Author comments to the reviewer 3 comments

We acknowledge the reviewer for the useful and constructive comments.

This paper presents results of implementation of a previously published lightning scheme into the RAMS model, and its application to several case studies of thunderstorms in central and northern Italy. It is certainly within the topics which NHESS deals with, and is well within the international standards that the journal aims for.
The paper is well-written, logically organized, thorough and lucid. The graphs are clear and properly show the main results, and the tables nicely summarize all the storms that were studied and the skills of the model. The authors discuss the results of 2 case studies in depth, and evaluate the performance of the model and the lightning prediction scheme, and offer possible reasons for the inconsistent accuracy for different cases. They also set a clear course for future development and it is likely that the insights gained in the initial study will be used to improve the performance of the RAMS and make it an operational tool for lightning forecasts, and to gauge the nature of convection based on offsets in the model-predicted and observed lightning activity. I have some comments and questions which the authors need to consider. Apart from some minor revisions, I recommend that this paper is accepted for publication in the NHESS.

Major comments

1. The Dahl et al. (2010) scheme over-simplifies the complex charge structure of thunderstorms and portrays it as two plates at different altitudes and temperature regimes within the cloud, each composed of a different class of particles and assigned with a different polarity. This should be clearly mentioned in section 2.2, and compared with "real-life" charge structures, such as those reported by Marshall et al. (JGR, 1995) and Stolzenburg et al. (1994). This can be one of the sources of the discrepancies found between the computed lightning activity and the observed one. Also, assigning a "fixed" polarity to each species is not accurate, as many laboratory experiments show (reviewed in Saunders, 2008), and certainly using a constant lightning efficiency of 0.9 all throughout the storm life-cycle is inaccurate. In the discussion section the authors need to evaluate their results in light of these shortcomings of the scheme (perhaps in p. 22, after lines 11-22).

- Thank you for pointing this out. We are aware that assuming two charged regions with fixed polarity is a simplification that may have a significant impact on the results. This will be clearly mentioned in Section 2.2. Moreover, the discussion about the drawbacks of the lightning scheme will be extended in the discussion section of the paper to include these issues and some additional points suggested by reviewer #2. The Conclusions section 4 that will be renamed “Discussion and Conclusions”.

2. Additional factors that may play into the performance of the lightning scheme is the time interval between two calls of the scheme – 10 minutes (p.10, lines 7-9) seem to be a very long time. Since the authors claim that the scheme is computationally efficient (p.22, lines 31-33), one wonders what shortening this interval will do in terms of efficiency and would it be critical. Have the authors conducted a sensitivity study for the performance of the scheme with varying times? If not, it is highly desirable to run al least one of the case studies (preferably the October 15th 2012 case, which had rather poor agreement in timing and total lightning densities).

- All the simulations and results, in the revised version of the paper, will be considered using a time interval of 5 minutes (T5), instead of 10 minutes (T10), between two calls of the lightning scheme. The Tables containing the scores of the discussion paper (Table 2 of the discussion paper with ETS replacing TS, as required by the reviewers 1 and 3), and the scores when using a time interval of 5 minutes between two calls of the lightning scheme are shown at the end of this authors' response. The differences between T10 and T5 are small, but not negligible. Moreover, calling the lightning scheme every 5 minutes improved the results (see the ETS values in the Tables T10 and T5 below), as expected.

3. The difference in model performance for the two case studies is puzzling and the authors rightly discuss it. However they seem to place the reason with the failure of the RAMS to correctly portray the convection, and hence the development of the thunderstorm cells and the ensuing lightning activity. This specific issue should be further discussed and there needs to be a suggested explanation: is this because the initialization was inaccurate? What in the synoptic condition caused such a difference? Why is the onset of convection on the October 15th 2012 case "missed by the model"? (p.16, line 15).

