Responses to the reviewers

Large impact of tiny model domain shifts for the Pentecost 2014 MCS over Germany

by Christian Barthlott and Andrew I. Barrett November 26, 2019

Dear Editor,

This letter accompanies our revised manuscript. We are grateful for the reviewer's helpful comments, and hope our revision addresses them all. Below we detail the changes made in our revision. We include the text of the reviews in black, our responses are in blue.

Reviewer 1

This study examines simulations of a rather impactful MCS over Germany in 2014. The COSMO model at convection-allowing resolutions is used to simulate the MCS in a control run, and additional simulations move the model domain one grid point in eight cardinal directions. The predictability and forecast errors of the MCS are largely dependent on convection initiation in western France and subsequent propagation of the system over land, as well as the environment the system encounters during the day before impacting Germany. Substantial variability exists amongst the simulations, whereby some instances of the MCS-induced rainfall is forecast reasonably over Germany and other instances there is no precipitation in this region at all. It doesn't appear that systematic movement of the domain resulted in clustering of forecast errors either.

Major Comments

1. My first and foremost major concern is the contribution of this work to the atmospheric science literature on MCSs, predictability, and forecast generation. I believe what the authors have described is merely a technique to perturb lateral boundary conditions, thereby producing spread in the initial states of the model simulations which filters into the forecasts of the MCS. If we assume the environment at the model boundaries is relatively homogeneous, at least within a few grid points which only amounts to < 10km, then what the authors have described here is essentially a 10-member ensemble forecast system with perturbed boundary conditions. This methodology is not in itself flawed by any means, but I do not believe it is innovative or new. See Torn et al. (2006), Gebhardt et al. (2011), and Romine et al. (2014) for other examples of boundary condition perturbation studies. Given this statement, however, I think the authors could easily address my concern by a number of avenues:

    a. If the authors believe this truly is an innovative technique to generate CAM ensembles, they should either more succinctly clarify this in the introduction with references supporting this claim or demonstrate the methodology alongside some of the more traditional techniques (e.g., covariance perturbations) for this case study.

    We do not believe that our technique is a new method to generate ensembles with perturbed initial/boundary conditions in operational convective-scale ensemble forecasting. However, we were surprised to see such a large influence of these tiny changes on the simulation results and strongly believe that this method should be tested for more cases (also with different extents of domain shifting) and other models. It may also be that the high sensitivity is a feature of days with low predictability only, which would be a useful information to have. Therefore, a more systematic evaluation is left for future work. We adapted the text to make that clearer.

Changes to paper

Abstract:

This study demonstrates the potentially huge impact of tiny model domain shifts on forecasting convective processes in this case, which suggests that ~~the inclusion of this simple method in convective-scale ensemble forecasting systems~~ *the sensitivity to similarly small initial condition perturbations* should be evaluated ~~for different cases, models~~ *across other cases, model* and weather regimes.

Summary:

The results of this work suggests that ~~the method of~~ model domain shifting could be used to ~~account for~~ *quantify how* uncertainties in the initial and boundary conditions ~~by introducing a small disturbance at model initialization~~ *contribute to the predictability of an event*. However, this single case study needs to be expanded to cover more cases ~~. Thus, it is of interest to further evaluate this simple approach of domain shifting,~~ ,for example in weather regimes with strong synoptic forcing and more stratiform precipitation and in other models such as ICON...

b. If the authors would still like to use the domain-shifting methodology to investigate the predictability of the MCS, I would caution attributing the methodology to why the MCS is inherently unpredictable. In order to scientifically attribute the poor forecast predictability to the domain-shifting methodology, substantially more analysis and simulations would need to be conducted. For instance, do you see the same poor predictability if the domain is moved 5, 10, or 20 grid points? What about if another perturbation technique is used? Can you reproduce the poor forecasts?

The goal of this paper was not to assess the impact of other perturbation techniques, as we already mentioned the poor forecast quality of the operational COSMO-DE-EPS of the German Weather Service in the introduction. While trying to find a model setup which could reproduce the MCS, we made a lot of tests, also with respect to domain size and domain location. When we changed the location of the domain by 2, 10, and 20 grid points, we already had successful and unsuccessful results. This is why we went to the minimal domain shifting possible, namely 1 grid point in eight cardinal directions. We believe that this is a good first step and this method should be evaluated as mentioned in the reply to the first comment. It is of special interest to see, if other cases with low predictability (i.e. forecast busts) show the same sensitivity. However, we think that such an analysis would not fit into the present paper and is therefore left for future work.

