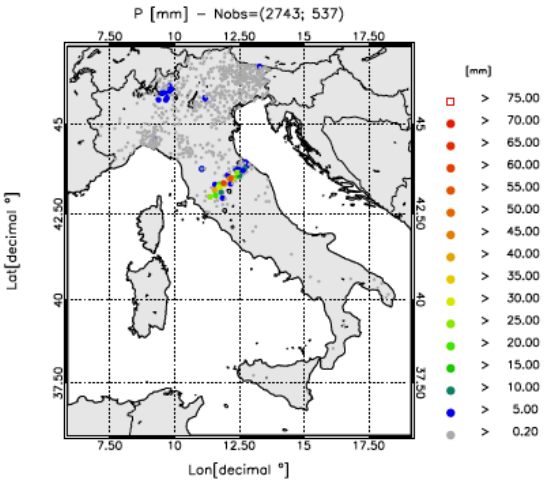
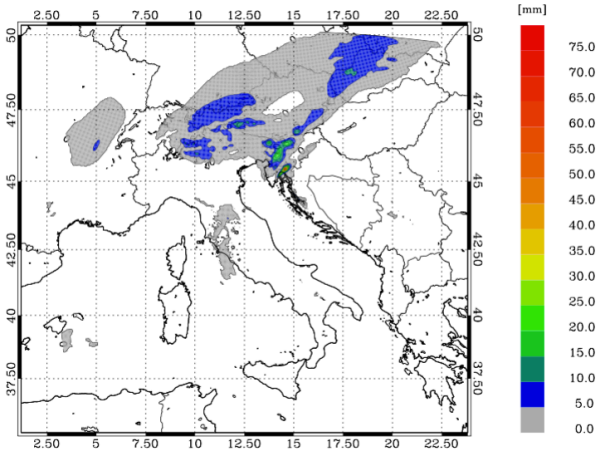


Figure 14: a) RAMS@ISAC reflectivity factor simulated 3 km above sea level at 06 UTC on 10 September 2017; b) relative humidity difference between the analysis and the background at 06 UTC at 3.2 km level in the terrain following vertical coordinate of RAMS@ISAC.

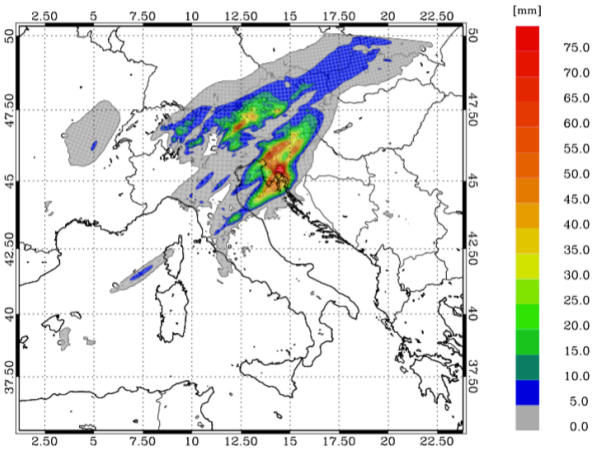
a)



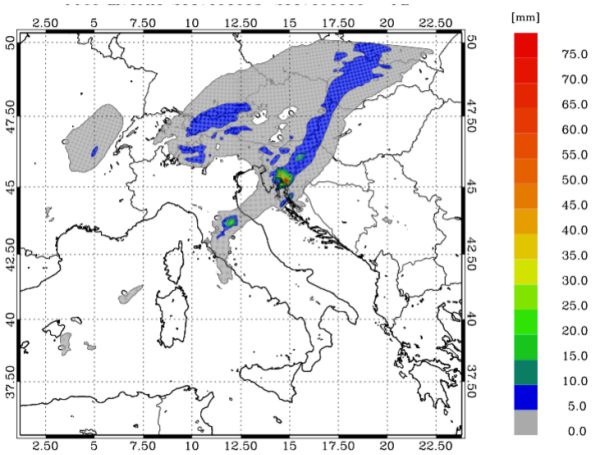
b)



c)



d)



e)