- As suggested by the reviewer there is a combination of reasons for the less satisfactory simulation of the 15 October 2012 case, compared to the 20 October 2011. In particular: weaknesses in the RAMS model, initial and boundary conditions, absence of local observations useful to trigger convection. It is also important to remark that the 15 October 2012 event was less intense than other cases considered in this paper, at least over the Lazio region, and was characterized by weak convective cells, whose simulation is difficult because the weak forcing causing them can be missed by the model.
  The onset of the convection during the night between 14 and 15 October 2012, for example, was caused by small convective cells (as evidenced by the LINET observations shown below, Figure 1, blue and green dots), over the sea and near the coast. Likely, this could have been simulated nudging local observations (not available) into the RAMS.
  Another important point to remark is that the convergence generated by the model at the surface (Figure 7 of the discussion paper), which is a key factor to initiate the convection, is close to the coastline and over the southern part of the Lazio region, where most of flashes are simulated (Figure 9b of the discussion paper). The development of the convergence zone near the coastline is caused by the flow deflection induced by the orography, which interacts with the undisturbed flow over the Sea. The absence of convergence over the sea is a key factor for the missing of the two bands of flashes over the sea shown by the LINET (Figure 9a of the discussion paper) in the afternoon and evening of the 15 October.
  The misplaced pattern of surface convergence simulated by the model is in turn the result of several factors: a) initial and dynamic boundary conditions, which can influence the specific model behaviour through the Froude number, or the lack of humidity, or absence of convergence lines; b) model deficiencies both physical and dynamical, and; c) the absence of local observations, which could trigger the convective cells.
  The above discussion will be included in the revised version of the paper in Section 3.2.

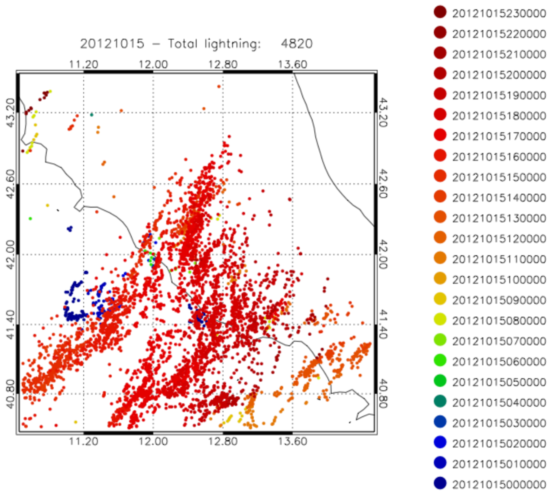


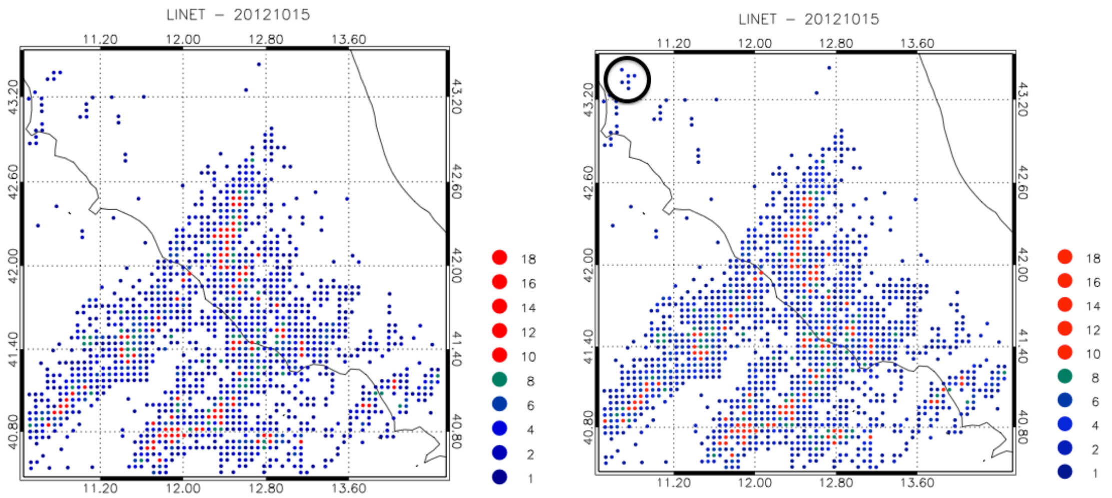
Figure 1. LINET flashes for the 15 October 2013 case. The colour bar shows the time of occurrence of the flashes.

4. In their explanation of the persistence of lightning long after the actual end of convection (Figure 10, and p. 17, lines 3-9) the authors suggest that there is a problem with the sedimentation scheme of the RAMS for the cell residing over the sea. This explanation needs further justification, because it is not entirely clear how the surface properties affect the microphysics aloft so strongly.