Changes to paper:

none

c. If the authors would rather focus on the predictability aspect of this event, I believe the authors could implement some other analysis techniques to derive some of the dynamic aspects for this case to complement what has been presented. Sensitivity approaches such as those demonstrated by Schumacher and Davis (2010) and Ancell and Hakim (2007) could be valuable additions to the analysis. I invite the authors to consult a number of papers that apply sensitivity analyses to convection-resolving forecasts as well: Bednarczyk and Ancell (2015), Hill et al. (2016), Limpert and Houston (2018), and Torn et al. (2016). Additionally, other aspects of predictability could be garnered through initializing ensemble forecasts at later times, which may answer the particular question of whether CI is the limiting factor of predictability.

We thank the reviewer for this useful hint and performed an ensemble sensitivity analysis for our model runs. We present the results in the new section 4.5.

Changes to paper:

We include the sensitivity analysis in a new section 4.5, this includes explanation of the method and interpretation of the results and includes discussion of the newly added Figure 9.

2. A second concern I have is in the presentation of the forecast itself. There is no mention of the upper-level dynamics that could be supporting MCS development, particularly since the orientation of development and distribution of environmental parameters conducive to MCS propagation are misaligned from traditional understanding. For instance, the MCS propagation within a region of predominantly northwesterly or westerly surface winds, which would not advect the CAPE-rich air from the southeast. Typically, we would expect a convergence of moisture and higher theta-e air just ahead of the MCS, but this is not the case. Also, there is no mention as to what causes the MCS to initiate so early in the day. My inclination from reading the forecast description is that the orientation of the longwave mid-tropospheric trough is supporting the traversal of short-wave troughs through western Europe. I suggest the authors add supporting evidence for how the MCS initiates, which could elucidate some other predictability elements that have not been considered, e.g. the position and placement of upper-level vorticity maxima.

   The MCS propagation has not been the subject of the paper so far, as we focused more on the fact why the precursors of the MCS dissipate or not. This is why Figure 9 only presents the period between 1000 and 1400 UTC. The system later evolves into an MCS as can be seen in Fig. R.1, which shows that the distribution of environmental parameters are not misaligned from traditional understanding. We observe exactly what the reviewer has anticipated, but was

