Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

We thank the reviewer for their ongoing evaluation of this work. Our response to the reviewer's comments is indicated in green for each comment.

Note that in accordance with the reviewer's comments, we have undertaken two additional simulations with the Morrison scheme (no hail and with hail) to examine the sensitivity of this event to reviewer suggestions for: vertical levels (increased from 41 to 65 levels); and damping. All other settings remain as before. We detail the analyses of the new simulations that are presented in the paper and below here.

We appreciate that our research design is not exactly what the reviewer would have done but note that some of the disagreements really lie in personal preference regarding scheme selection and model configuration. All schemes employed here are accepted and well used schemes in WRF for many version iterations and have been subjected to extensive testing and development. We have generated a uniquely comprehensive ensemble suite and rigorously evaluated it and sought to link performance to dynamical drivers. We believe this work is valid and useful and hope that given the extraordinary lengths we have taken to address the reviewer's concerns (many additional simulations), this matter can be brought to a positive and successful conclusion.

Reviewer 1 comments:

The namelist that the authors provided shows some problems to me. It validates my concerns about how the simulations were conducted.

1. No shallow cumulus parameterization is used, which is surely a big problem for domain 1 which is 12 km resolution. For the 4-km, I would still recommend using it for convective system simulations.

We have used the shallow convection parameterization in the 12 km simulations. The Kain-Fritsch scheme used in the 12 km simulations (like all convective schemes) runs shallow convection by default. The user must change the Kain-Fritsch parameterization code manually and recompile WRF to switch shallow convection off (see lines 1393-94 of module_cu_kfeta.F).

NB: Perhaps the reviewer has misread the namelist provided and is reading from the line above which indicates the settings for boundary layer physics calls (0). The namelist states quite clearly that convective parameterization is used on the 12 km domain:

```
bldt                     = 0,     0,     0,
cu_physics               = 1,     0,     0,
cudt                     = 5,
```

While opinions vary, multiple previous studies have used no cumulus scheme at 4 km or below (i.e. "convection permitting" simulations) (see an example citation list below). In addition, Romine et al. (2013) states that: "*Several efforts in convection-permitting NWP (CP; horizontal grid spacing of 4 km or less), both operational and experimental, have commenced in the last decade. This is motivated in part by evidence that resolution sufficient to allow explicit representation of convection improves simulation of both the convection's diurnal characteristics (Done et al. 2004; Liu and Moncrieff 2007) and its mode and structure (Clark et al. 2007; Kain et al. 2008; Schwartz et al. 2009; Sobash et al. 2011).*"

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

Example citations of studies that have used only a microphysics scheme within the innermost domain:

- Khai Shen Sow, Liew Juneng, Fredolin T. Tangang, Abdul Ghapor Hussin, Mastura Mahmud, Numerical simulation of a severe late afternoon thunderstorm over Peninsular Malaysia, *Atmospheric Research*, 99, 2, 2011, https://doi.org/10.1016/j.atmosres.2010.10.014: model domains of 27, 9, 3km. "*In the third domain, several cumulus schemes were examined but the simulation result is the best with the cumulus process explicitly resolved*".

- Rajeevan, M., Kesarkar, A., Thampi, S. B., Rao, T. N., Radhakrishna, B., and Rajasekhar, M.: Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over Southeast India, Ann. Geophys., 28, 603–619, https://doi.org/10.5194/angeo-28-603-2010 2010: BMJ at 12, 8km, no cumulus parameterization at 2km.

- Garcia-Carreras, L., A. J. Challinor, B. J. Parkes, C. E. Birch, K. J. Nicklin, and D. J. Parker, 2015: The Impact of Parameterized Convection on the Simulation of Crop Processes. *J. Appl. Meteor. Climatol.*, 54, 1283–1296, https://doi.org/10.1175/JAMC-D-14-0226.1: "*The parameterized runs produced an unrealistic distribution of rainfall frequencies and intensities, a well-known issue with parameterizations of convection. The parameterization leads to too frequent light rainfall events (in time and space), with too few heavy rainfall showers and days with no rainfall at all. The two convection-permitting runs, on the other hand, had rainfall distributions much closer to observations, although peak rainfall intensities and rainfall totals were overestimated.*"

- Halder, M., Mukhopadhyay, P. Microphysical processes and hydrometeor distributions associated with thunderstorms over India: WRF (cloud-resolving) simulations and validations using TRMM. Nat Hazards 83, 1125–1155 (2016). https://doi.org/10.1007/s11069-016-2365-2: *Nested domains with increasing resolution of 27-, 9-, 3- and 1-km grid spacing ... Kain-Fritsch used in the 27 and 9 km domain only.*

- Xing Yu & Tae-Young Lee (2010) Role of convective parameterization in simulations of a convection band at grey-zone resolutions, Tellus A: Dynamic Meteorology and Oceanography, 62:5, 617-632, DOI: 10.1111/j.1600-0870.2010.00470.x: "*The impacts of resolution and convective parameterization at grey-zone resolutions (i.e. 3, 6 and 9 km) are then investigated. Results indicate that a grid size of 3 km is sufficient to resolve the convection band and CP for this size grid is not necessary. With 6 and 9 km grids, explicit simulations or those based on a Kain–Fritsch CP scheme do not simulate the atmospheric structure surrounding the band accurately*".

- Bodine, D. J., and K. L. Rasmussen, 2017: Evolution of Mesoscale Convective System Organizational Structure and Convective Line Propagation. Mon. Wea. Rev., 145, 3419–3440, https://doi.org/10.1175/MWR-D-16-0406.1. *27, 9, 3km domains with 44 vertical levels, and Kain-Fritsch used in the outer domains only (i.e. 27, 9 km).*

Weisman et al. (2013) NB: already cited in our study, used only a microphysics scheme in a 3km domain WRF simulation of an unusually intense bow echo and associated mesoscale vortex that were responsible for producing an extensive swath of high winds across Kansas, southern Missouri, and southern Illinois on 8 May 2009 (superderecho).

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

2. The radiation timesteps (Radt) are not set appropriately. It is recommended to be 1 min per km of dx based on the guide so the values for Radt should be 12, 4, and 1.3 for Domain 1, 2, and 3. The authors use 10 min for all three domains, which is particularly not appropriate for Domain 3 where the results of the study are mainly obtained from.

We have not made changes in response to this suggestion. Our reasoning is that the radiation scheme time step is set on domain 1 only and cannot be set to a different value on nested domains. From the WRF user guide on page 5-78: “minutes between radiation physics calls. Recommended 1 minute per km of dx (e.g. 10 for 10 km grid); use the same value for all nests”.
See:
https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/user_guide_V3.8/ARWUsersGuideV3.8.pdf

3. The vertical resolution of 41 is too coarse for convective cloud system simulations.

We have implemented this suggestion and have included two new simulations using 65 vertical levels.

We appreciate the reviewer’s concern. We note that using 41 levels is more than what is used by default in WRF operational/research application namelists provided by NCAR (e.g. 1 – 3 km convection permitting or hurricane applications). There is precedent in the literature for using ~40 vertical levels for WRF simulation of mesoscale convective systems (e.g. Bodine and Rasmussen (2017) used 44 vertical levels).