- Actually the explanation provided in the paragraph (p. 17, lines 3-9) is limited and the sentence about the problem with the sedimentation scheme is misleading. The all paragraph will be modified pointing out the points outlined in the answer to point 3 (previous page). What we meant about the sedimentation scheme of RAMS is that any weakness would be more evident over the sea, compared to the land, because of the absence of orographic forcing on the convection.

5. The binning of lightning data from the LINET network assumes that all strokes occurring within 1 second and 10 km radius from an initial strike point are defined as a single flash (p. 11, lines 25-27). These are the conservative values used by the NLDN. Have the authors considered other thresholds for grouping strokes into flashes? (For example, in a recent paper Yair et al. (NHESS, 2013) showed that using stricter ranges (5 km, 0.5 s) is sufficient to discriminate between successive flashes in most cases). Could it perhaps improve the skill if the actual numbers reported by the number were different?

- As pointed out by the reviewer, we designed as a single flash all the strokes observed within 10 km and 1 s from an initial strike point. The use of stricter ranges (5 km , 0.5 s instead of 10 km and 1 s) for the binning of the strokes has two effects on the results: a) increase of the number of total lightning, which spread over a wider area of the verification domain (see, for example, the area inside the circle in the two figures below, Figure 2); b) increase of the number of the cells in the verification domain with more than 10 flashes per day (MLT10, see the red dots in Figure 2).



a) b)

Figure 2: Flashes observed for the 15 October 2012 case study: a) choosing larger ranges (10km , 1 s) for binning the strokes from an initial strike point recorded by LINET; b) as in a) but choosing stricter ranges (5km , 0.5 s).

From Figure 2 it is evident that the differences in the results obtained for the two binning criteria (10 km, 1 s ; 5km, 0.5 s) are small, nevertheless choosing stricter ranges (5km , 0.5 s) gives better scores in terms of higher hits and lower false alarms and misses. Table T5b, at the end of this authors' response, shows the scores obtained using the stricter ranges. The largest differences between the results of tables T5 and T5b are in the bias score (sometimes better sometimes worse). Nevertheless, it should be noticed the slight increase of the ETS values for the T5b table (i.e. using the stricter ranges of 5 km and 0.5 s) compared to T5 (i.e. using the conservative ranges of 10 km and 1 s). These differences are less significant than those found when comparing T10 (calling the lightning scheme every 10 minutes) and T5 (calling the lightning scheme every 5 minutes).

In the paper of Dahl et al. (2011b) (referred to in the discussion paper) the 10 km and 1 s thresholds are used for grouping the LINET strokes. Therefore, we prefer to maintain the 10 km and 1 s as binning thresholds in our study, as the results will be more comparable with those presented in Dahl et al. (2011b), however, a comment about the impact of choosing the stricter ranges (5 km and 0.5 s) will be included in the sections "3.3 Scores", after presenting the results of Figure 13, and a table showing the results for the 20 October 2011 and 15 October 2012 cases (i.e. a portion of the Table T5b) will be added. We believe that this approach will give more emphasis to this important issue raised by the reviewer.

Minor comments

1. P. 4, line 18: suggest including Barth et al., 2010 2. P.4, line 20 and throughout: intra cloud -> intracloud 3. P.10, lines 23-27: please explain how the distribution of the lightning is made to follow the shape of the convective cell. How do you determine the shape? Is it contours of certain values of graupel concentrations, radar reflectivity or. . .? 4. P. 19, line 15: studies that occurred in fall 2012. . . 5. P. 21, line 22: six case studies that occurred over the. . . 6. P. 21, line 23: The fist one occurred. . ..

- These corrections will be included in the revised version of the paper.

T10: Skill score statistics of the six case studies. Date of forecast and number of flashes observed (LINET) and simulated (RAMS) for each case study are shown in the first column. POD, FAR, Bias, and ETS are given for the MLT1 and MLT10 (in parentheses) for the 25 km, 12.5 km and 5 km overlays superimposed to the 2.5 km RAMS grid. The area considered for the statistics is the area shown in Fig. 4 (10.5 – 14.5 E, 40.5 – 43.5 N). The spatial and temporal thresholds used for the binning of the strokes from an initial strike point recorded by LINET are 10 km and 1 s, respectively. The time interval between two calls of the lightning scheme is 10 minutes.