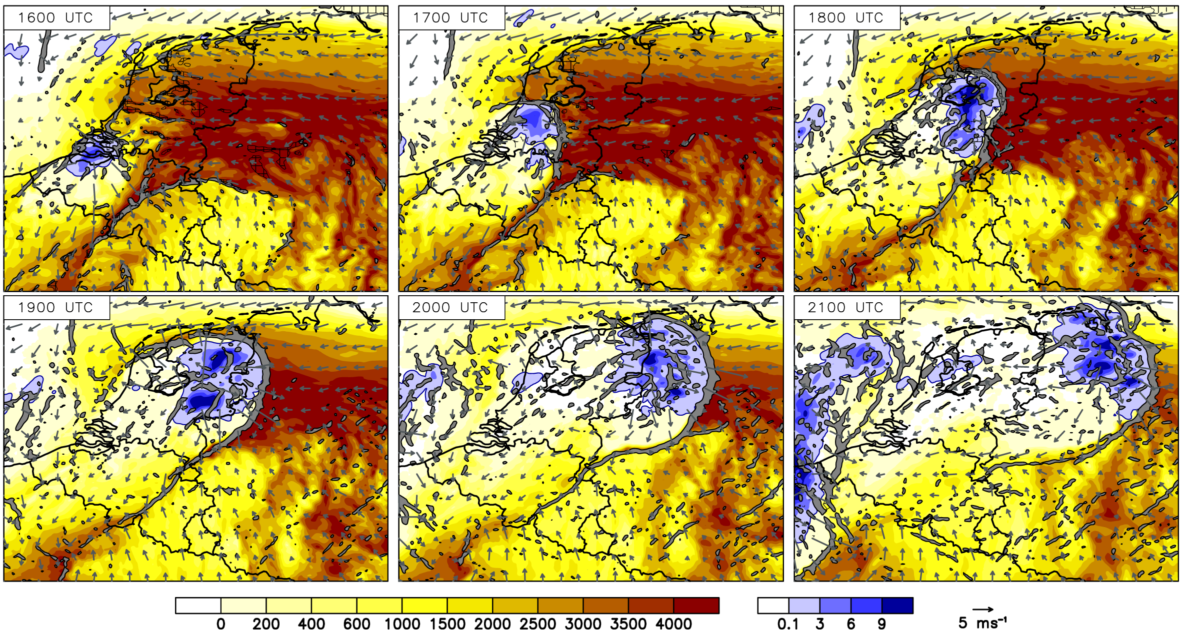

Figure R.1: Convective available potential energy (colour shading, in J kg^{-1}, 30-min precipitation rate (blue colour shading, in mm (30 min)$^{-1}$, and 10-m wind field (arrows) between 1600–2100 UTC on 9 June. Gray areas indicate low-level wind convergence larger than $0.35{\cdot}10^{-3}$ m s^{-1} and hatched areas represent regions where convective inhibition is smaller than 5 J kg^{-1}.

   not able to see in the Figure 9: an advection of CAPE-rich air from the East, with well-defined region of low-level wind convergence at the outflow boundary. The MCS cleary moves into the region with high CAPE which corresponds to high values of equivalent potential temperature.

We included this Figure in a new subsection 4.7 in the manuscript.

Changes to paper:
new Figure 11 and new subsection 4.7:

"Having established a possible explanation for the decay of the precursors of the MCS in the previous section, we now analyze the further evolution of the system into a MCS using the reference simulation (Fig. 11). To the east of the system, the model simulates an east-west oriented region of high low-level equivalent potential temperature in the north-central part of Germany, which corresponds to CAPE values between 3000–4000 $J\,kg^{-1}$. This CAPE-rich air is advected with easterly winds towards the convective system over the Netherlands. Colliding with the cell's outflow, a strong low-level mass and moisture convergence occurs, which fosters the evolution into a MCS. As already discussed in section 4.4, the 0-6 km deep layer shear shows suitable conditions for highly-organised convection (27–30 $m\,s^{-1}$). The maximum rain intensities reach locally up to 22 mm $(30\,min)^{-1}$ with a weakly defined bow-like structure of precipitation, typical of storms with an intense rear-inflow jet. In the wake of the MCS, CAPE is almost entirely consumed. From 23:00 UTC onwards, the MCS is decaying while further travelling towards Poland (not shown)."

3. Why is accumulated rainfall used as the sole metric of forecast evaluation? I would think observed radar reflectivity compared to simulated reflectivity would be a better metric for comparing model runs. Comparing reflectivity would better illuminate the intensity and structure of the MCS between observations and simulations; accumulated rainfall doesn't discriminate these differences well.

   We believe that rainfall at the ground is a suitable metric to assess the sensitivity of the model in simulating an MCS. Even if the model shows some discrepancies with respect to location and propagation speed, an overall good agreement between simulation and observation exists. Moreover, the focus of our paper lies on the sensitivity of the model to domain shifting and in-depth comparison of the MCS of the reference run with radar observations is not necessary for the reader to follow our story. Also, radar reflectivities are not available to the authors and the simulations would have to be done again with a radar-forward operator. Having said that, we think that an evaluation with rainfall is sufficient for our purpose.

Minor Comments:

1. Why was the predictability low for the operational prediction systems? Do those systems parameterize convection or is it explicitly solved?

   The origin of the low predictability of this case was unknown so far. This paper is another contribution to that topic. While in the Barthlott et al. (2017) paper, an enlargement of the model domain, a higher grid spacing and a single/double moment microphysics scheme were addressed, this study has shown show that small displacements of the convective system over France can lead to a decaying system or to a system developing into an MCS later on. The operational prediction systems used the same grid spacing as in our study, so deep convection was resolved and shallow convection parameterized. We added remarks in the model description and in the summary.

   Changes to paper
   Model description:

   Deep convection is resolved explicitly and a modified Tiedtke-scheme (Tiedtke, 1989) is used to parameterize shallow convection *(as did the operational deterministic and ensemble prediction system at that time)*.

Summary:

However, the predictability of this event was very low; neither the operational deterministic nor the ensemble prediction system *(both convection resolving)* captured the event with more than 12 hours lead time.

2. "However, the low predictability of the event was evident by the surprisingly large impact of tiny changes to the model domain": Is the argument there is low predictability because of dynamics or because of the model configuration? There appears to be two separate statements of predictability related to this event, but it is unclear what statements the authors really want to make. I'm assuming the main predictability element comes through the numerical (domain) aspect.

   As outlined in the reply to the previous comment, we do not know the origin of the low predictability. In our study, the low predictability is reflected by the domain effect.

   Changes to paper:
   none

3. Introduction Lines 28-30: Were the German Weather Service operational models convection resolving? Is there any indication as to "why" the deterministic and ensemble systems failed to produce convection over Germany? This piece of discussion would be a good addition to the manuscript to help explain "why" the model forecasts failed and potentially motivate the use of convection-allowing models.

   The operational models were convection-resolving, i.e. with a horizontal grid spacing of 2.8 km. As already mentioned earlier, we do not know why these model runs failed to produce convection over Germany. We added two remarks in the manuscript about this fact and the operational resolution.

   Changes to paper
   Model description:
   Deep convection is resolved explicitly and a modified Tiedtke-scheme (Tiedtke, 1989) is used to parameterize shallow convection *(as did the operational deterministic and ensemble prediction system at that time)*.

   Summary:
   However, the predictability of this event was very low; neither the operational deterministic nor the ensemble prediction system *(both convection resolving)* captured the event with more than 12 hours lead time.

4. Line 35: First bullet point: what operational model is being discussed here? Second bullet point: Is the COSMO model being discussed here? Please be explicit about what model and associated configuration is being altered.

   We did not mean the operational model, but the COSMO model in an operational setup. We modified the text to make that clearer.

   Changes to paper:
   A series of different numerical simulations for the convective events of 8 and 9 June 2014 were performed *"with the COSMO model"*, the main findings were:

   - The ~~operational model~~ *COSMO model (in quasi operational set-up, without data assimilation)* initialized at 00:00 UTC reproduced the events on 8 June only, but not the mesoscale convective system (MCS) on 9 June.

5. Line 49: What is COSMO-DE? While the COSMO acronym has been properly described, I don't know what "DE" references.

   COSMO-DE is the name of operational configuration at DWD over Germany. This information is not needed here. So instead of explaining it, we just replaced *"COSMO-DE domain"* with *"model domain"*.

6. Lines 85-91: What benefit does "domain shifting" have over other traditional lateral boundary perturbation techniques (e.g., Torn et al. 2016)? I have not been convinced in the introduction that there is significant benefit in developing a new technique to perturb boundary conditions. Would it be appropriate to compare the described "domain shifting" technique with other perturbation techniques? Including this type of analysis would presumably shift the focus of your manuscript to an evaluation of ensemble-generation techniques for a specific MCS case study. Alternatively, the authors could instead focus on the true predictability of the event (rather than the domain shifting idea) and include some additional predictability analysis (e.g., ensemble sensitivity). See Schumacher and Davis (2010), Ancell and Hakim (2007), Bednarczyk and Ancell (2015), Hill et al. (2016), and Torn et al. (2016) for some examples of sensitivity analysis for precipitation and high-impact weather forecasts. (Major comment above)

   It was not our goal to evaluate different ensemble-generation techniques. Our case study is a first step, but needs evaluation with more cases before it can be compared to different methods to introduce uncertainties in the initial and boundary conditions. However, the study of Henneberg et al. (2018) showed, that by shifting the model domain, by ten to 30 grid points, an estimate of the uncertainty of the model results can be achieved with a sufficient large model spread. We believe that the large impact of these tiny changes need further evaluation with more cases, different extents of domain shifting, and other models. In several places in the manuscript, we now state that this method needs further evaluation and that suitability for representing uncertainties should be compared to traditional lateral boundary perturbation techniques.

   We are grateful for the examples of the sensitivity analysis. We performed such an analysis, the results are presented in the new section 4.5.

   Changes to paper:
   We include the sensitivity analysis in a new section 4.5, this includes explanation of the method and interpretation of the results and includes discussion of the newly added Figure 9.

7. I would suggest leaving the descriptive nouns out of the manuscript, and let the reader decide what is "surprising" or not (e.g., Line 92).

   We do believe that our technique of domain shifting of just 1 grid point provides surprising or at least unexpected results for this particular case. Given the large model domain and the minor changes at the boundaries, we would not have anticipated such a large dependency. Therefore we like to keep our phrasing in the current form.

   Changes to paper:
   none

Section 4

1. What is the source of radar observations? Would be appropriate to add this into the manuscript for reproducibility.

   The radar observation come from the radar network of the German Weather Service, the product is called RADOLAN. We added this sentence at the beginning of section 4.1.

Changes to paper
"Here we compare our simulations to radar-derived precipitation from the precipitation analysis algorithm RADOLAN (Radar Online Adjustment), which combines weather radar data with hourly surface precipitation observations of about 1300 automated rain gauges to get quality-controlled, high-resolution (1 km) quantitative precipitation estimations."

2. Line 144: should be (Fig. 4a)

   Done

3. Lines 158-159: I actually do not agree with this statement. I think the reference forecast has some glaring errors that do not make this a particular good forecast. Consider revising or removing this statement.

   We agree with the reviewer that the bow-like structure is not well-defined in our simulations. We therefore removed that sentence. Otherwise, we think that the model is doing a reasonable job, despite the differences already described in the text.

   Changes to paper:
   ~~However, the model succeeds in producing the bow-like structure of precipitation, typical of storms with an intense rear-inflow jet.~~

4. Lines 179-181: All the eastward shift simulations have poorer prediction though.

   We agree with the reviewer as the precipitation in the E-run is more to the North and does not extend as much to the East as the other successful runs.

   Changes to paper:
   "However, as the precipitation in the E run is more to the North and does not extend as much to the East as the other successful runs, all the eastward shift simulations have poorer prediction."

5. Lines 270-272: I think a reasonable counter argument could be that the W run initiated the convection well to the east (east of the red circle) and therefore had an earlier impact over Germany than the reference run, making it a "poor" forecast. Additionally, this forward storm system appeared to greatly impact the CAPE field in Figure 9, which seemingly had an impact on the development of upstream convection in the red circle. Furthermore, there is clearly a neutral to slightly negatively-tilted mid-tropospheric trough to aid in the propagation of short-waves (hard to tell where these might exist in the coarse resolution of Figure 1): what role did mid-tropospheric dynamics play in this system? I suggest a more thorough evaluation of the simulations and discussing all aspects of the environment more thoroughly, including any convection that might have influenced convection initiation (CI) in the focus area.

   The reviewer is right about the fact that in the W run, a convective cell occurs east of the red circle. We already mentioned this at the end of section 4.5:

   "The isolated cell, to the north west of these plots between 1000-1100 UTC, does not appear to be important to the decay of the cell of interest. It is located approximately 150 km upstream. The cell is stronger in the W run leading to a slight reduction of CAPE and therefore creating slightly less favorable environmental conditions in the area into which the main cell would later move. However, it appears that the weakening of the main cell occurred independently of the cell upstream and can rather be attributed to the proximity to the colder sea surface."

   Our main counter argument would be that this convection does indeed reduce the CAPE locally, but the reduction in CAPE does not reach the region where the cell of interest is decaying. At 1130 UTC, the cell in the W run is decaying although further downstream there is still a tongue of air with higher CAPE values as in the REF run. Moreover, in the SE run there is no cell to the east of the system of interest, and convection dies out anyway, in spite of the unaltered CAPE field downstream. We therefore conclude that the weakening of the main cell occurred independently of the cell upstream and can rather be attributed to the proximity to the colder sea surface.

Changes to paper:
none

6. Line 284: The sea surface temperatures have not been described in detail yet. How do we know these SSTs are the limiting factor? We do not know what the SSTs from each simulation are or how they dynamically are impacting the simulation convection. Seems like a reaching statement without any evidence and I would suggest revising or providing more concrete, quantitative support.

   The surface temperature and CAPE are depicted in Fig. R.2. The sea surface temperature is much lower than the land surface temperature, at least in the northwestern coast of France where no significant amounts of rain was simulated in the last hours. As a result of these lower temperatures, CAPE is significantly reduced over sea. Along the coastline, there is a strong gradient in temperature (23 → 15 deg C) and CAPE. These statements also hold true for the remaining model runs. We added some remarks on that in the text, but decided not to provide an extra figure.

