Nonlin. Processes Geophys. Discuss., 1, C384–C401, 2014
www.nonlin-processes-geophys-discuss.net/1/C384/2014/

Nonlinear Processes in Geophysics Discussions
Open Access

Interactive comment on "Hybrid variational-ensemble assimilation of lightning observations in a mesoscale model" *by* K. Apodaca et al.

K. Apodaca et al.

karina.apodaca@colostate.edu

Received and published: 22 August 2014

Development of a hybrid variational-ensemble data assimilation technique for observed lightning tested in a mesoscale model

Karina Apodaca1, Milija Zupanski1, Mark DeMaria2,*, John A. Knaff2, and Lewis D. Grasso1

[1]{Colorado State University/Cooperative Institute for Research in the Atmosphere, Fort Collins, Colorado, USA} [2]{NOAA Center for Satellite Research and Applications, Fort Collins, Colorado, USA} [*]{now at: Technology and Science Branch, National Hurricane Center, Miami, Florida, USA} Correspondence to: K. Apodaca (ka-

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

rina.apodaca@colostate.edu)

Answers to reviewer's comments

Reviewer no. 1

GENERAL COMMENTS The manuscript describes the application of a (previously documented) hybrid ensemble-variational assimilation method to observations of lightning rate from a surface lightning detection network, intended as a proxy of a future satellite observing instrument, which should be able to provide similar data in areas presently uncovered, such as the oceans. The assimilation of lightning rate data is applied to a model where convection is parameterised, so it is aimed at improving, rather than the description of convection processes, that of convection "environment", i. e. values of model prognostic variables locally involved in the computation of model-estimated lightning rate, through the estimated vertical velocity (itself not a prognostic variable in this parameterised convection system). After the definition of the observation operator, described in some detail in the appendix, the application of the assimilation method is straightforward. The test case is an important event of a front impacting on South-Eastern and Eastern U.S.A., causing many convection episodes associated with heavy rain, wind gusts and tornadoes. The motivation, context and development of this work appear relevant and well described. However the results appear much weaker than what is suggested in the manuscript title and what is commented in the body of the text and in the conclusions. Indeed it is appropriately shown that, at analysis time, the lightning rate observations are able to provide information and that this is transferred to model state variables. The interpretation of the effect of assimilation on forecast, though, is too optimistic. My indication is that a major revision is needed. The authors should revise the manuscript, including the title, and present their results under a much more cautious shade. An improvement in the presentation of results is also desirable to make their interpretation clearer.

Response: We would like to thank the reviewer for his/her valuable input. Revisions

NPGD

1, C384–C401, 2014

to the manuscript were done as suggested, in the “Abstract”, “Results,” and “Summary and Conclusions.” An alternative title is now proposed: Development of a hybrid variational-ensemble data assimilation technique for observed lightning tested in a mesoscale model

Minor Points:

1. P. 923, Sect. 2.2. Please motivate the choice of computing observed lightning rates on a grid different from the model grid.

Following the Reviewer’s suggestion, a paragraph has been added to Sect. 2.2 (after “. . . model configuration that will be discussed in Sect. 3.2.”: “The choice of a regular grid that is not identical to the model grid is arbitrary. In our case, it was motivated by a desire to keep the observation information formally independent from the model, i.e. to not use any information about the model when defining observations and observation errors.”

2. * P.923, line 24-26. This choice should be synthetically motivated. In areas interested by convection (in either the model or reality) and in locations nearby, observations of zero lightning might in principle represent information about misplaced convection events. Moreover: how does this choice affect the PDF of innovation vectors discussed in Sect. 2.3 ?

Although we agree with the Reviewer’s suggestion that zero lightning observations can be important in pointing the locations of misplaced convection events, it is not clear how this information would impact the convection events that are not characterized by strong lightning. It is likely that additional information would be needed in order to selectively define zero lightning observations. We feel that this information is important, but needs further investigation, and thus it is left for the future.