We have, however, as part of the two additional simulations, increased the number of vertical levels to 65. Figure 1 below shows a comparison of model vertical resolution and height above the surface between the 11 original ensemble members and the two new 65 level simulations. The change in resolution with height is more gradual and there is increased resolution above 0.5 km.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions* (nhess-2021-373)

Response to reviewer #1:

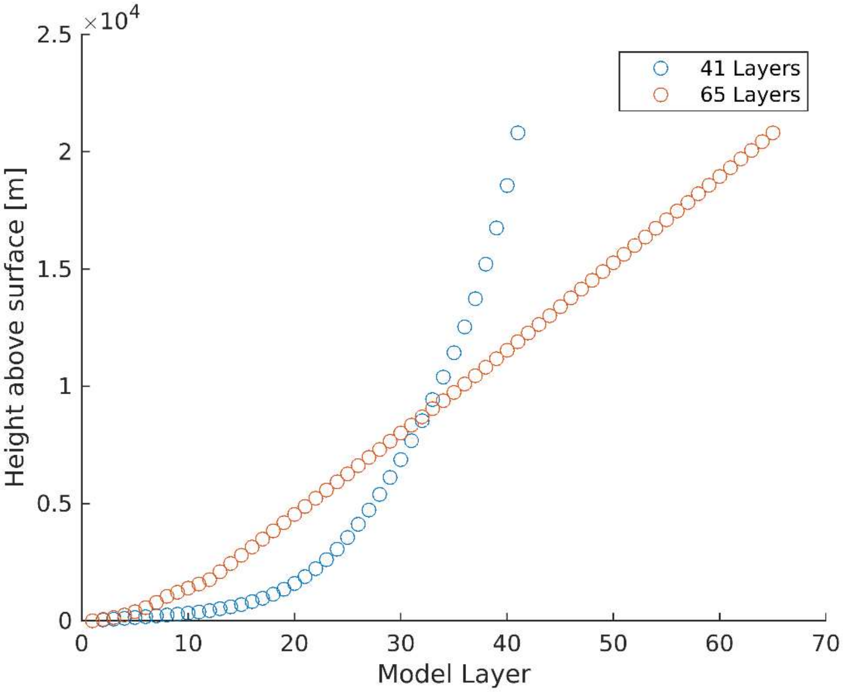

Figure 1: Comparison of model vertical resolution between the 11 original 41 vertical level ensemble members, and the two new 65 level Morrison simulations.

We now have an ensemble of 15 simulations. We reiterate that this study's objective is to examine the sensitivity of a real event to the model configuration. We have tested several hypotheses, and at the request and suggestion of the reviewer, added additional ensemble members under the hypothesis that with each iteration, the simulation fidelity should improve. We now have an increasingly large and comprehensive ensemble where we have tested sensitivities and which member most closely approaches reality. We would suggest that this is sufficient to provide a useful and interesting contribution for the current work to be published.

In summary:

1. The original work was an 11-member ensemble with 41 vertical levels, using the GNU Fortran compiler and initialized with ERA5 that varied only in terms of microphysics package. As part of this ensemble:
    a. The Morrison hail flag was also tested.
    b. Lateral boundary condition sensitivity (ERA-I vs. ERA5) was also explored using one microphysics scheme (Milbrandt).
    c. Sensitivity to model start time was explored using one microphysics scheme (Milbrandt).
2. After initial peer-review, compiler sensitivity was tested. One microphysics scheme was selected (Morrison), one LBC chosen (ERA5), with two simulations (no hail flag, with hail flag) using the Intel compiler. There were now 13 ensemble members.
3. After second peer-review, the sensitivity to the vertical resolution (and damping – see reviewer comment #4 below) has been tested, using the method in 2 above. There are now 15 ensemble members.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

The two new simulations using 65 vertical levels with damping have been integrated into the manuscript. Each manuscript figure now contains all 15 ensemble members. Tables 3 – 5 have also been updated to include the two new simulations using 65 vertical levels.

In the following 5 comparison figures (Figure 2 – 6 in this reply) and subsets of the manuscript tables 3 – 5 we compare the new simulations (both have '65L' in their displayed names in the figures) to the Morrison and Morrison+Hail results that were part of the original 11-member ensemble in the paper. Note that Figures 2 – 6 in this reply correspond to Figures 4 – 8 from the manuscript.

Along with the comparison figures below, a subset for each of Tables 3 – 5 is shown here to illustrate how the two 65 level simulations compare to the previous Morrison simulations. From Table 3, the simulations with 65 levels do not result in improved agreement in terms of the spatial fields of accumulated precipitation (e.g. see ASOS/RADAR column for precipitation, for example: Morrison 0.187 vs. Morrison 65levels 0.1591). The only significant improvement is against ASOS wind gusts in the Morrison 65levels+Hail simulation vs. the original Morrison+Hail simulation.

Table 3: Spearman rank correlations for the spatial fields of maximum wind gusts in domain d03 during the derecho (Derecho period: 29-Jun-2012 21:30:00 to 30-Jun-2012 13:30:00) from WRF and ASOS observations. In this analysis WRF output for maximum time step wind speeds (dt = 6 sec) is sampled at the 34 ASOS locations and compared with the maximum 3-second ASOS wind gusts measurements (see spatial fields in Figure 11). Also shown are the Spearman rank correlations between spatial fields of total accumulated precipitation from WRF output relative to RADAR estimates and ASOS in situ measurements. In these analyses the correlations between WRF and the RADAR data are for all WRF grid cells sampled by the RADAR (99.4% of d03), while the comparison with ASOS measurements is for the 34 ASOS stations. The final column shows the correlations between the spatial fields of maximum composite reflectivity cREF (again in any time step during the Derecho period) from the WRF ensemble members and RADAR.

Ensemble member	ASOS		RADAR	
	Wind gusts	Precipitation	Precipitation	cREF
Morrison	0.312	0.063	0.187	0.199
Morrison+Hail	-0.557	0.138	0.181	0.255
Morrison 65levels	0.3418	0.0059	0.1591	0.2054
Morrison 65levels+Hail	0.2030	0.1522	0.2604	0.2561

Table 4: Number of grid cells in domain d03 where hail is indicated by the RADARs or present in the WRF simulations during the derecho (Derecho period: 29-Jun-2012 21:30:00 to 30-Jun-2012 13:30:00) and the frontal passage (Front period: 30-Jun-2012 15:20:00 to 01-Jul-2012 14:50:00). Also shown is the number of grid cells with Maximum Estimated Size of Hail (MESH) above 25 mm from the RADAR or WRF. Recall: RADAR detection of hail is re-gridded onto the WRF grid used for domain d03 prior to use in the model evaluation.