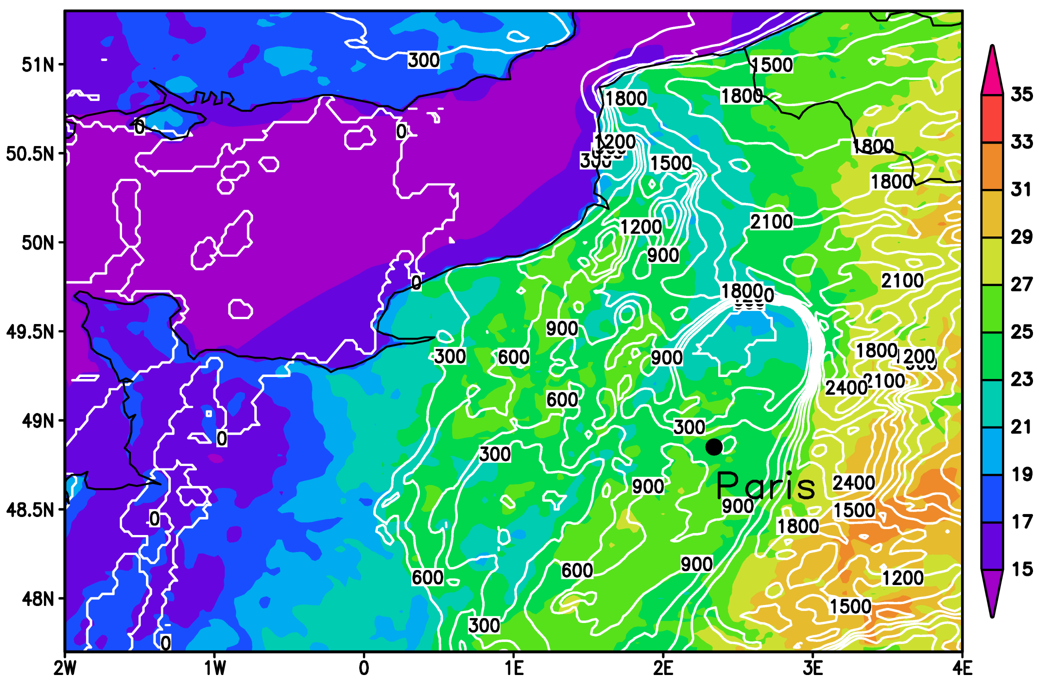

Figure R.2: Surface temperature of the REF run at 1000 UTC (colours, in deg C) and CAPE (white contours, in $\mathrm{J\,kg^{-1}}$).

Changes to paper:
"The sea surface temperatures along the French coast lie around 15°C and are much lower than the land surface temperatures (around 23°C, not shown). This temperature distribution is similar in all model runs for the preconvective environment."

References:

Ancell B. C. and G. J. Hakim, 2007: Comparing Adjoint- and Ensemble-Sensitivity Analysis with

Applications to Observation Targeting. Monthly Weather Review, 135, 4117-4134

Bednarczyk C. N. and B. C. Ancell, 2015: Ensemble sensitivity analysis applied to a southern Plains convective event. Monthly Weather Review, 143, 230-249

Gebhardt C., S. E. Theis, M. Paulat, and Z. B. Bouallgue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmospheric Research, 100, 168-177.

Hill A. J., C. C. Weiss, and B. C. Ancell, 2016: Ensemble senstivity analysis for mesoscale forecasts of dryline convection initiation. Monthly Weather Review, 144, 4161-4182.

Limpert G. L. and A. L. Houston, 2018: Ensemble senstivity analysis for targeted observations of supercell thunderstorms. Monthly Weather Review, 146, 1705-1721.

Romine G. S, C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. Monthly Weather Review, 142, 4519-4541.

Schumacher R. S. and C. A. Davis, 2010: Ensemble-Based Forecast Uncertainty Analysis of Diverse Heavy Rainfall Events. Weather and Forecasting, 25, 1103-1122.

Torn. R., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. Monthly Weather Review, 134, 2490-2502.

Torn R., G. S. Romine, and T. J. Galarneau, Jr., 2016: Sensitivity of dryline convection forecasts to upstream forecast errors for two weakly forced MPEX cases. Monthly Weather Review, 145, 1831-1852.

Additional changes to the paper:

1. We included a new sentence in the introduction about two recent papers:

   "Recent studies of Schneider et al. (2019) and Keil et al. (2019) have also shown that different assumptions for the amount of cloud condensation nuclei could be included in convective-scale ensemble forecasting, but only if the model employs a double-moment microphysics scheme."

2. Old Figure 9 was enhanced by increasing the size, length, and density of the wind arrows.

3. Information about the financial support was added.