Since using strictly positive-definite lightning observations introduces a skewness that can be also seen as a positive bias in a typically used Gaussian PDF, it is possible that

NPGD
1, C384–C401, 2014

adding zero lightning observations would somewhat reduce the current observational bias. However, even with zero lightning the innovation vectors are still non-negative, thus observations will require a bias correction under the typical Gaussian assumptions of data assimilation. Probably, the most adequate solution is to treat the PDF of lightning observations as non-Gaussian, which is left for future investigation.

Modifications were made in the manuscript in P.923, L26.

3. * P.924, line 2-4 and Eq. 2. The definitions if h1 and h2 in line 2-4 are not consistent with Eq. 2 and with the equation (line 3) h= h1 h2. Which of the two is the "transformation"? Which is applied first?

The Reviewer is correct, we are thankful for pointing this out. Corrections were made to the manuscript accordingly. The corrected sentences are: L.2-4: "In this study the forward lightning transformation operator (h2) was adopted by exploiting the relationship between lightning and vertical velocity." L.8-9: "A bi-linear interpolation technique was used to interpolate the guess lightning flash rates to observation location (h1)."

4. * P.924, lines 8-9. Interpolation FROM observation points (the "grid" used to compute lightning rate as described in Sect. 2.2) TO model grid points? Since the definition of an observation operator goes from model state variables to observation estimates, I expect that the interpolation would go the same way. Did you instead choose to build pseudo-observations at model grid points? In this case this should be said in a more explicit and clear way (while motivating the initial choice of using a different grid, see above).

The Reviewer is correct, we only do a standard interpolation from the model grid to observation locations. The sentence now reads (also please see the answer to previous comment):

L.8-9: "A bi-linear interpolation technique was used to interpolate the guess lightning flash rates to observation location (h1)."

NPGD

1, C384–C401, 2014

5. * P.925 line 3. The abbreviation "CWM" has not been defined.

Thanks for pointing this out. CWM is now defined.

Cloud Water Mass (CWM – total cloud condensate in WRF-NMM)

6.* P.925 line 4. "...and neighbouring points": why? how many?

Considering the Reviewer's comments, the following sentence was added on L.4:

“We defined a 5 x 10 grid point area (approximately a square domain in Arakawa E-grid staggering used in WRF-NMM) surrounding the central point in order to introduce a smooth transition for the calculation of wmax .”

7. * P.925 lines 11 and following. How does the choice of discarding zero observations affect the innovation PDFs?

We believe that discarding zero observations would not dramatically change the skewed character of the PDF since the innovations are still a non-negative and thus incompatible with a typical Gaussian PDF assumption used in data assimilation. Adding zero innovations may impact the choice of a skewed PDF that allows zero innovation values, but not the need for bias correction in (standard) Gaussian data assimilation. (Also, please see the response to the comment (2))

8. * P.927 line 10 and line 17: "the number of ensembles". Do you mean "the number of ensemble members"?

Yes, thanks for noticing. Changes have been done in the manuscript: P.927 line 10: Replace “of ensembles” by “number of ensemble members” P.927 line 17: Replace “of ensembles” by “number of ensemble members”

9. * P.928 Sect 3.1. Please convert units, use m sËĘ-1 for wind and C for temperature.

Thank you, units have been converted as suggested. P.928, line 7, change: 80 kn to 41.15 m sec-1, line 9: 70°F to 21°C, and 15kn to 7.72 m sec-1

10. * P.928 Sect 3.1. The event description should be completed with a more detailed discussion of the evolution in time of the front and of the convective activity, and this should be later used in the discussion of results (Sect. 4) at different times, with regards to forecast improvements and deterioration.

We followed the Reviewer's suggestions. A more detailed event description has been done in the manuscript in Sect.3.1.

11. * P. 931 line 14. Why at 17:00 UTC? Please make explicit reference to the evolution of the event of Sect. 3.1 (after improving it, see above).

The time should read 18:00 UTC, thanks for spotting this. This time was chosen because tornados started developing over northern Alabama just a couple of hours before. The text has been changed accordingly.

12. * P 931, Sect.4.1. It would be useful to explicitly give the size of the spatial extension of the signals appearing in Fig. 6, and compare it with the model grid size.