	# Grid cells with hail		# Grid cells with hail values > threshold	
	Derecho	Front	Derecho	Front
RADAR	3078	2152	824	813
Ensemble member				
Morrison	0	24	0	0
Morrison+Hail	3000	74398	0	0
Morrison 65levels	2	36	0	0

Morrison 65levels+Hail	37030	72997	1	75

For Table 4, to recap from the manuscript: *"When remapped to the WRF grid, the RADAR data indicate 824 of the almost 90,000 grid cells experienced severe hail during the Derecho period (Table 4). These locations identified by the RADAR detection algorithm as exhibiting hail and MESH > 25 mm …"*
The new Morrison simulation using 65 levels with the hail flag on (Morrison 65levels+Hail), shows greater spatial coverage of hail compared to the original Morrison+Hail simulation (# grid cells with hail for the derecho period is 37030 vs. 3000). This is reflected below in Figure 3 in this reply and in the revised manuscript. Note that greater spatial coverage does not equate to improved fidelity. The number of grid cells with RADAR detection of hail (3078) still shows closest agreement with the Morrison+Hail simulation (3000) (Table 4). The systematic finding of more grid cells with hail when using the hail flag turned on vs. leaving the flag off is also maintained. The front period remains largely unchanged (Table 4 and Figure 4 below), as does the number of grid cells when the threshold of MESH > 25 mm is applied (0 vs. 1 for the derecho period).

A subset of Table 5 in this reply directly compares the Morrison 41 vs. 65 level simulations for the simulation fidelity and convection metrics. The Milbrandt-626-ERA-I and Milbrandt-628-ERA-I metrics are included for reference as these were the best performing ensemble members. The addition of the two new Morrison 65 levels simulations does not change that conclusion. Comparing the Morrison cases directly, the 4 metrics for simulation fidelity confirm that using 65 levels for the Morrison scheme results in enhanced simulation fidelity with and without the hail flag. Compared to the two original Morrison simulations, these two new simulations have the highest agreement with observations of the spatial extent of high cREF, total precipitation accumulation, maximum wind gusts and large hail. The convection metrics are less conclusive, but there is comparatively high values for the 95th percentile downward vertical velocities in the 65 levels simulations.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

Table 5: Metrics of simulation fidelity relative to observations, and convection metrics derived from output from each original Morrison WRF member compared to the new Morrison 65 level simulations during the period of the derecho passage (Derecho period: 29-Jun-2012 21:30:00 to 30-Jun-2012 13:30:00). The metrics of simulation fidelity are described in section 2.2 and are as follows: The Max Gust Ratio: the ratio of the maximum wind gust in any land grid cell from WRF output and observations at the ASOS stations. Total Precip. Ratio: the ratio of the spatial mean total accumulated precipitation from WRF to RADAR and STAGE IV, respectively, for any grid cell with common coverage. cREF>40 dBZ: the ratio of the spatial extent of grid cells with cREF above 40 dBZ at the peak coverage in WRF and RADAR. The color-coding used in this table is as follows; for the measures of simulation fidelity table cells colored red have low fidelity, and those indicated by cyan exhibit relatively high fidelity. For all other cells in the table, a background of orange indicates low values, while blue indicates comparatively high values. The saturation of the color indicates relative ordering of the values. The definitions of each convection metric are given in section 2.3.

	Simulation Fidelity				Convection Metric					
	Max Gust Ratio	Total Precip. Ratio (RADAR)	Total Precip. Ratio (Stage IV)	cREF>40 dBZ Ratio	95% Temperature deviation [-K]	95% SLP deviation [hPa]	Median CAPE loss [J kg^{-1}]	95% -W [ms^{-1}]	Max std(w) height [km]	Z_{R20} [km]
Morrison	0.673	0.413	0.435	0.788	3.29	1.850	1532	0.151	8.0	15.3
Morrison +Hail	0.460	0.016	0.017	0.102	2.12	-0.756	175	0.117	8.0	9.0
Morrison 65levels	0.693	0.454	0.479	0.972	4.35	2.726	1706	0.263	8.0	13.9
Morrison 65levels+Hail	0.820	0.508	0.536	0.850	4.47	-0.193	1620	0.238	8.5	13.1
Milbrandt-626-ERA-I	0.633	0.566	0.597	0.844	5.44	2.360	1960	0.152	8.9	14.4
Milbrandt-628-ERA-I	0.695	0.636	0.671	0.945	5.58	2.790	2030	0.146	7.1	14.9

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

Figure 2 (below) shows a comparison between the original Morrison simulations with 41 levels and new Morrison simulations with 65 levels and damping for the number of grid cells in domain d03 with composite reflectivity (cREF) > 40 dBZ (Figure 2a). In the derecho period it is evident that the simulations with 65 levels show a greater number of grid cells with cREF > 40 dBZ.

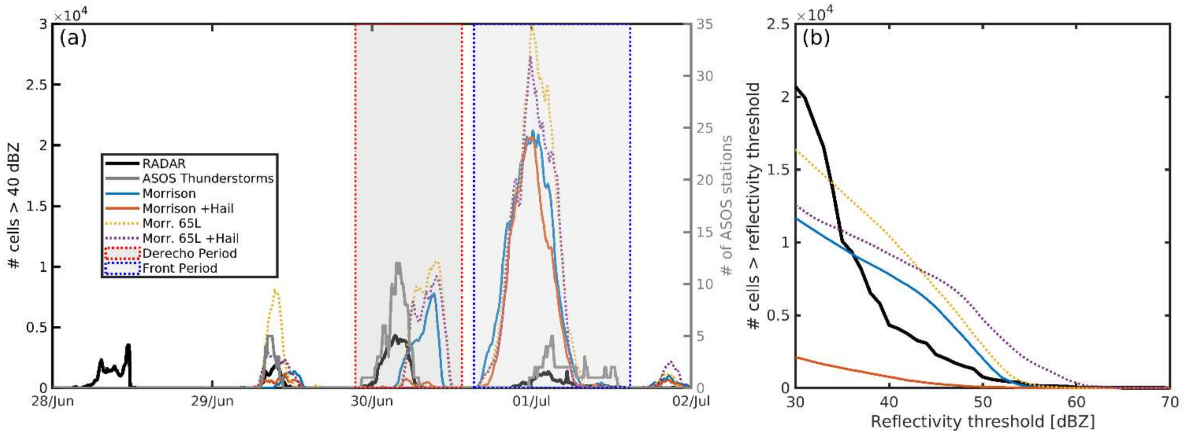

Figure 2: Time series of number of grid cells in domain d03 with composite reflectivity (cREF) > 40 dBZ from RADAR and 4 WRF ensemble members (original Morrison simulations with 41 levels, and new Morrison simulations with 65 levels and damping). The number of the 34 ASOS stations in domain d03 reporting thunderstorms is shown in grey (right axis). The timing of the (Derecho period: 29-Jun-2012 21:30:00 to 30-Jun-2012 13:30:00) and the frontal passage (Front period: 30-Jun-2012 15:20:00 to 01-Jul-2012 14:50:00) are denoted by the grey backgrounds. (b) The number of grid cells in domain d03 where output from each WRF ensemble member or the RADARs exceeded the specified threshold during the time step within the derecho period when the maximum number of grid cells exceeded the threshold. For example, in the RADAR observations there is a single 10-minute period during which approximately 5000 grid cells exhibit a value above 40 dBZ.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