	25 km overlay				**12.5 km overlay**				**5 km overlay**			
Case study	POD	FAR	Bias	ETS	POD	FAR	Bias	ETS	POD	FAR	Bias	ETS
20111020 LINET: 16231 RAMS: 18631	0.73 (0.76)	0.21 (0.17)	0.92 (0.91)	0.46 (0.57)	0.66 (0.68)	0.18 (0.17)	0.80 (0.82)	0.47 (0.54)	0.58 (0.46)	0.18 (0.51)	0.71 (0.95)	0.45 (0.28)
20120903 LINET: 6666 RAMS: 6496	0.85 (0.58)	0.09 (0.33)	0.93 (0.88)	0.63 (0.30)	0.68 (0.43)	0.25 (0.55)	0.91 (0.97)	0.38 (0.21)	0.43 (0.18)	0.56 (0.84)	0.99 (1.07)	0.18 (0.01)
20120930 LINET: 7073 RAMS: 7635	0.90 (0.85)	0.13 (0.32)	1.03 (1.25)	0.68 (0.48)	0.79 (0.46)	0.27 (0.62)	1.08 (1.20)	0.46 (0.18)	0.53 (0.05)	0.54 (0.96)	1.16 (1.01)	0.22 (0.01)
20121015 LINET: 4820 RAMS: 6554	0.80 (0.68)	0.20 (0.38)	1.00 (1.09)	0.53 (0.39)	0.70 (0.40)	0.29 (0.67)	0.99 (1.23)	0.42 (0.15)	0.42 (0.01)	0.58 (0.99)	1.01 (1.97)	0.19 (0.00)
20121111 LINET: 9030 RAMS: 12308	0.79 (0.76)	0.05 (0.05)	0.84 (0.80)	0.69 (0.67)	0.73 (0.65)	0.06 (0.25)	0.78 (0.87)	0.61 (0.47)	0.62 (0.31)	0.24 (0.80)	0.81 (1.53)	0.44 (0.12)
20121128 LINET: 14357 RAMS: 13842	0.90 (0.79)	0.02 (0.06)	0.92 (0.84)	0.81 (0.64)	0.80 (0.46)	0.06 (0.29)	0.85 (0.65)	0.63 (0.27)	0.52 (0.15)	0.25 (0.83)	0.69 (0.94)	0.31 (0.01)

T5: Skill score statistics of the six case studies. Date of forecast and number of flashes observed (LINET) and simulated (RAMS) for each case study are shown in the first column. POD, FAR, Bias, and ETS are given for the MLT1 and MLT10 (in parentheses) for the 25 km, 12.5 km and 5 km overlays superimposed to the 2.5 km RAMS grid. The area considered for the statistics is the area shown in Fig. 4 (10.5 – 14.5 E, 40.5 – 43.5 N). The spatial and temporal thresholds used for the binning of the strokes from an initial strike point recorded by LINET are 10 km and 1 s, respectively. The time interval between two calls of the lightning scheme is 5 minutes.