Agreed, the following sentences have been included in Sect. 4.1.:

"The spatial extension of the impact of assimilating a single lightning strike on some of the dynamical variables of the model in D02 (9 km resolution) was: (i) on specific humidity the impact extends to approximately 12 grid points ($\sim$ 110 km), (ii) for temperature to 20 grid points ($\sim$ 180 km), and (iii) for wind approximately 30 grid points ($\sim$270 km). "

13. * P932 line 12-13. The agreement is not very good for Cycle 3 (middle panel of Fig. 7). Please acknowledge this. Moreover: the maxima of degrees-of-freedom-for-signal appear to be somewhat off with respect to the distribution of observed lightning rate. The authors should perhaps discuss how this is related to the effect of observations on model prognostic variables, with regard to the main flow.

Following the Reviewer's suggestion, the following paragraph has been included in Sect. 4.2.:

NPGD

1, C384–C401, 2014

Interactive Comment

“Note however that the agreement in cycle 3 was not very good, it is possible that ensemble perturbations where not large enough over northeastern Alabama where another maximum was missing. This lack of agreement can arise from the use of a reduced rank ensemble approach and consequently not having enough spread in the ensembles. However, the agreement improved in subsequent cycles (e.g., shown for cycle 5).

14. * P.932 line 18 and line 19: "the number of ensembles". Do you mean "the number of ensemble members"?

Thank you. “The number of ensembles” has been replaced by “number of ensemble members” on lines 18 and 19.

15. * P 933 Sect.4.3 You should include, in Figure 8 and 10, a panel with the same field (of 8a and 10a) for the NODA experiment. I think that the discussion would benefit from a LIGHT-NODA comparison (here too, not only in Sect. 4.4).

We followed the Reviewer’s suggestions and included the figures of wind and CAPE for the NODA experiment, made corrections to the description of results, and changed figure captions accordingly.

The following are the figure captions for Fig. 8. and Fig. 10.:

Fig. 8. (a) Background (forecast) winds at 850 hPa at 0000 UTC 28April 2011 (cycle 5) from the experiment without lightning (NODA), (b) background (forecast) winds at 850 hPa at 0000 UTC 28April 2011 (cycle 5) from the lightning data assimilation experiment (LIGHT) and (c) GOES IR and observed 6-hour WWLLN lightning flash rates at the same time (Courtesy of Gregory DeMaria and Jack Dostalek). The core of strong wind speed matches the region of high lightning flash rate density in the observations, but note that the core of maximum wind speed has a larger spatial coverage in the LIGHT experiment (b) and based on computed differences, stronger winds in the order of 4 m sec-1 were found in the LIGHT experiment.

NPGD
1, C384–C401, 2014

Fig. 10. Background CAPE for (a) NODA and (b) LIGHT experiments, and (c) observed CAPE from the Storm Prediction Center's Surface Mesoanalysis at 0000 UTC 28 April 2011 (cycle5). A region of high CAPE gradient is observed in the upper-left hand side of the domain, indicating the presence of a well-defined dry line, in agreement with observations, but there are no significant differences between both experiments. One reason is that there are no lightning observations in the region where the strongest CAPE was observed. Lightning data was not able to impact CAPE.

16. * Sect. 4.4. Please refer the assimilation cycles to time and event evolution

The suggested changes were made in Section 4.4.

17. * P 934 lines 6-12. In Fig 11-a, an improvement can be seen in 5 cases, however not for the 6th. Is it possible to investigate possible reasons for that?

The following paragraph was included in Sect. 4.4.:

"A possible reason for not seeing an improvement in the 6th cycle could be that the system was exiting the model domain at that time. Since the strongest convection and cold front moved away from the domain, there was no significant lightning activity over the region. Consequently, the number of lightning observations available for data assimilation significantly decreased and the impact of lightning data assimilation was reduced."

18. * P 934 line 8. From Fig 11-b, no systematic improvement can be seen in the 6-h forecast. What appears here is that there is some improvement in 2 cases out of 6, some deterioration in 1 case, and no change in 3 cases. I don't think this is enough to consider this a "partial improvement".

We agree with the Reviewer's assessment of Fig.11b and following the reviewer's suggestion, the sentence now reads: "This result is not very well retained in the forecast (Fig. 11b). This issue definitely requires further investigation."