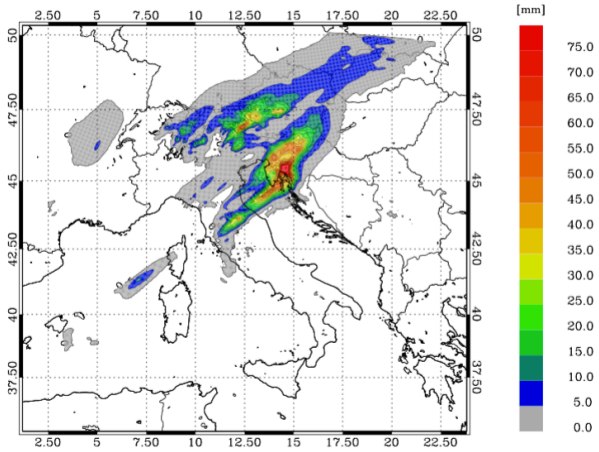
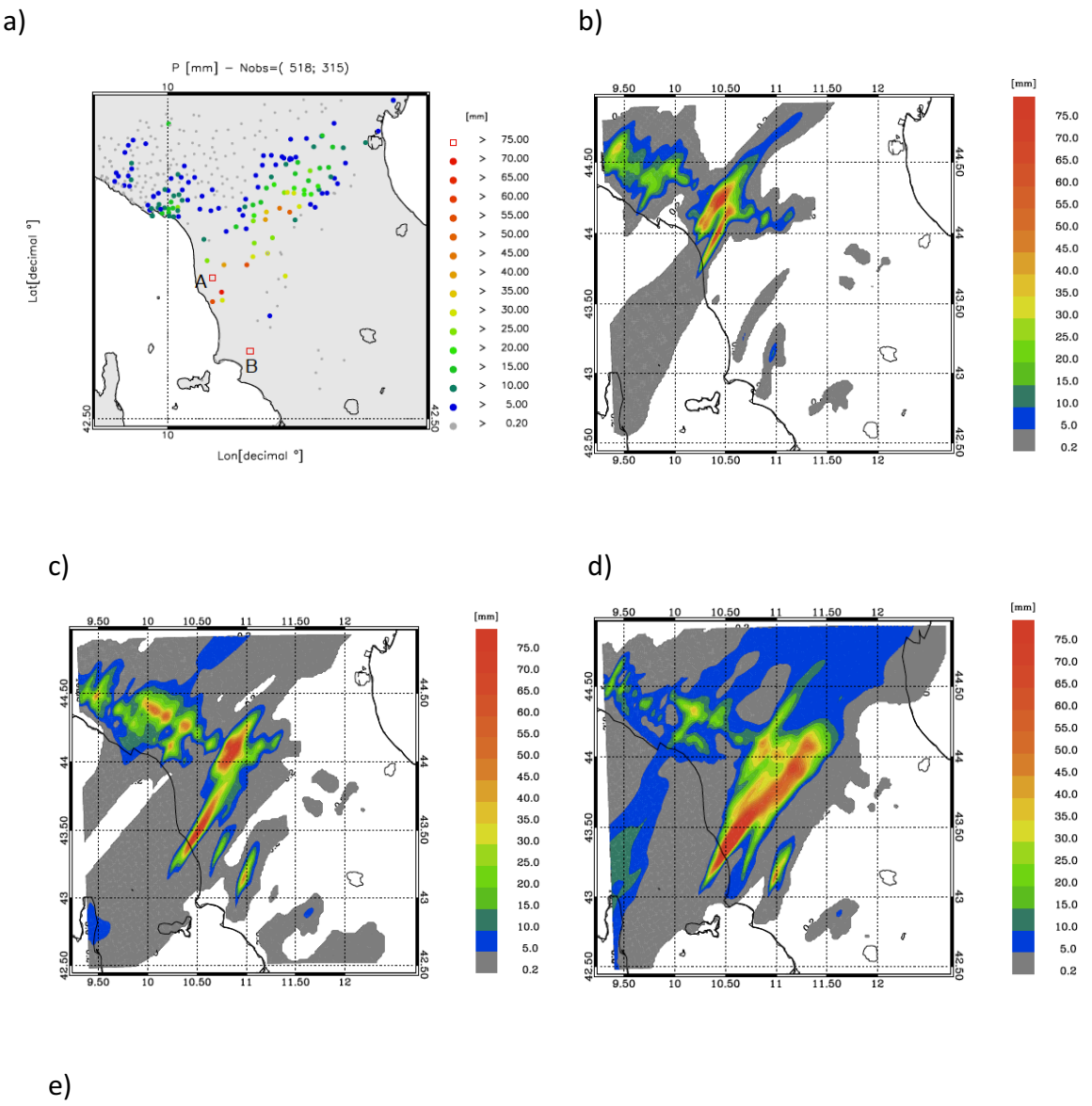


Figure 15: a) rainfall reported by raingauges between 03 and 06 UTC on 16 September 2017. Only raingauges observing at least 0.2 mm/day are shown. The first number in the title within brackets represents the available raingauges, while the second number represents those observing at least 0.2 mm/3h; b) as in a) for the CTRL forecast; c) as in a) for the RAD forecast; d) as in a) for the LIGHT forecast; e) as in a) for the RADLI forecast.



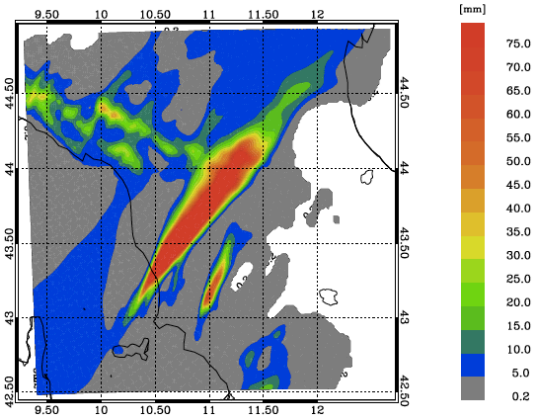
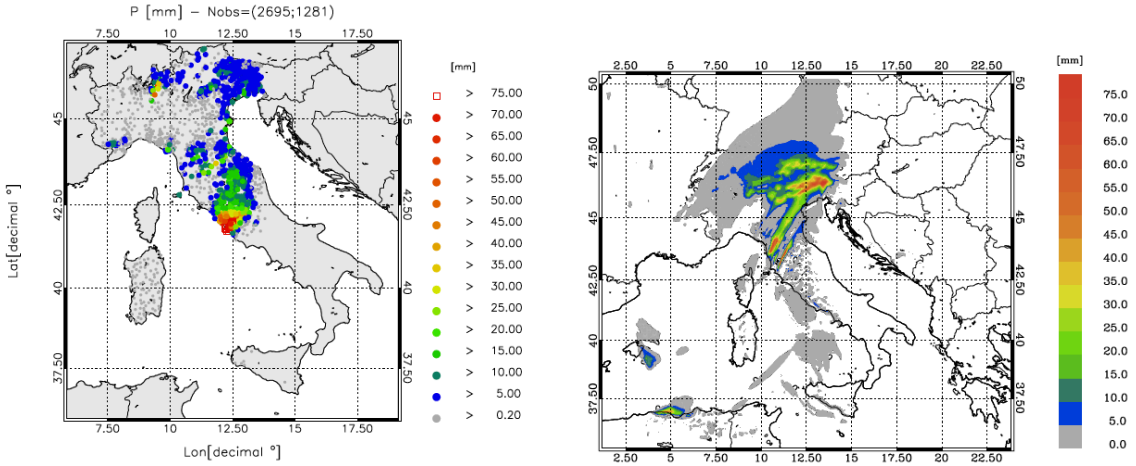


Figure 16: a) rainfall reported by raingauges between 00 and 03 UTC on 10 September 2017. Only stations reporting at least 0.2 mm/3h are shown. The first number in the title within brackets represents the number of raingauges available over the domain, while the second number shows those observing at least 0.2 mm/3h; b) as in a) for the CTRL forecast; c) as in a) for the RAD forecast; d) as in a) for the LIGHT forecast; e) as in a) for the RADLI forecast.

a)

b)