	25 km overlay				**12.5 km overlay**				**5 km overlay**			
Case study	POD	FAR	Bias	ETS	POD	FAR	Bias	ETS	POD	FAR	Bias	ETS
20111020 LINET: 16231 RAMS: 16002	0.65 (0.74)	0.08 (0.09)	0.72 (0.82)	0.50 (0.62)	0.64 (0.65)	0.08 (0.12)	0.70 (0.75)	0.51 (0.55)	0.58 (0.38)	0.15 (0.55)	0.68 (0.84)	0.46 (0.23)
20120903 LINET: 6666 RAMS: 5844	0.79 (0.59)	0.08 (0.32)	0.86 (0.88)	0.58 (0.31)	0.64 (0.43)	0.25 (0.58)	0.86 (1.03)	0.35 (0.19)	0.42 (0.17)	0.57 (0.81)	0.98 (0.88)	0.17 (0.01)
20120930 LINET: 7073 RAMS: 8080	0.90 (0.90)	0.17 (0.32)	1.10 (1.33)	0.63 (0.50)	0.85 (0.51)	0.28 (0.64)	1.18 (1.40)	0.49 (0.18)	0.61 (0.05)	0.55 (0.94)	1.36 (0.84)	0.23 (0.02)
20121015 LINET: 4820 RAMS: 6794	0.77 (0.65)	0.15 (0.28)	0.91 (0.90)	0.56 (0.43)	0.62 (0.34)	0.26 (0.67)	0.84 (1.02)	0.40 (0.14)	0.38 (0.06)	0.55 (0.97)	0.85 (2.35)	0.18 (0.01)
20121111 LINET: 9030 RAMS: 13626	0.80 (0.78)	0.10 (0.07)	0.90 (0.84)	0.65 (0.66)	0.77 (0.75)	0.09 (0.26)	0.85 (1.02)	0.63 (0.53)	0.71 (0.31)	0.25 (0.82)	0.93 (1.72)	0.49 (0.13)
20121128 LINET: 14357 RAMS: 16152	0.90 (0.89)	0.02 (0.03)	0.91 (0.91)	0.81 (0.80)	0.86 (0.62)	0.05 (0.30)	0.90 (0.89)	0.70 (0.36)	0.67 (0.17)	0.26 (0.86)	0.92 (1.25)	0.40 (0.06)

T5b : Skill score statistics of the six case studies. Date of forecast and number of flashes observed (LINET) and simulated (RAMS) for each case study are shown in the first column. POD, FAR, Bias, and ETS are given for the MLT1 and MLT10 (in parentheses) for the 25 km, 12.5 km and 5 km overlays superimposed to the 2.5 km RAMS grid. The area considered for the statistics is the area shown in Fig. 4 (10.5 – 14.5 E, 40.5 – 43.5 N). The spatial and temporal thresholds used for the binning of the strokes from an initial strike point recorded by LINET are 5 km and 0.5 s, respectively. The time interval between two calls of the lightning scheme is 5 minutes.

	25 km overlay				**12.5 km overlay**				**5 km overlay**			
Case study	POD	FAR	Bias	ETS	POD	FAR	Bias	ETS	POD	FAR	Bias	ETS
20111020 LINET: 18373 RAMS: 16002	0.64 (0.74)	0.09 (0.09)	0.70 (0.82)	0.49 (0.62)	0.63 (0.65)	0.08 (0.10)	0.69 (0.72)	0.51 (0.55)	0.57 (0.39)	0.13 (0.47)	0.66 (0.75)	0.46 (0.26)
20120903 LINET: 7291 RAMS: 5844	0.79 (0.59)	0.08 (0.32)	0.86 (0.87)	0.58 (0.30)	0.64 (0.42)	0.25 (0.56)	0.85 (0.97)	0.36 (0.20)	0.42 (0.17)	0.55 (0.79)	0.93 (0.77)	0.18 (0.01)
20120930 LINET: 8055 RAMS: 8080	0.91 (0.90)	0.16 (0.32)	1.08 (1.33)	0.64 (0.50)	0.85 (0.52)	0.26 (0.60)	1.15 (1.29)	0.50 (0.20)	0.62 (0.05)	0.51 (0.93)	1.27 (0.69)	0.25 (0.02)
20121015 LINET: 5337 RAMS: 6794	0.77 (0.64)	0.15 (0.26)	0.91 (0.87)	0.56 (0.44)	0.62 (0.33)	0.25 (0.65)	0.83 (0.96)	0.40 (0.14)	0.38 (0.06)	0.53 (0.97)	0.80 (1.98)	0.19 (0.02)
20121111 LINET: 11155 RAMS: 13626	0.80 (0.77)	0.10 (0.07)	0.90 (0.83)	0.65 (0.65)	0.77 (0.74)	0.09 (0.24)	0.85 (0.98)	0.63 (0.54)	0.70 (0.32)	0.21 (0.76)	0.89 (1.32)	0.51 (0.13)
20121128 LINET: 17100 RAMS: 16152	0.90 (0.89)	0.02 (0.02)	0.91 (0.91)	0.81 (0.81)	0.86 (0.62)	0.04 (0.27)	0.90 (0.84)	0.72 (0.37)	0.67 (0.16)	0.23 (0.83)	0.87 (0.93)	0.41 (0.06)