19. * P 934 lines 9-12 . Supplementary observations should in principle improve

NPGD

1, C384–C401, 2014

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

the assimilation results, but they could mask the effect of assimilating lightning rate. Shouldn't this new observational source be useful to compensate for lack of information in under-observed areas?

We agree with the Reviewer's assessment. The lines below were added to Sec. 4.4.:

"Note that lightning is just an additional type of observation. All available observations have to be in agreement with each other at the same location. Therefore, in regions where lightning observations are not in agreement with other types of observations, the data assimilation algorithm will create the optimal observation impact based on uncertainty of all observations in the region. In areas where lightning observations are not available other measurements should help."

20. * P 935 line 14-15. From the results shown, it is not possible to affirm that the analysis improvement is partially retained in the forecast. Omitting in the conclusions that a short range forecast (+6h only) was considered is also misleading.

We agree with the Reviewer. The text now includes: "However, the 6 h forecast errors after assimilation did not show any clear improvement in terms of the RMS errors. This requires further investigation."

21. * P.941, reference list, line 8. Is Rodgers' book title correct? Please check.

Thanks for pointing this out, the correct reference was replaced in the manuscript: Rodgers C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice. World Scientific, Singapore, pp 256, 2000.

22. * P.941, reference list, line 12. There is a misprint for Zupanski et al. (2007): delete "bibitem 26"

Thanks for noticing. This misprint was corrected.

23. * Caption of Fig. 7. The agreement is not very good for Cycle 3 (middle panel). Please acknowledge.

NPGD

1, C384–C401, 2014

Interactive Comment

Following Reviewer's suggestion, the caption of Fig. 7 was replaced with:

"Fig. 7. Degrees of freedom for signal (top-three plots) of assimilated lightning data and observed GOES IR and WWLLN lightning flash rates (bottom-three plots, courtesy of Gregory DeMaria and Jack Dostalek) for cycles 1, 3, and 5. The areas of highest density of lightning observations are in general agreement with information content, implying that the flow-dependent ensemble forecast error covariance geographically coincides with throughout most of the assimilation period. Note, however, that the agreement for cycle 3 is not very good, implicitly confirming that ensemble forecast uncertainty is not always sufficient to represent the true forecast uncertainty."

24. * Caption of Fig. 11. It is not correct to comment this Figure saying that the error reduction obtained at analysis time is kept, even partially, in the 6-h forecast.

We agree with the reviewer. The caption of Fig.11 was replaced with the following: "Fig. 11. Root mean square (RMS) errors with respect to lightning flash rate observations during six assimilation cycles at 6-hour intervals: (a) Analysis RMS error. The RMS error reduction was achieved during the first 5 cycles of the assimilation period, while there is deterioration in the last cycle, possibly due to the fact that the system was exiting the model domain. (b) 6-hour forecast RMS error. There is no clear improvement in the forecast, suggesting that additional development of the assimilation system might be required, such as an improvement of the observation operator, adding new observations, and possibly improving the forecast uncertainty estimation."

Reviewer No. 2

It was generally a pleasant read for me on the assimilation of lightening data with the WRF-NMM at a 9-km resolution for a single severe-weather case. The paper is generally well written, and the topic is very interesting. Since this study is the first step to lightning data assimilation, some limitations of the paper, such as a single case study rather than comprehensive statistical comparisons, are understandable. The paper is within the scope of the journal and worth publication in general. However, it is not very

NPGD

1, C384–C401, 2014

straightforward to find the exact impact of assimilating lightning data. I would suggest to improve the overall presentation of results. Also, some important details need to be provided. I would list specific points as below. Once these points are addressed properly, it may be acceptable for publication.

We want to thank the reviewer for the valuable comments we received. The manuscript has been revised to address shortcomings as suggested.

Specific comments:

1. The impact of assimilating lightening data is not clearly presented in the results. For example, Fig. 9 shows analysis increments, but I found almost no increment where the dense lightening data exist. Also, Fig. 11 shows better analysis fit to lightening observations, simply meaning that the system is working properly, but not necessarily meaning the lightening data help improve NWP. I would suggest to provide analysis/forecast fit to other (unused) observations such as surface stations. Overall revisions on the presentation of results, in particular, clearer presentation of the impact due to additional lightening data (i.e., LIGHT vs. NODA) is strongly recommended.