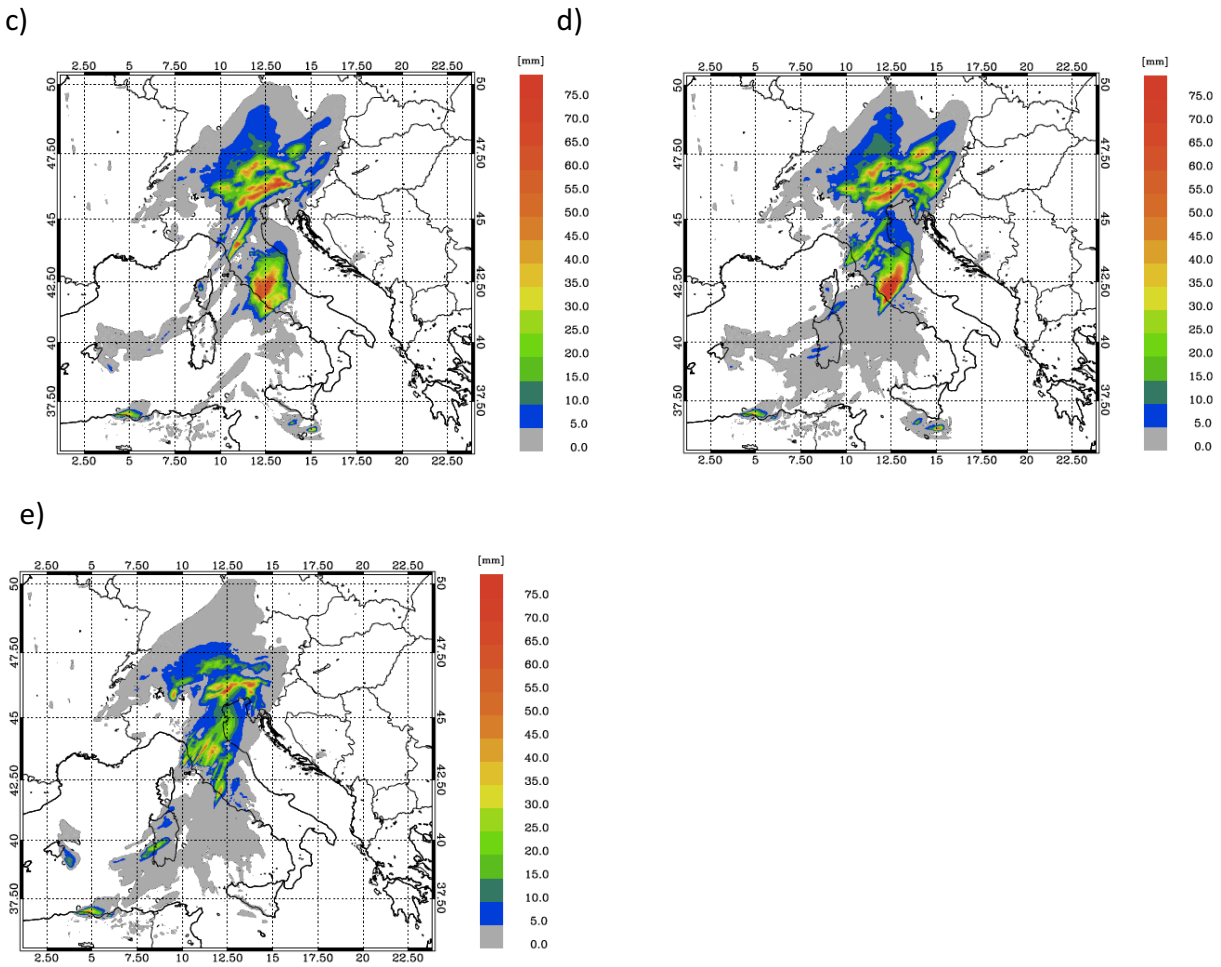


Figure 17: a) rainfall reported by raingauges between 06 - 09 UTC on 10 September 2017. For this time period 2695 raingauges reported valid observations in the domain, however only stations reporting at least 0.2 mm/3h are shown The first number in the title within brackets represents the number of raingauges available over the domain, while the second number shows those observing at least 0.2 mm/3h; b) as in a) for the CTRL forecast; c) as in a) for the RAD forecast; d) as in a) for the LIGHT forecast; g) as in a) for the RADLI forecast.

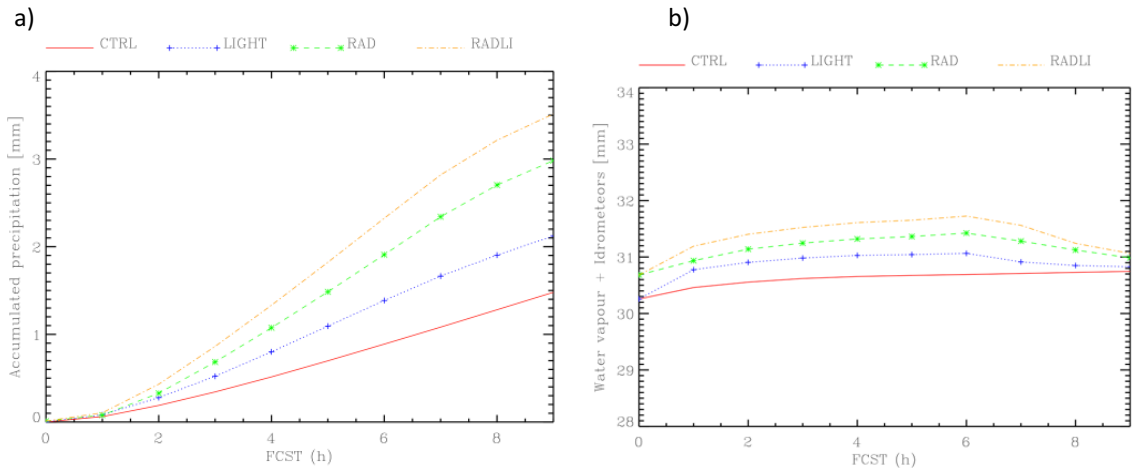


Figure 18: a) Evolution of the accumulated precipitation for different model configurations and for all forecast hours; b) as in a for the hydrometeor mass plus the water vapour equivalent mass per unit area. All quantities are expressed in [mm] and are averaged over the number of grid columns.