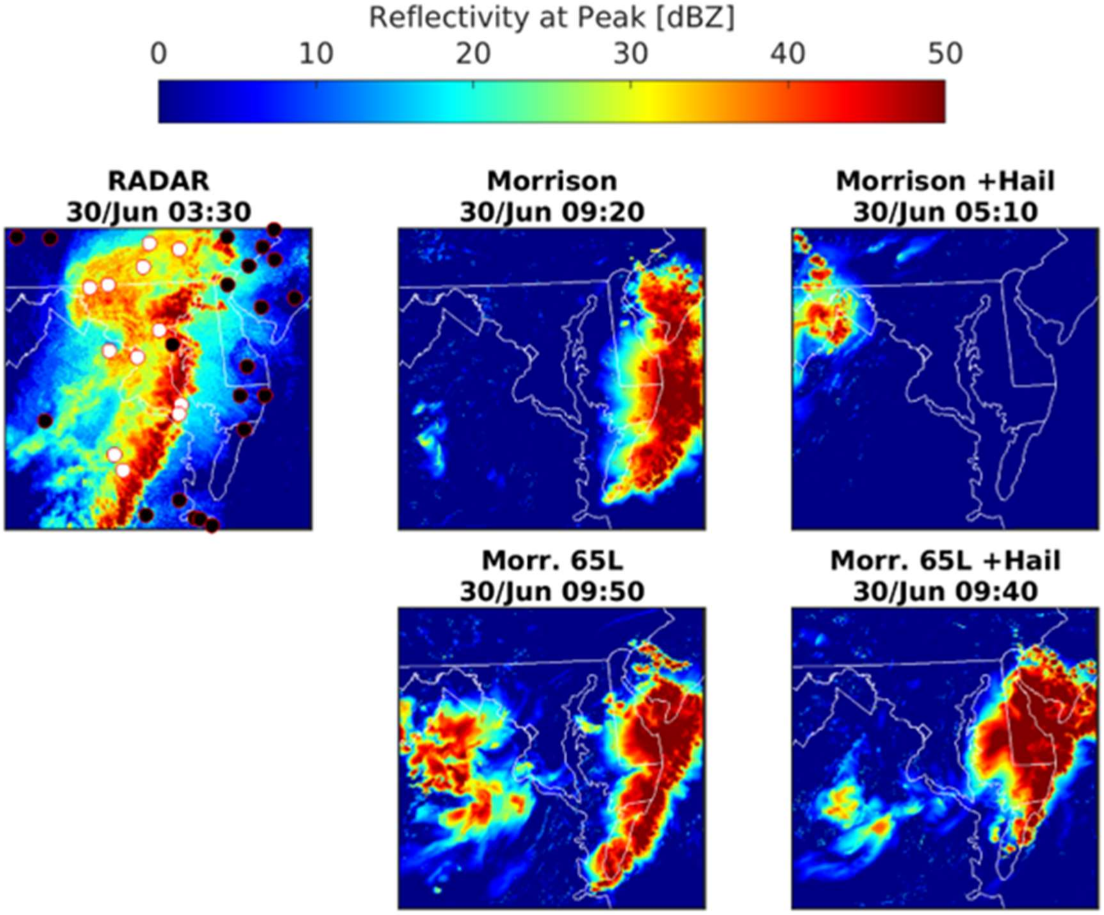

Figure 3: Composite reflectivity (cREF) in domain d03 at t_p (the time when values from the maximum number of grid cells exceeded 40 dBZ) during the Derecho period from RADAR and each WRF ensemble member (times are noted in panel titles). The RADAR panel includes markers showing the presence (white) and absence (black) of thunderstorm reports from ASOS stations in domain d03 in the hour surrounding 03:30 UTC 30 June 2012.

Response to reviewer #1:

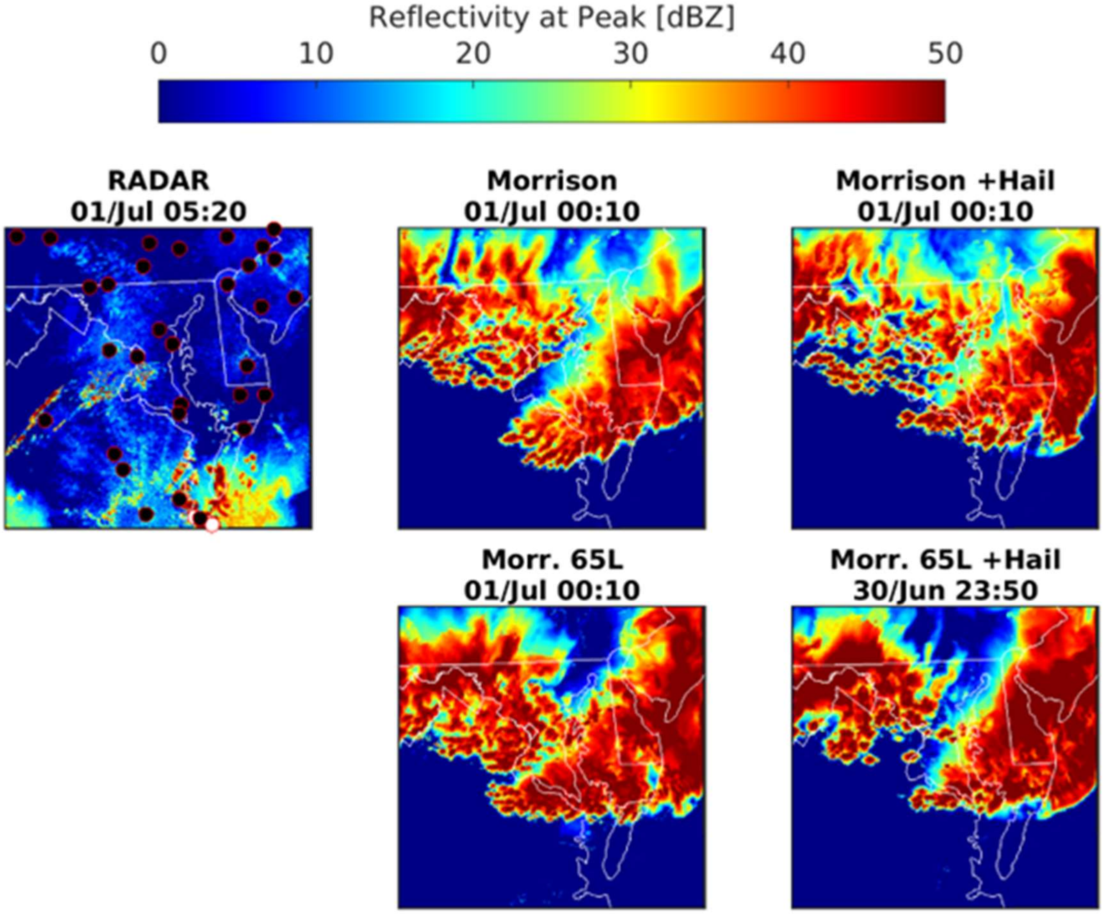

Figure 4: Composite reflectivity (cREF) in domain d03 at t_p (the time when values from the maximum number of grid cells exceeded 40 dBZ) during the Front period from RADAR and each WRF ensemble member (times are noted in panel titles). The RADAR panel includes markers showing the presence (white) and absence (black) of thunderstorm reports from ASOS stations in domain d03 in the hour surrounding 05:20 UTC 1 July 2012.

Response to reviewer #1:

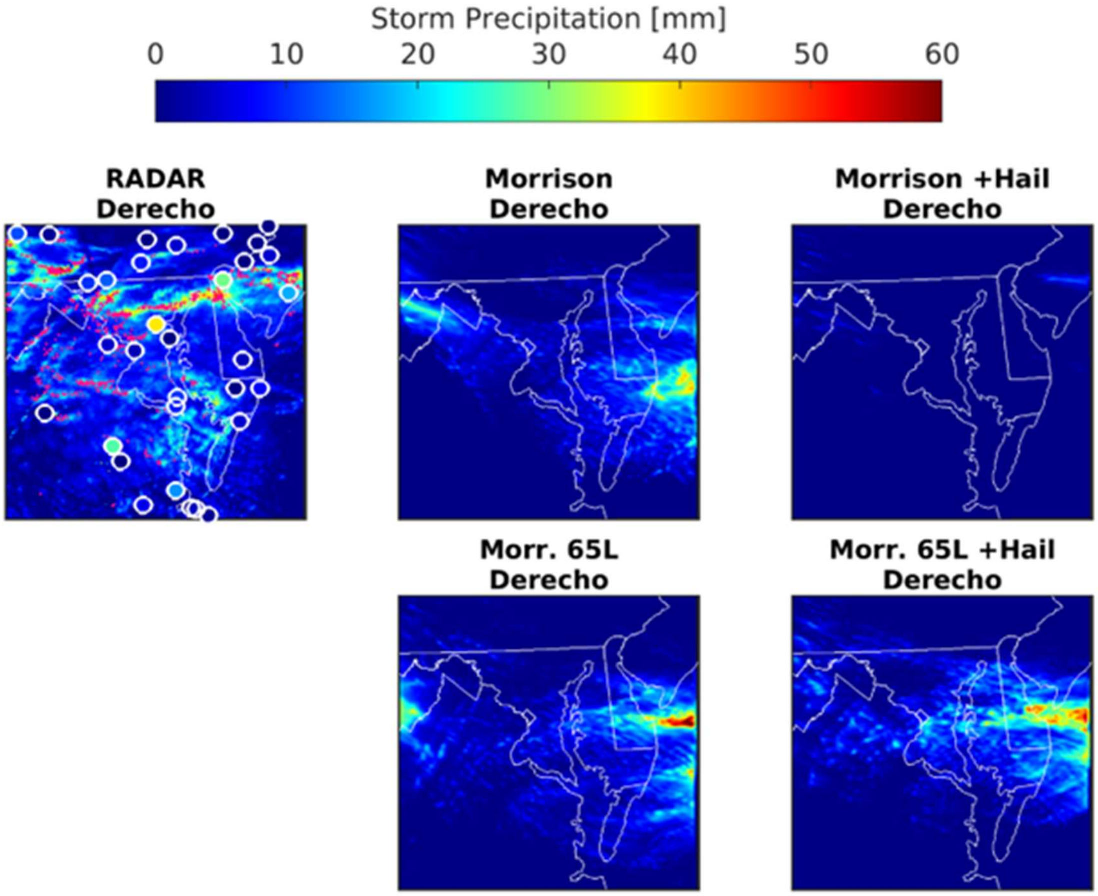

Figure 5: Total accumulated precipitation (mm) from RADAR observations and each WRF ensemble member during the Derecho period. Grid cells with MESH>25mm are marked in magenta.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

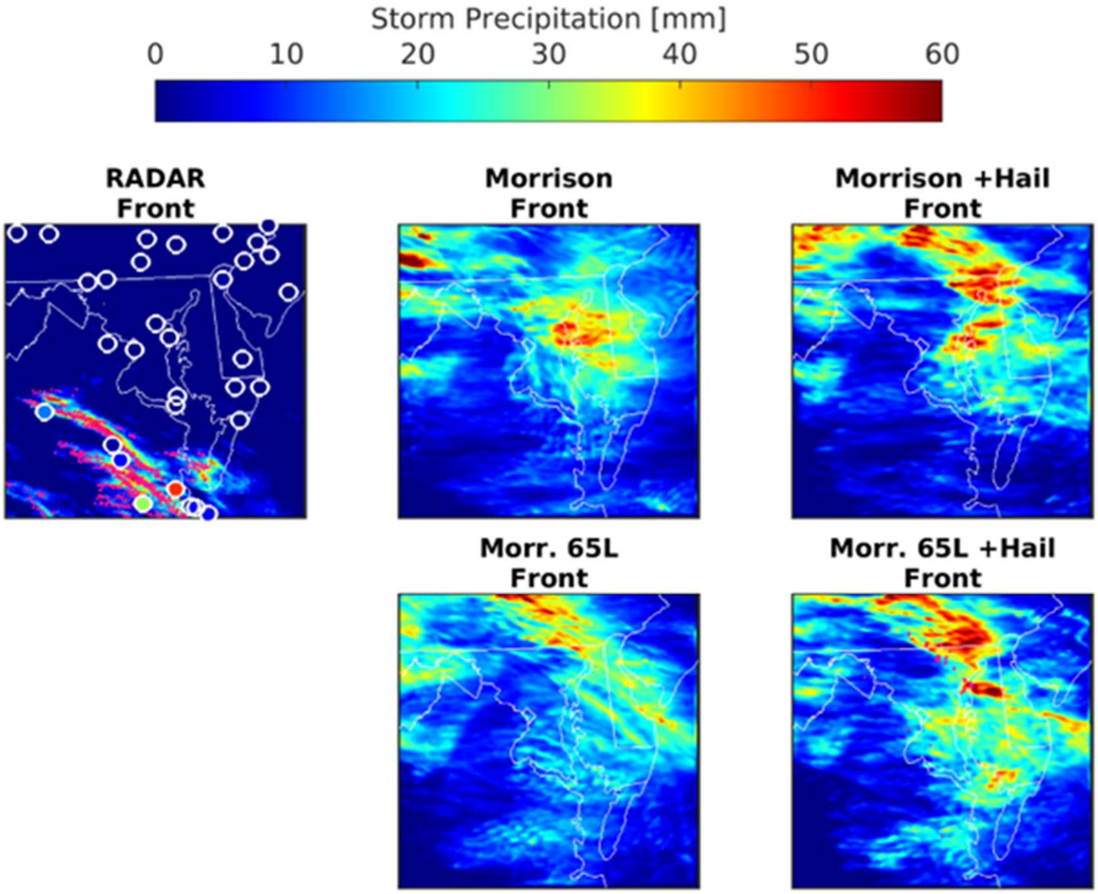

Figure 6: Total accumulated precipitation (mm) from RADAR and each WRF ensemble member during the Front period. Grid cells with MESH>25mm are marked in magenta.