Following the Reviewer's suggestions, improvements on the presentation of results were made throughout the manuscript, with particular emphasis on Sect. 4 through 5.

The following paragraph was added to Sect. 4.3: Almost no analysis increments can be found in the region where the densest lightning observations are located (Alabama). Among possible reasons, we can mention the following: (i) the largest forecast uncertainty (i.e. ensemble perturbations) typically occurs in the areas of strongest dynamical instability, in this case, in the region where a dry line was present over the states of Louisiana, Mississippi, Arkansas, and Missouri. Even though, the dry line may not be characterized by the strongest lightning activity, there were still some isolated lightning observations present over the domain as seen in Fig. 8c, (ii) alternatively, it may be a consequence of using an ensemble-based forecast error covariance that was not able to produce sufficient uncertainty in all relevant areas.

We made several efforts to fit the analysis and forecast with respect to several unused observations including those from surface stations. For example, time series of temperature, pressure, humidity, the U, V components of the wind, and accumulated precipitation were calculated at different locations within the inner nest in the experiments. However, we felt this approach was unpractical and inconclusive at this stage.

2. Some important details are missing.

(a) In section 2.3, it is unclear how to relate w to wmax. wmax = w? If so, it should be explicitly stated. I guess usually wmax > w, and we could define wmax = αw, with the multiplicative factor $\alpha > 1$, possibly being a function of w.

Following the reviewer's comment, we added some clarification to the definition of wmax in Sect. 2.3:

"The observation operator transformation (e.g., Eq. (6)) is defined over a 2-dimensional horizontal domain only since flash rate f is a horizontal field (e.g., number of hits per area and time). This requires wmax to be 2-dimensional as well. Therefore, wmax is defined for each horizontal grid point, as the maximum value of vertical velocity (w) over all vertical levels."

(b) In section 3.3, it is unclear what are the ensemble initial and boundary conditions, what is the localization setting for ensemble-based covariance, what are the hybrid settings such as weight between ensemble-based and static background error covariances. Also, it should be mentioned in this section that no other observations are assimilated.

Following the Reviewer's suggestions the following is added in Sect. 3.3" "The ensemble boundary conditions are obtained from the NCEP Global Forecast System (GFS) using the WRF preprocessing system (WPS). With the exception of the initial ensemble preparation (i.e. cycle 0 in our terminology), the initial conditions for the ensemble members are obtained through the MLEF algorithm by adding the analysis square root

NPGD
1, C384–C401, 2014

Interactive Comment

error covariance columns to the analysis. Further information about the MLEF methodology can be found in Zupanski (2005) and Zupanski et al. (2008). The localization setting for the ensemble-based covariance includes a de-correlation length of 90 km.".

Note that MLEF is considered a hybrid approach since it is using a nonlinear minimization from variational methods in the ensemble context, and it does not include the static error covariance yet (ongoing work). This has also been acknowledged by researchers developing other approaches to hybrid data assimilation (e.g., Wang, 2010). We mention in section 3.3 that no other types of observations were used in these experiments. Thanks for pointing this out.

Wang, X., 2010: Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation Variational Minimization: A Mathematical Framework. Mon. Wea. Rev., 138, 2990-2995.

3. P.934, L.22, I am not convinced if this is really "new information." Sampling errors due to a limited ensemble size are treated as if they were "information." It is unclear if what is shown here is "information" rather than spurious sampling noise.

Thank you for pointing out this issue. Note that our calculation of Degrees of Freedom for Signal (DFS), Section 4.2., Fig.7) implicitly confirms that there is a new information added to the system by assimilating lightning observations, since by definition (e.g., Rogers 2000, Zupanski et al. 2007) DFS measures the information content (e.g., entropy change), not the noise. If the impact of observations were to be negligible, DFS would be close to zero. Additional clarification regarding this issue is added to Section 4.2.

4. Section 4.2 describes d.o.f. for signal. I understood intuitively that the d.o.f. for signal mean the number of independent pieces of information from a certain number of observations considered here. If my understanding is correct, it is important to provide how many observations are considered for each d.o.f. value.