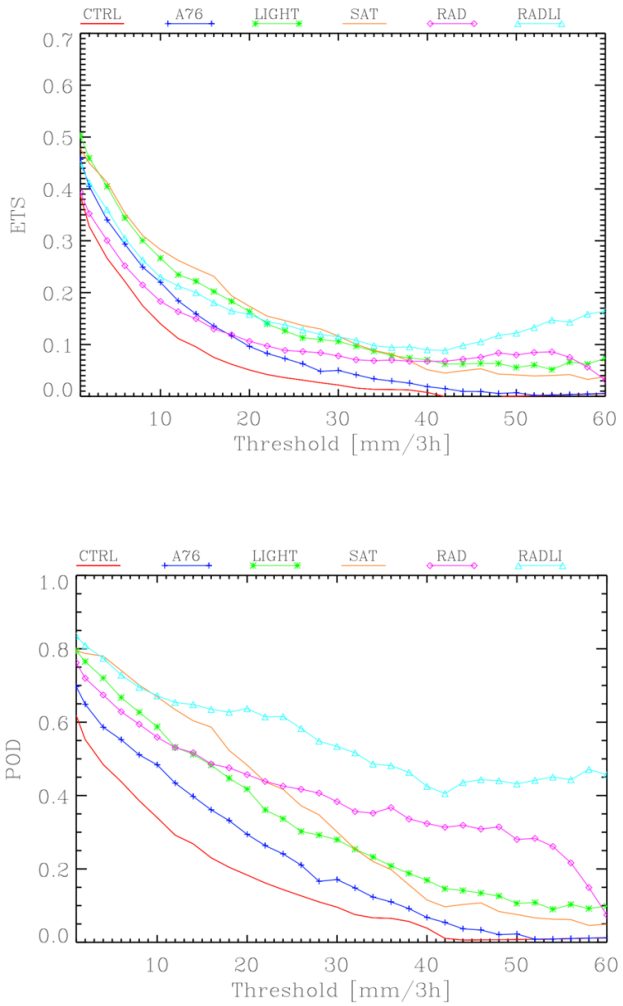


Figure 19: a) ETS score for all VSF considered in this paper; b) as in a) for the POD score.

SUPPLEMENTAL MATERIAL

Serano 2017-09-16 03-06 Domain D2

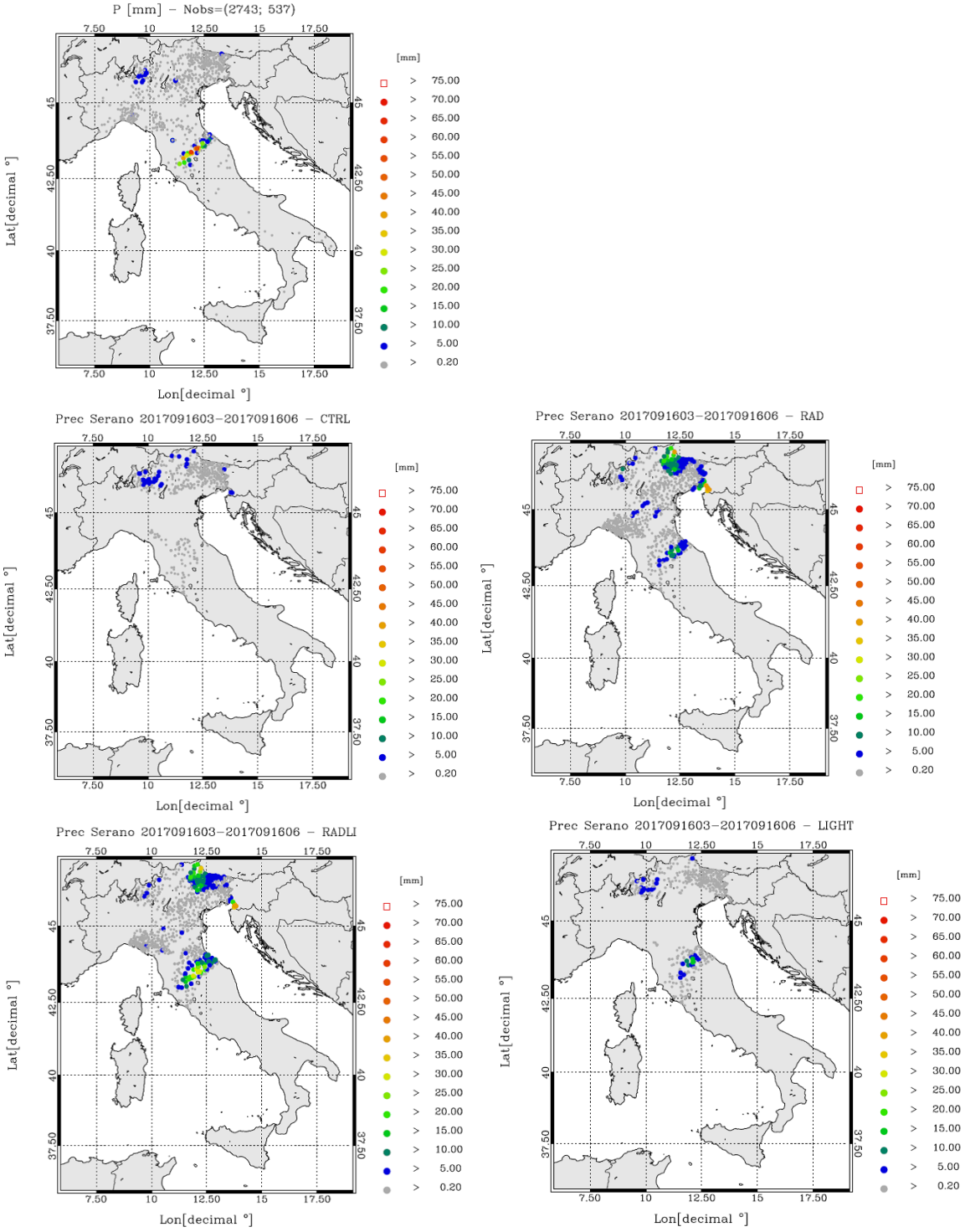
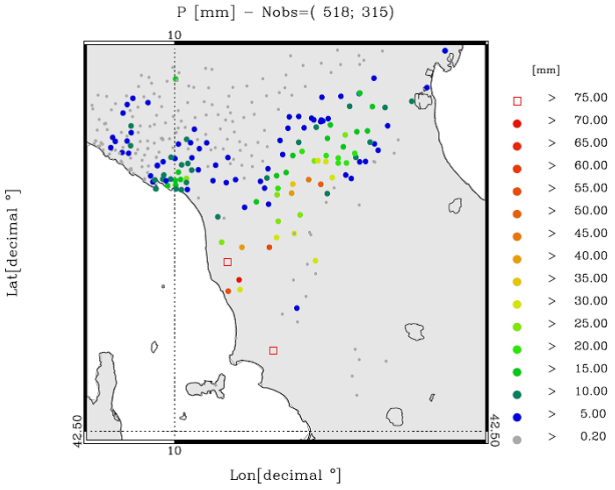
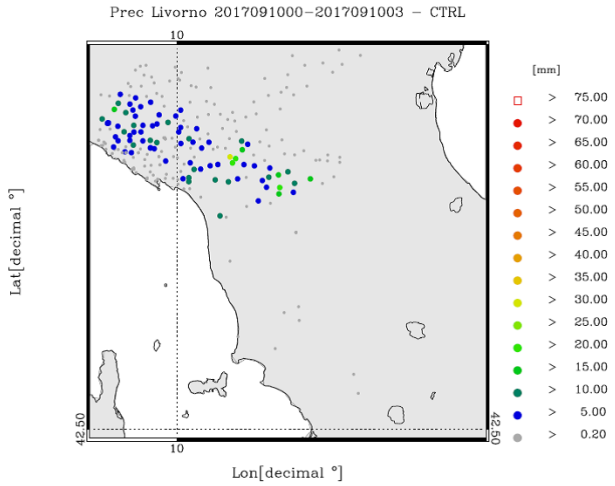
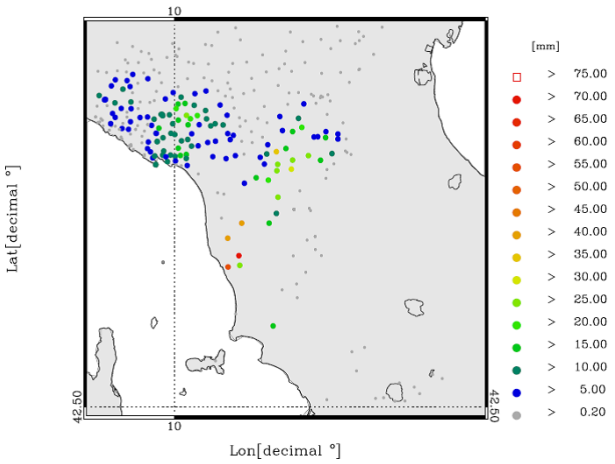


Figure S1: a) rainfall reported by raingauges between 03 and 06 UTC on 16 September 2017. Only raingauges observing at least 0.2 mm/day are shown. The first number in the title within brackets represents the available raingauges, while the second number represents those observing at least 0.2 mm/3h; b) as in a) for the CTRL forecast; c) as in a) for the RAD forecast; d) as in a) for the LIGHT forecast; e) as in a) for the RADLI forecast.

Livorno 2017-09-10 00-03 - Domain D3


P [mm] – Nobs=(518; 315)
Lat[decimal °]
Lon[decimal °]
10
42.50
[mm]
> 75.00
> 70.00
> 65.00
> 60.00
> 55.00
> 50.00
> 45.00
> 40.00
> 35.00
> 30.00
> 25.00
> 20.00
> 15.00
> 10.00
> 5.00
> 0.20


Prec Livorno 2017091000–2017091003 – CTRL
Lat[decimal °]
Lon[decimal °]
10
42.50
[mm]
> 75.00
> 70.00
> 65.00
> 60.00
> 55.00
> 50.00
> 45.00
> 40.00
> 35.00
> 30.00
> 25.00
> 20.00
> 15.00
> 10.00
> 5.00
> 0.20


Prec Livorno 2017091000–2017091003 – RAD
Lat[decimal °]
Lon[decimal °]
10
42.50
[mm]
> 75.00
> 70.00
> 65.00
> 60.00
> 55.00
> 50.00
> 45.00
> 40.00
> 35.00
> 30.00
> 25.00
> 20.00
> 15.00
> 10.00
> 5.00
> 0.20

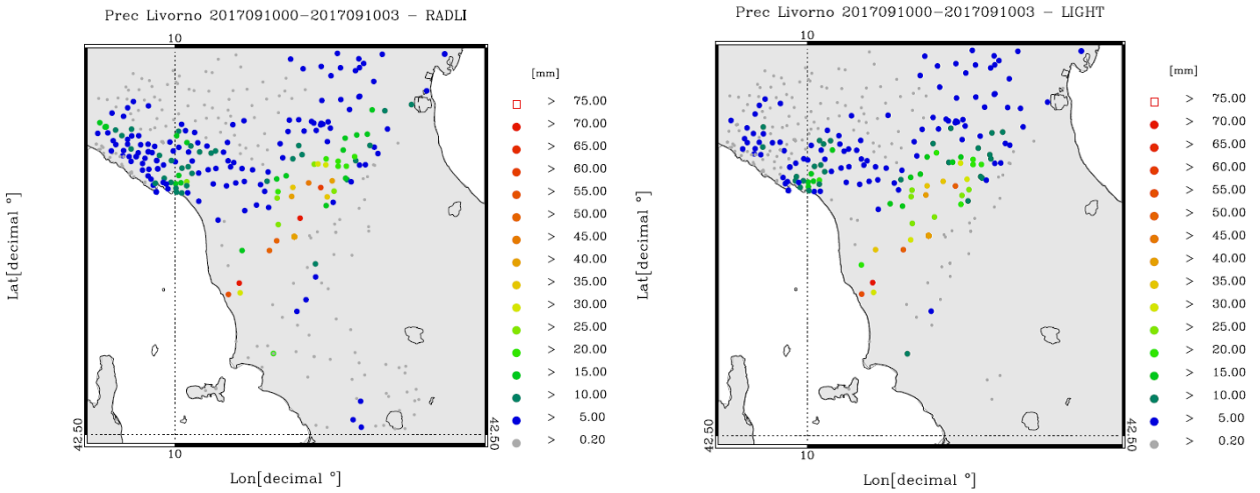
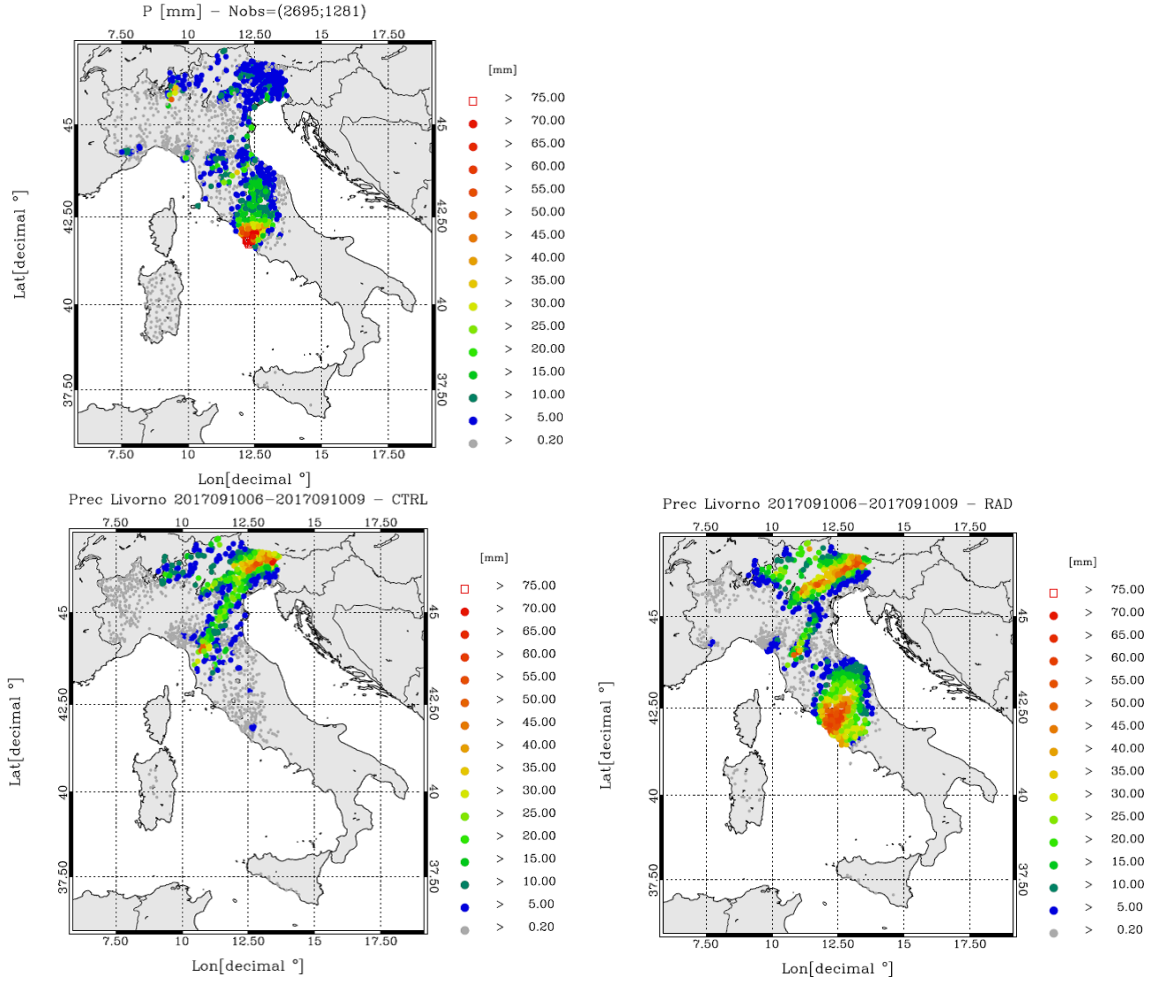


Figure S2: a) rainfall reported by raingauges between 00 and 03 UTC on 10 September 2017. Only stations reporting at least 0.2 mm/3h are shown. The first number in the title within brackets represents the number of raingauges available over the domain, while the second number shows those observing at least 0.2 mm/3h; b) as in a) for the CTRL forecast; c) as in a) for the RAD forecast; d) as in a) for the LIGHT forecast; e) as in a) for the RADLI forecast.

Livorno 2017-09-16 03-06 Domain D2



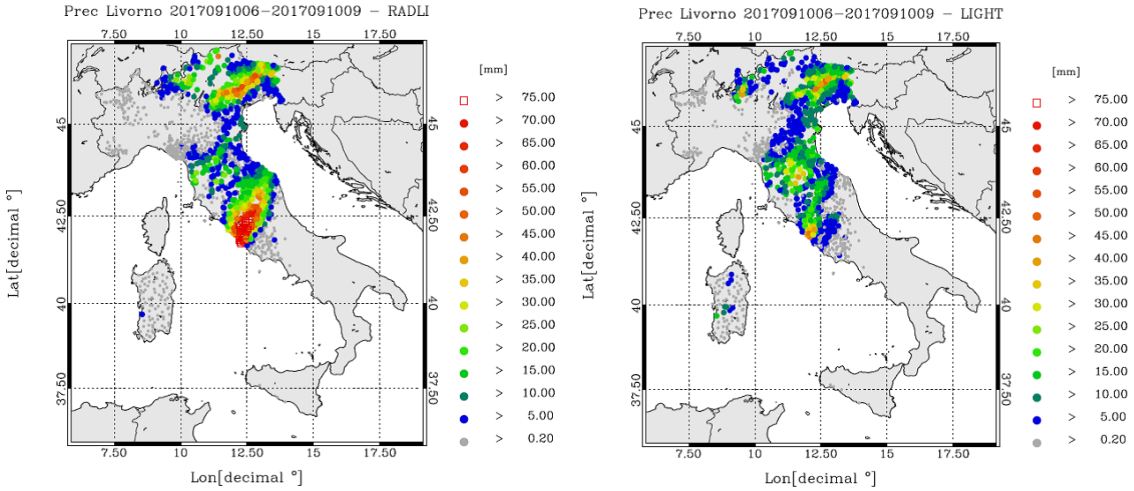


Figure S3: a) rainfall reported by raingauges between 06 - 09 UTC on 10 September 2017. For this time period 2695 raingauges reported valid observations in the domain, however only stations reporting at least 0.2 mm/3h are shown The first number in the title within brackets represents the number of raingauges available over the domain, while the second number shows those observing at least 0.2 mm/3h; b) as in a) for the CTRL forecast; c) as in a) for the RAD forecast; d) as in a) for the LIGHT forecast; g) as in a) for the RADLI forecast.

REFERENCES (it misses some title referenced by the reviewer)

Alexander, G. D., Weinman, J. A., Karyampoudi, V. M., Olson, W. S., and Lee, A. C. L.: The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 superstorm, Mon. Weather Rev., 127, 1433–1457, 1999.

Avolio,E., Federico,S. : WRF simulations for a heavy rainfall event in southern Italy: Verification and sensitivity tests, Atmospheric Research, Volume 209, 2018, Pages 14-35, ISSN 0169-8095, https://doi.org/10.1016/j.atmosres.2018.03.009.

Barker, D. M., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg ,S., Ajjaji, R., Bourgeois, A., Bray, J., Chen ,Y., Demirtas, M.,. Guo, Y.-R, Henderson, T., Huang, W, Lin, H.C., Michalakes, J., Rizvi, S., and Zhang, X.: The Weather Research and Forecasting (WRF) Model's Community Variational/Ensemble Data Assimilation System: WRFDA. Bull. Amer. Meteor. Soc., 93, 831–843, 2012.

Barker, D.M., Huang, W., Guo, Y.-R., and Xiao, Q.N.: A Three-Dimensional Variational Data Assimilation System For MM5: Implementation And Initial Results, Monthly Weather Review, 132, 897-914, 2004.

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation– forecast cycle: The RUC. Mon. Wea. Rev., 132, 495–518, doi:10.1175/1520-0493(2004)132,0495:AHACTR.2.0.CO;2.

Betz, H.-D., Schmidt, K., Laroche, P., Blanchet, P., Oettinger, P., Defer, E., Dziewit, Z., and Konarski, J.: LINET-an international lightning detection network in Europe, Atmos. Res., 91, 564– 573, 2009.

Betz, H. D., Schmidt, K., Oettinger, P.,Wirz, M. (2004), Lightning detection with 3D-discrimination of intracloudandcloud-to-grounddischarges. J.Geophys. Res. Lett. 31, L11108. doi:10.1029/2004GL019821.

Boccippio, D. J., K. L. Cummins, H. J. Christian, and S. J. Goodman, 2001: Combined satellite- and surface-based estimation of the intracloud–cloud-to-ground lightning ratio over the continental United States. Mon. Wea. Rev., 129, 108–122, doi:10.1175/1520-0493(2001)129,0108:CSASBE.2.0.CO;2

Carey, L. D., and S. A. Rutledge, 1998: Electrical and multiparameter radar observations of a severe hailstorm. J. Geophys. Res., 103, 13 979–14 000, doi:10.1029/97JD02626

Caumont, O., Ducrocq, V., Wattrelot, E., Jaubert, G., and Pradier-Vabre, S.: 1D+3DVar assimilation of radar reflectivity data: a proof of concept, Tellus A: Dynamic Meteorology and Oceanography, 62:2, 173-187, https://www.tandfonline.com/doi/abs/10.1111/j.1600-0870.2009.00430.x, 2010.

Chang, D. E., Weinman, J. A., Morales, C. A., and Olson, W. S.: The effect of spaceborn microwave and ground-based continuous lightning measurements on forecasts of the 1998 Groundhog Day storm, Mon. Weather Rev., 129, 1809–1833, 2001.

Deierling, W., and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. J. Geophys. Res., 113, D16210, doi:10.1029/2007JD009598.

Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansa, A., Kalthoff, N., Richard, E., Taupier-Letage, I., Ayral, P.-A., Belamari, S., Berne, A., Borga, M., Boudevillain, B., Bock, O., Boichard, J.-L., Bouin, M.-N., Bousquet, O., Bouvier, C., Chiggiato, J., Cimini, D., Corsmeier, U., Coppola, L., Cocquerez, P., Defer, E., Delanoë, J., Di Girolamo, P., Doerenbecher, A., Drobinski, P., Dufournet, Y., Fourrié, N., Gourley, J.J., Labatut, L., Lambert, D., Le Coz, J., Marzano, F.S., Molinié, G., Montani, A., Nord, G., Nuret, M., Ramage, K., Rison, W., Roussot, O., Said, F., Schwarzenboeck, A., Testor, P., Van Baelen, J., Vincendon, B., Aran, M., and Tamayo, J.: HYMEX-SOP1 The Field Campaign Dedicated to Heavy Precipitation and Flash Flooding in the Northwestern Mediterranean. Bull. Amer. Meteor. Soc., 95, 1083–1100, https://doi.org/10.1175/BAMS-D-12-00244.1 , 2014.

Emersic, C., and C. P. R. Saunders, 2010: Further laboratory investigations into the relative diffusional growth rate theory of thunderstorm electrification. Atmos. Res., 98, 327–340, doi:https://doi.org/10.1016/j.atmosres.2010.07.011.

Federico, S., Petracca, M., Panegrossi, G., and Dietrich, S.: 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, 2017a.

Federico, S., Petracca, M., Panegrossi, G., Transerici, C., and Dietrich, S.: Impact of the assimilation of lightning data on the precipitation forecast at different forecast ranges. Adv. Sci. Res., 14, 187–194, 2017b.

Federico, S.: Implementation of a 3D-Var system for atmospheric profiling data assimilation into the RAMS model: Initial results, Atmospheric Measurement Techniques, 6(12), 3563-3576, 2013.

Ferretti, R., Pichelli, E., Gentile, S., Maiello, I., Cimini, D., Davolio, S., Miglietta, M. M., Panegrossi, G., Baldini, L., Pasi, F., Marzano, F. S., Zinzi, A., Mariani, S., Casaioli, M., Bartolini, G., Loglisci, N., Montani, A., Marsigli, C., Manzato, A., Pucillo, A., Ferrario, M. E., Colaiuda, V., and Rotunno, R.: Overview of the first HyMeX Special Observation Period over Italy: observations and model results, Hydrol. Earth Syst. Sci., 18, 1953–1977, https://doi.org/10.5194/hess-18-1953-2014, 2014.

Fierro, A. O., M. S. Gilmore, E. R. Mansell, L. J. Wicker, and J. M. Straka, 2006: Electrification and lightning in an idealized boundary-crossing supercell simulation of 2 June 1995. Mon. Wea. Rev., 134, 3149–3172, doi:10.1175/MWR3231.1.

Fierro, A. O., Mansell, E., Ziegler, C., and MacGorman, D.: Application of a lightning data assimilation technique in the WRFARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011, Mon. Weather Rev., 140, 2609–2627, 2012

Fierro, A. O., J. Gao, C. Ziegler, E. R. Mansell, D. R. MacGorman, and S. Dembek, 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. Wea. Rev., 142, 183–202, doi:10.1175/ MWR-D-13-00142.1..

Fierro, A. O., A. J. Clark, E. R. Mansell, D. R. MacGorman, S. Dembek, and C. Ziegler, 2015: Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. Mon. Wea. Rev., 143, 757–777, doi:https://doi.org/10.1175/MWR-D-14-00183.1.

Fierro, A.O., Gao, l., Ziegler, C. L., Calhoun, K. M., Mansell, E. R., and MacGorman, D. R.: Assimilation of Flash Extent Data in the Variational Framework at Convection-Allowing Scales: Proof-of-Concept and Evaluation for the Short-Term Forecast of the 24 May 2011 Tornado Outbreak. Mon. Wea. Rev., 144, 4373–4393,https://doi.org/10.1175/MWR-D-16-0053.1, 2016.

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). Atmos. Res., 125–126, 34–49, doi:10.1016/j.atmosres.2013.01.006.

Kuhlman, K. M., C. L. Zielger, E. R. Mansell, D. R. MacGorman, and J. M. Straka, 2006: Numerically simulated electrification and lightning of the 29 June 2000 STEPS supercell storm. Mon. Wea. Rev., 134, 2734–2757, doi:10.1175/MWR3217.1

Kummerow, C., Hong, Y., Olson, W.S., Yang. S., Adler, R.F., McCollum, J., Ferraro, R., Petty, G., Shin. D.-B., and Wilheit, T.T.: The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. J. Appl. Meteor., 40, 1801–1820, 2001.

Lagouvardos, K., Kotroni, V., Betz, H.-D., and Schmidt, K.: A comparison of lightning data provided by ZEUS and LINET networks over Western Europe, Nat. Hazards Earth Syst. Sci., 9, 1713–1717, https://doi.org/10.5194/nhess-9-1713-2009, 2009

Lynn, B. H., G. Kelman, and G. Ellrod, 2015: An evaluation of the efficacy of using observed lightning to improve convective lightning forecasts. Wea. Forecasting, 30, 405-423 doi:10.1175/ WAF-D-13-00028.1 .

MacGorman, D. W. Burgess, V. Mazur, W. D. Rust, W. L. Taylor, and B. C. Johnson, 1989: Lightning rates relative to tornadic storm evolution on 22 May 1981. J. Atmos. Sci., 46, 221–251, doi:10.1175/ 1520-0469(1989)046,0221:LRRTTS.2.0.CO;2.

MacGorman, W. D. Rust, P. Krehbiel, W. Rison, E. Bruning, and K. Wiens, 2005: The electrical structure of two supercell storms during STEPS. Mon. Wea. Rev., 133, 2583–2607, doi:10.1175/MWR2994.1.

MacGorman, I. R. Apostolakopoulos, N. R. Lund, N. W. S. Demetriades, M. J. Murphy, and P. R. Krehbiel, 2011: The timing of cloud-to-ground lightning relative to total lightning activity. Mon. Wea. Rev., 139, 3871–3886, doi:10.1175/MWR-D-11-00047.1

Marchand, M., and H. Fuelberg, 2014: Assimilation of lightning data using a nudging method involving low-level warming. Mon. Wea. Rev., 142, 4850–4871, doi:10.1175/MWR-D-14-00076.1.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B. A.: Does increasing horizontal resolution produce more skilful forecasts?, B. Am. Meteorol. Soc., 83, 407–430, 2002.

Mittermaier, M., N. Roberts, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. Meteor. Appl., 20, 176–186, doi:https://doi.org/10.1002/met.296.

Olson, W. S., Kummerow, C. D. , Heymsfield, G. M., and Giglio, L.: A method for combined passive-active microwave retrievals of cloud and precipitation profiles. J. Appl. Meteor., 35, 1763-1789, 1996.

Papadopoulos, A., Chronis, T. G., and Anagnostou, E. N.: Improving convective precipitation forecasting through assimilation of regional lightning measure-ments in a mesoscale model, Mon. Weather Rev., 133, 1961–1977, 2005.

Parrish, D.F., and Derber, J.C.: The National Meteorological Center's Spectral Statistical Interpolation analysis system, Monthly Weather Review, 120, 1747-1763, 1992.

Pessi, A. T., and S. Businger, 2009: The impact of lightning data assimilation on a winter storm simulation over the North Pacific Ocean. Mon. Wea. Rev., 137, 3177–3195, doi:10.1175/ 2009MWR2765.1

Petracca, M., L. P. D'Adderio, F. Porcù, G. Vulpiani, S. Sebastianelli, and S. Puca, 2018: Validation of GPM Dual-Frequency Precipitation Radar (DPR) rainfall products over Italy. J. Hydrometeor., 19, 907–925, https://doi.org/10.1175/JHM-D-17-0144.1

Ricciardelli, E.; Di Paola, F.; Gentile, S.; Cersosimo, A.; Cimini, D.; Gallucci, D.; Geraldi, E.; Larosa, S.; Nilo, S.T.; Ripepi, E.; Romano, F.; Viggiano, M. Analysis of Livorno Heavy Rainfall Event: Examples of Satellite-Based Observation Techniques in Support of Numerical Weather Prediction. Remote Sens. 2018, 10, 1549.

Rohn, M., Kelly, G., Saunders, R. W.: Impact of a New Cloud Motion Wind Product from Meteosat on NWP Analyses and Forecasts, Monthly Weather Review, 129, 2392-2403, 2001.

Stensrud, D. J., and Fritsch, J. M.: Mesoscale convective systems in weakly forced large-scale environments. Part II: Generation of a mesoscale initial condition, Mon. Weather Rev., 122, 2068-2083, 1994.

Takahashi, T., 1978: Riming electrification as a charge generation mechanism in thunderstorms. J. Atmos. Sci., 35, 1536–1548, doi:https://doi.org/10.1175/1520 0469(1978)0352.0.CO;2.

Vulpiani, G., A. Rinollo, S. Puca, and M. Montopoli, 2014: A quality-based approach for radar rain field reconstruction and the H-SAF precipitation products validation. Proc. Eighth European Radar Conf., Garmish-Partenkirchen, Germany, ERAD, Abstract 220, 6 pp., http://www.pa.op.dlr.de/erad2014/programme/ ExtendedAbstracts/220_Vulpiani.pdf (last access January 2019).

Wiens, K. C., S. A. Rutledge, and S. A. Tessendorf, 2005: The 29 June 2000 supercell observed during STEPS. Part II: Lightning and charge structure. J. Atmos. Sci., 62, 4151–4177, doi:10.1175/JAS3615.1