Response to reviewer #1:

4. Upper-level damping is set to be zero, which is needed particularly for convective system simulations.

Past research (e.g. Chen et al. 2019) has indicated damping methods have proven less successful for real atmospheric NWP as opposed to idealized studies, although it is recommended e.g. see: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019jd030968

Nevertheless, on the advice of the reviewer, we have incorporated the reviewer's suggestion and the recommendation in Chen et al. (2019) and have set upper-level damping (damp_opt = 3) for Rayleigh damping (dampcoef inverse time scale [1/s], e.g. 0.2; for real-data cases – see page 5-104 of the WRF v3.8.1 users guide) in the two new simulations undertaken for this response.

The result of using 65 levels and damping was described in comment #3 above along with the accompanying figures.

5. The authors used some physical scheme combinations that I usually do not use, such as the combination they used for the shortwave and longwave radiation, surface layer, and PBL. Therefore, I am not very sure if all those schemes are well coupled together and work well.

One of the many goals of sensitivity studies is to advance the understanding of scheme coupling and interaction, but according to Kehler-Poljak (2017), based on a literature review of mixed physics studies examining sensitivity to convective parameterization (Mayor and Mesquita 2015), microphysics (Weisman et al. 2008; Givati et al. 2012; Horvath and Vilibić 2014), radiation (Kleczek et al. 2014), land surface scheme (Mohan and Bhati 2011) and PBL scheme (Weisman et al. 2008; Acs et al. 2014; Gómez-Navarro et al. 2015; Cohen et al. 2015; Milovac et al. 2016), while the authors of such studies observed general trends, no single combination of scheme coupling for the simulation of convective events was recommended. According to Clark et al. (2010b), ensemble design and model setup should consider the type of forecast fields desired, regional area, similar types of events, but recognize that these schemes interact in a non-linear manner. Note that our study does not attempt to test multiple ensembles of mixed physics for every type of physics scheme – such a study like Clark et al. (2010b) is not the goal here and would require significant computational resources. As stated, our goal was to assess microphysics, with additional simulations to complement this and assess model configuration.

We have employed schemes that are widely used in literature. In response to the reviewer's comment, using the original publications for the schemes used (shortwave: Dudhia (Dudhia 1989); longwave: RRTM (Mlawer et al. 1997); surface layer: revised MM5 Monin-Obukhov (Jimenez et al. 2012); PBL: MYNN level 2.5 (Nakanishi and Niino 2006)), we have undertaken a citation search for deep convection, mesoscale convective system and derecho papers from the last approx. 10 years that reference those schemes, or where possible, use them as a combination. This list is not meant to be exhaustive, but indicative of the breadth of previous studies. Where possible, we have constrained this list to studies over the US. From the studies below, one can observe that even these published studies do not have a common combination of physics schemes.

Citation search results summary:

Response to reviewer #1:

- Duda and Gallus (2013): used the RRTM and Dudhia radiation schemes with the Thompson microphysics scheme, the MYJ PBL scheme, and the Monin-Obukhov surface layer scheme in a 3 km WRF simulation of two mesoscale convective systems.

- Gensini and Mote (2014): used Dudhia and RRTM radiation schemes alongside MYJ PBL and WSM6 microphysics in dynamically downscaled WRF simulations to resolve March–May hazardous convective weather east of the U.S. Continental Divide for a historical climate period (1980–90). A hazardous convective weather model proxy is used to depict occurrences of tornadoes, damaging thunderstorm wind gusts, and large hail at hourly intervals during the period of record.

- Hariprasad et al. (2014): used RRTM, Dudhia, and revised MM5 in a study examining PBL scheme sensitivity for tropical conditions.

- Bodine and Rasmussen (2017): used RRTM and Dudhia alongside Morrison microphysics, Kain-Fritsch cumulus, and YSU PBL in a high-resolution WRF simulation of the 6 July 2015 MCS in South Dakota. Simulations with Thompson microphysics were also run, but Thompson microphysics produced less realistic stratiform rain regions and did not capture the discrete propagation events as well as the simulations using Morrison microphysics.

- Yang et al. (2017): used RRTM, YSU PBL, and Goddard radiation in a study examining mesoscale convective systems over the central US.

- Wang et al. (2022): used RRTM for longwave and shortwave, YSU PBL, revised MM5 in a study examining lake surface temperature impacts on summertime climate over the Great Lakes region. It was shown that warmer lake surface temperature reduces mesoscale convective precipitation upstream of the Great Lakes region, however, isolated deep convective precipitation is enhanced downstream.

In addition, a paper we already cite in our study (Weisman et al. 2013) notes:
"Since 2003, real-time explicit convective forecasts at 3–4-km grid spacing have been run at NCAR during each spring to establish the capabilities of the WRF-ARW modeling system (e.g., Skamarock and Klemp 2008) to forecast severe convective events and to test recent model improvements (e.g., Done et al. 2004; Weisman et al. 2008). These forecasts have also been evaluated yearly alongside a host of forecasts obtained from differing modeling configurations and dynamical cores by a variety of modeling groups as part of the Storm Prediction Center (SPC) and National Severe Storms Laboratory (NSSL) Hazardous Weather Test Bed (HWT) Spring Experiment (e.g., Weiss et al. 2004; Kain et al. 2005, 2006, 2008; Coniglio et al. 2010; Clark et al. 2011)." From Weisman et al. (2013), these well evaluated, real-time explicit convective forecasts use the MYJ PBL scheme, the Noah land surface model, Thompson microphysics, RRTM and Dudhia for longwave and shortwave radiation respectively. Retrospective sensitivity testing between the MYJ and YSU schemes was also undertaken. A discussion of PBL schemes is below.

We appreciate that the combination of surface layer scheme (revised MM5 Monin-Obukhov) and PBL (MYNN) is potentially a combination used in the present study where most uncertainty lies, but our selections are based on supported literature for those schemes and use of these options as standard in WRF (Samelson et al. 2020). There is greater consistency in the use of combined Dudhia and RRTM for radiation. According to Onwukwe and Jackson (2020), the surface-layer schemes compute friction velocity and other exchange coefficients for estimating sensible and latent heat fluxes, and momentum flux from land surface models, and surface stress in the PBL schemes. Furthermore, both the Eta surface layer scheme and revised MM5 are based on similarity theory. This fact accounts for nearly half of the surface layer schemes available in v3.8.1 - the other schemes are less cited in the literature (Pleim-Xiu, QNSE, and TEMF). The Eta similarity, however, includes a parameterization of viscous sublayer according to Janjić (2002), but the underlying theory for each scheme is based on similarity theory. For these reasons, we do not believe the combination used in the present study is a cause for significant concern, especially in the context of existing literature.

We note that several studies have used these schemes in combination. This is a small selection:

- Kumar et al. (2024): used RRTM, Dudhia, MYNN, and revised MM5 in a WRF simulation of a Southern California wildfire event.

- Huang et al. (2023): used revised MM5 and MYNN in convection-permitting simulations of precipitation over the Peruvian Central Andes.

- Minder et al. (2020): used revised MM5 and MYNN in a multi-scheme sensitivity study to help separate sensitivity to PBL and surface layer scheme.

We note that we have previously published this combination of schemes for deep convection simulations over the southern Great Plains – see peer reviewed AMS publications (Letson et al. (2020b); Pryor et al. 2023). Those simulations exhibited some degree of fidelity for key meteorological properties.

Notably, from the citation search, there is a split between use of the YSU and MYJ PBL schemes for the mesoscale convective system studies, with no recommended preference, further reiterating the arbitrary nature of scheme selection by the user. These two PBL schemes use different formulations (non-local [YSU] vs. local closure [MYJ/MYNN]). According to Minder et al. (2020), these variations in PBL scheme formulation have been found to have large impacts on PBL structure and convective storms (e.g., Coniglio et al. 2013; García-Díez et al. 2013; Cohen et al. 2015; Milovac et al. 2016). There are documented biases in the literature for different types of PBL scheme. Weisman et al. (2008) identified biases in a sensitivity study between YSU and MYJ: *"The YSU scheme tends to create boundary layers that are deeper and drier, and is also very aggressive in eliminating capping inversions. On the other hand, the MYJ scheme tends to deepen the boundary layer more slowly, resulting in PBL conditions that are characteristically cooler, moister, and more strongly capped." and "...cases run with the MYJ scheme show that it was consistently better at maintaining the boundary layer moisture than YSU".*

Note that MYNN uses a similar baseline as MYJ, but with considerable modifications of certain parameterizations, e.g., mixing length scale, eddy diffusivities, and stability functions in the stable PBL. The scheme has been modified and improved since its inclusion in WRF. The fact it is based on MYJ makes it a reasonable analog to biases in MYJ, but should be viewed in the context of improvements made to MYNN in recent years.

A study examining the sensitivity and interaction between the land surface schemes (i.e. Noah) and PBL schemes in a case study over western Germany (Milovac et al. 2016), showed that the representation of land surface processes significantly impacts the simulation of mixing properties within the convective boundary layer. In addition, the nonlocal PBL schemes (YSU) simulated a deeper and drier boundary layer than the local schemes (MYJ and MYNN), which is consistent with Weisman et al. (2008). Note that our land surface scheme selection in the present study (Noah) is consistent with the literature.

In summary, we appreciate the reviewer's comments about scheme selection and coupling – our choices were not random, but based on existing literature, known trade-offs and biases, and ultimately informed by science.

6. Another note is that the radiation schemes they only take the hydrometeor mass from the cloud microphysics calculation. The cloud droplet and ice effective radii are fixed in the radiation scheme and not from cloud microphysics calculation. For better coupling of cloud microphysics with radiation, the default WRF is not enough and the users need to do the coupling in cloud and ice effective radii.

We appreciate the reviewer's comment, but such an exercise is a model/scheme development study and is outside the scope of the current study (an ensemble sensitivity study). Indeed, previous sensitivity studies like the present study run the microphysics schemes with no user-specific code modification. The literature we have cited here that uses the specific radiation schemes, includes no specific modification to the default WRF schemes.

The large differences (100% or even more) in their tests between NERSC Cori and their university supercomputer for the Morrison scheme with the hail option (Figure 13b) is an indication that the code or computer is not stable. Changing compilers or computer would never cause such large differences which basically make the model simulations not trustable. Note the code instability may not be shown in any tests and might only be triggered in some conditions. Since WRF has so many options for physics and model configurations, and many options were plugged in only for some specific schemes and might not be tested for other schemes, users need to be very careful and better work with experienced WRF modelers or relevant scheme developers to get things figured out when unexpected/unusual results are obtained.

We showed in our previous response that the simulations between platforms, and the Intel vs GNU tests are quite similar. We have worked with scheme developers– we spoke with Hugh Morrison to gain an appreciation for the difference between the Morrison simulations and we communicated that in our previous response.

With appreciation for the reviewers careful attention to our paper, and our lengthy and detailed responses and additional simulations we hope that this process can now be drawn to a conclusion and our paper move into publication.
Thanks to everyone for their time and effort on this.

References

Acs, F., Gyöngyösi, A. Z., Breuer, H., Horváth, A., Mona, T., & Rajkai, K. (2014). Sensitivity of WRF-simulated planetary boundary layer height to land cover and soil changes. Meteorologische Zeitschrift, 23, 279–293.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

Bodine, D. J., and K. L. Rasmussen, 2017: Evolution of Mesoscale Convective System Organizational Structure and Convective Line Propagation. Mon. Wea. Rev., 145, 3419–3440, https://doi.org/10.1175/MWR-D-16-0406.1.

Chen, Y., Yuan, H., and Gao, S.: A High-Resolution Simulation of Roll Convection Over the Yellow Sea During a Cold Air Outbreak, Journal of Geophysical Research: Atmospheres, 124, 10608-10625, https://doi.org/10.1029/2019JD030968, 2019.

Clark, A. J., W. Gallus, and T. Chen, 2007: Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models. Mon. Wea. Rev., 135, 3456–3473.

Clark, A. J., W. A. Gallus , M. Xue, and F. Kong, 2010b: Growth of Spread in Convection-Allowing and Convection-Parameterizing Ensembles. Wea. Forecasting, 25, 594–612, https://doi.org/10.1175/2009WAF2222318.1.

Cohen, A. E., Cavallo, S. M., Coniglio, M. C., & Brooks, H. E. (2015). A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. Journal of Applied Meteorology and Climatology, 30, 591–612.

Coniglio, M. C., J. Correia Jr., P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. Wea. Forecasting, 28, 842–862, https://doi.org/10.1175/WAF-D-12-00103.1.

Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecast (WRF) model. Atmos. Sci. Lett., 5, 110–117.

Duda, J. D., and W. A. Gallus , 2013: The Impact of Large-Scale Forcing on Skill of Simulated Convective Initiation and Upscale Evolution with Convection-Allowing Grid Spacings in the WRF. Wea. Forecasting, 28, 994–1018, https://doi.org/10.1175/WAF-D-13-00005.1.

Garcia-Carreras, L., A. J. Challinor, B. J. Parkes, C. E. Birch, K. J. Nicklin, and D. J. Parker, 2015: The Impact of Parameterized Convection on the Simulation of Crop Processes. *J. Appl. Meteor. Climatol.*, 54, 1283–1296, https://doi.org/10.1175/JAMC-D-14-0226.1

García-Díez, M., J. Fernández, L. Fita, and C. Yagüe, 2013: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. Quart. J. Roy. Meteor. Soc., 139, 501–514, https://doi.org/10.1002/qj.1976.

Gensini, V. A., and T. L. Mote, 2014: Estimations of Hazardous Convective Weather in the United States Using Dynamical Downscaling. J. Climate, 27, 6581–6589, https://doi.org/10.1175/JCLI-D-13-00777.1.

Givati, A., Lynn, B., Liu, Y., & Rimmer, A. (2012). Using the WRF model in an operational stream forecast system for the Jordan River. Journal of Applied Meteorology and Climatology, 51, 285–299.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

Gómez-Navarro, J. J., Raible, C. C., & Dierer, S. (2015). Sensitivity of the WRF model to PBL parametrisations and nesting techniques: Evaluation of wind storms over complex terrain. Geoscientific Model Development, 8, 3349–3363.

Halder, M., Mukhopadhyay, P. Microphysical processes and hydrometeor distributions associated with thunderstorms over India: WRF (cloud-resolving) simulations and validations using TRMM. Nat Hazards 83, 1125–1155 (2016). https://doi.org/10.1007/s11069-016-2365-2

Hariprasad, K.B.R.R., C.V. Srinivas, A.Bagavath Singh, S. Vijaya Bhaskara Rao, R. Baskaran, B. Venkatraman, Numerical simulation and intercomparison of boundary layer structure with different PBL schemes in WRF using experimental observations at a tropical site, Atmospheric Research, 2014, https://doi.org/10.1016/j.atmosres.2014.03.023.

Horvath, K., & Vilibić, I. (2014). Atmospheric mesoscale conditions during the Boothbay meteotsunami: a numerical sensitivity study using a high-resolution mesoscale model. Natural Hazards, 74, 55–74.

Huang, Y., M. Xue, X. Hu, E. Martin, H. M. Novoa, R. A. McPherson, A. Perez, and I. Y. Morales, 2023: Convection-Permitting Simulations of Precipitation over the Peruvian Central Andes: Strong Sensitivity to Planetary Boundary Layer Parameterization. J. Hydrometeor., 24, 1969–1990, https://doi.org/10.1175/JHM-D-22-0173.1.

Janjić, Z. I., 2002: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. NCEP Office Note 437, 61 pp., https://repository.library.noaa.gov/view/noaa/11409.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. Wea. Forecasting, 23, 931–952.

Kehler-Poljak, G., Telišman Prtenjak, M., Kvakić, M. et al. Interaction of Sea Breeze and Deep Convection over the Northeastern Adriatic Coast: An Analysis of Sensitivity Experiments Using a High-Resolution Mesoscale Model. Pure Appl. Geophys. 174, 4197–4224 (2017). https://doi.org/10.1007/s00024-017-1607-x

Khai Shen Sow, Liew Juneng, Fredolin T. Tangang, Abdul Ghapor Hussin, Mastura Mahmud, Numerical simulation of a severe late afternoon thunderstorm over Peninsular Malaysia, *Atmospheric Research*, 99, 2, 2011, https://doi.org/10.1016/j.atmosres.2010.10.014

Kleczek, M. A., Steeneveld, G.-J., & Holtslag, A. A. M. (2014). Evaluation of the weather research and forecasting mesoscale model for GABLS3: Impact of boundary-layer schemes, boundary conditions and spin-up. Boundary-Layer Meteorology, 152, 213–243.

Kumar, M., Kosović, B., Nayak, H. P., Porter, W. C., Randerson, J. T., and Banerjee, T.: Evaluating the performance of WRF in simulating winds and surface meteorology during a Southern California wildfire event, Frontiers in Earth Science, 11, 10.3389/feart.2023.1305124, 2024.

Letson, F., T. J. Shepherd, R. J. Barthelmie, and S. C. Pryor, 2020b: WRF modelling of deep convection and hail for wind power applications. J. Appl. Meteor. Climatol., 59, 1717–1733, https://doi.org/10.1175/JAMC-D-20-0033.1.

Liu, C., and M. W. Moncrieff, 2007: Sensitivity of cloud-resolving simulations of warm-season convection to cloud microphysics parameterizations. Mon. Wea. Rev., 135, 2854–2868.

Mayor, Y. G., & Mesquita, M. D. S. (2015). Numerical simulations of the 1 May 2012 deep convection event over Cuba: Sensitivity to cumulus and microphysical schemes in a high-resolution model. Advances in Meteorology, 2015. doi:10.1155/2015/973151

Milovac, J., Warrach-Sagi, K., Behrendt, A., Späth, F., Ingwersen, J., & Wulfmeyer, V. (2016). Investigation of PBL schemes combining the WRF model simulations with scanning water vapor differential absorption lidar measurements. Journal of Geophysical Research, 121, 624–649 (2016). doi:10.1002/2015JD023927.

Minder, J. R., W. M. Bartolini, C. Spence, N. R. Hedstrom, P. D. Blanken, and J. D. Lenters, 2020: Characterizing and Constraining Uncertainty Associated with Surface and Boundary Layer Turbulent Fluxes in Simulations of Lake-Effect Snowfall. Wea. Forecasting, 35, 467–488, https://doi.org/10.1175/WAF-D-19-0153.1.

Mohan, M., & Bhati S., (2011). Analysis of WRF model performance over subtropical region of Delhi, India. Advances in Meteorology, 2011. doi:10.1155/2011/621235.

Onwukwe, C., and P. L. Jackson, 2020: Meteorological Downscaling with WRF Model, Version 4.0, and Comparative Evaluation of Planetary Boundary Layer Schemes over a Complex Coastal Airshed. J. Appl. Meteor. Climatol., 59, 1295–1319, https://doi.org/10.1175/JAMC-D-19-0212.1.

Pryor, S. C., F. Letson, T. Shepherd, and R. J. Barthelmie, 2023: Evaluation of WRF Simulation of Deep Convection in the U.S. Southern Great Plains. J. Appl. Meteor. Climatol., 62, 41–62, https://doi.org/10.1175/JAMC-D-22-0090.1.

Rajeevan, M., Kesarkar, A., Thampi, S. B., Rao, T. N., Radhakrishna, B., and Rajasekhar, M.: Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over Southeast India, Ann. Geophys., 28, 603–619, https://doi.org/10.5194/angeo-28-603-2010 2010

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model Bias in a Continuously Cycled Assimilation System and Its Influence on Convection-Permitting Forecasts. Mon. Wea. Rev., 141, 1263–1284, https://doi.org/10.1175/MWR-D-12-00112.1.

Samelson, R. M., L. W. O'Neill, D. B. Chelton, E. D. Skyllingstad, P. L. Barbour, and S. M. Durski, 2020: Surface Stress and Atmospheric Boundary Layer Response to Mesoscale SST Structure in Coupled Simulations of the Northern California Current System. Mon. Wea. Rev., 148, 259–287, https://doi.org/10.1175/MWR-D-19-0200.1.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. Mon. Wea. Rev., 137, 3351–3372.

Shepherd T.J., Letson F., Barthelmie R.J. and Pryor S.C. How well are hazards associated with derechos reproduced in regional climate simulations? *Natural Hazards and Earth System Sciences Discussions (*nhess-2021-373)

Response to reviewer #1:

Sobash, R., J. Kain, D. Bright, A. Dean, M. Coniglio, and S. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. Wea. Forecasting, 26, 714–728.

Wang, J., Xue, P., Pringle, W., Yang, Z., & Qian, Y. (2022). Impacts of lake surface temperature on the summer climate over the Great Lakes Region. Journal of Geophysical Research: Atmospheres, 127, e2021JD036231. https://doi.org/10.1029/2021JD036231

Weisman, M. L., Christopher, D., Wang, W., Manning, K. W., & Klemp, J. B. (2008). Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. Weather and Forecasting, 23, 407–437.

Xing Yu & Tae-Young Lee (2010) Role of convective parameterization in simulations of a convection band at grey-zone resolutions, Tellus A: Dynamic Meteorology and Oceanography, 62:5, 617-632, DOI: 10.1111/j.1600-0870.2010.00470.x