Separating the number of observations per DFS is not possible because the observation information matrix transformation (Eq. 9 from Zupanski et al. 2007) uses the total number of observations in a local domain. This is formally a consequence of the fact that analysis error covariance combines all uncertainties from the forecast and from the observations. As a result, there is no possibility of separating their impacts per DFS.

5. It may be interesting to discuss about the potential non-Gaussianity of the lightening data. Lien et al. (2013, Tellus A) suggested an approach to deal with observations that have non-Gaussian error PDF, and taking such an approach may improve the use of lightening data.

Following the Reviewer's suggestion we added a brief discussion on the non-Gaussian character of lightning observations.

"The non-negative character of lightning observations introduces a skewness that points out to a need for a non-Gaussian PDF in data assimilation (e.g., Fletcher and Zupanski, 2007; Lien et al. 2013). This issue will be examined in the future since it can potentially improve the utility of lightning data."

Fletcher, S.J. and M. Zupanski, 2006: A Data Assimilation Method for Lognormally Distributed Observational Errors. Q.J. Roy. Meteorol. Soc., 132, 2505-2519.

Lien, G.-Y., Kalnay E., and Miyoshi T.: Effective assimilation of global precipitation: Simulation experiments. Tellus A, 65, 19915. doi:10.3402/tellusa.v65i0.19915, 2013.

Minor points:

1. P.919, L.25, "void" sounds too strong. There are observations such as AMVs and aircraft data. We omitted the use of "void".

2. P.920, L.2, "emissions" -> "production"?

Thank you, we replaced the word "emissions" by "production".

3. P.924, L.17, what is the unit of c? counts/hour/m/m or such?

NPGD

1, C384–C401, 2014

Interactive Comment

The parameter c is dimensionless. This was also added in the text.

4. P.924, L.24, a dot is missing on top of .

Thank you; this was corrected in the manuscript.

5. P.924, L.25, I thought WRF-NMM is a nonhydrostatic model that has prognostic vertical velocity.

Although WRF-NMM is a non-hydrostatic model, vertical velocity is a diagnostic field, unlike WRF-ARW, where it is a prognostic variable.

6. P.925, L.3, what is the unit of CWM?

The CWM units are kilograms per kilogram of dry air. These units were added in the manuscript (P.925, L.3)

7. P.925, L.18, observed lightning rates were considerably larger than the guess possibly because wmax = w is assumed?

As we explained earlier, the maximum vertical velocity is calculated for each horizontal point and represents the maximum value of vertical velocity over all vertical levels. Formally, wmax is a 2-d variable, while w is a 3-d variable. We believe that this was the consequence of using the observation transformation formula (Eq. (2)), with coefficients estimated from a large-scale climatological study, not really representative of short-term assimilation intervals used in this work.

8. P.927, L.8, “the diagonal elements of the eigenvalues matrix” -> “the eigenvalues”

Changed as suggested, thank you.

9. P.929, L.16, “was” -> “were”

Thank you for spotting this, “was” was replaced by “were.”

10. P.934, L.9, “no other types of observations being assimilated” may be bad for LIGHT if we think about relative impact. With many other observations, lightening

NPGD

1, C384–C401, 2014

Interactive Comment

observations may not have as distinct impact.

Thank you for the comment. Although this may be true, it is a consequence of using data assimilation. It may be possible, however, that lightning data are the only observations in a remote region such as open oceans. The changed text includes: “Lightning is just an additional type of observations. All available observations have to be in agreement with each other at the same location. Therefore, in regions where lightning observations are not in agreement with other types of observations, the data assimilation algorithm will create the optimal observation impact based on uncertainty of all observation in the region. In regions where lightning observations are not available other measurements should help.”

Please also note the supplement to this comment:
http://www.nonlin-processes-geophys-discuss.net/1/C384/2014/npgd-1-C384-2014-supplement.pdf

---

Interactive comment on Nonlin. Processes Geophys. Discuss., 1, 917, 2014.

NPGD

1, C384–C401, 2014

Interactive Comment

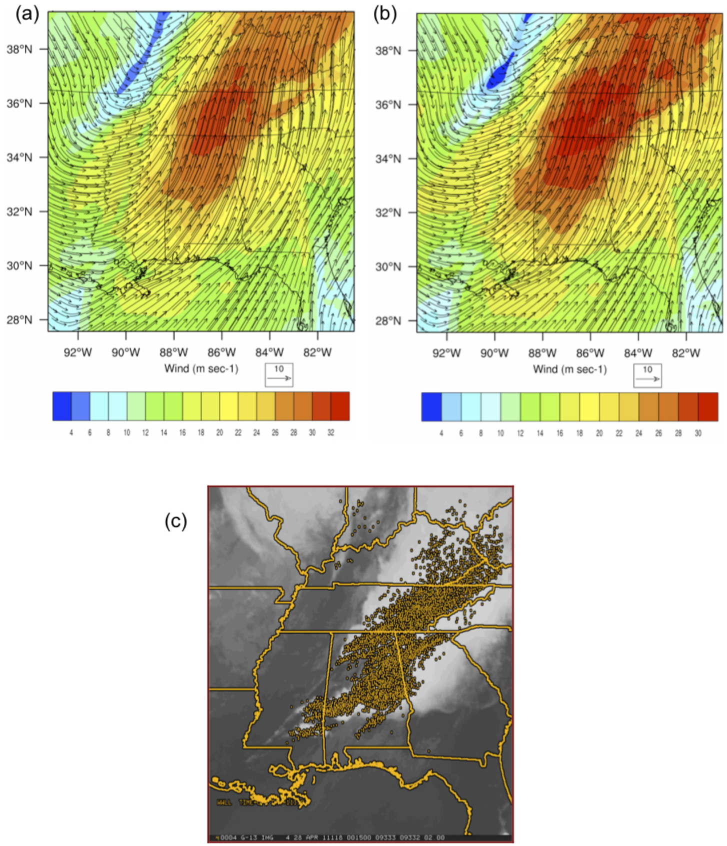

Fig. 1. Fig. 8

NPGD

1, C384–C401, 2014

Interactive Comment

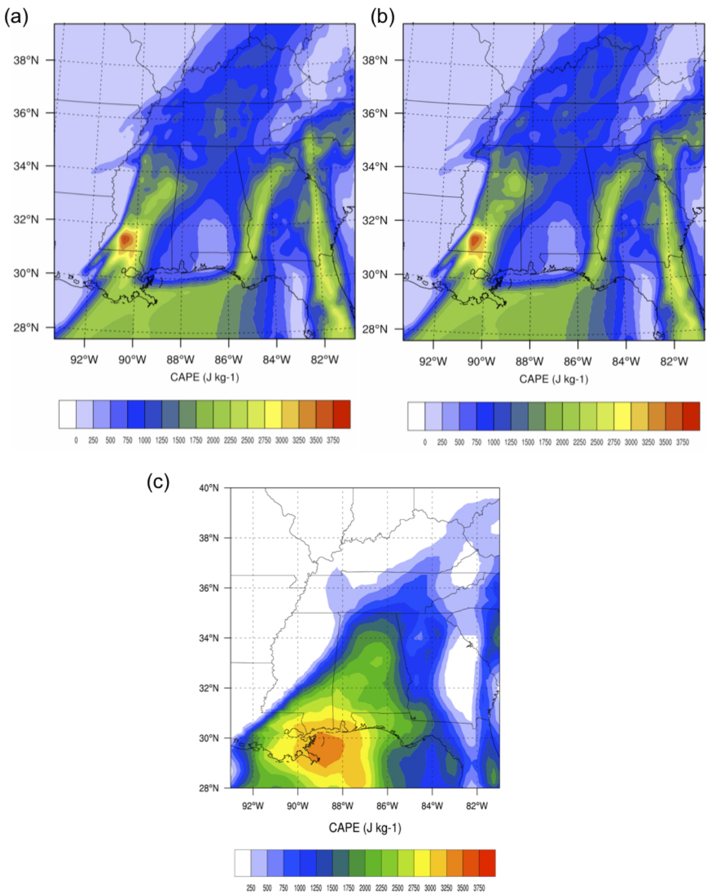

Fig. 2. Fig. 10

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper