We thank both referees for their comments on our paper. We found them very helpful in improving the manuscript.

Comments from Referee #1

This manuscript evaluates a commonly used parameterization for lightning flash rate to determine its applicability in mesoscale models. The Price and Rind (1992) parameterization uses the cloud top height as the predictor of total flash rate and a formulation using cold cloud depth (Price and Rind, 1993) to determine the IC/CG flash ratio. These schemes are tested in the WRF model at 36- and 12-km horizontal resolution. A 4-km (cloud-resolving) resolution run was also conducted which used a related parameterization scheme involving the maximum vertical velocity as the flash rate predictor. The model flash rates are compared with observations from the National Lightning Detection Network (NLDN) and the Earth Networks Total Lightning Network (ENTLN). These parameterizations had not previously been thoroughly evaluated in a mesoscale model, and thus this work provides valuable advice in their application, especially in terms of the IC/CG ratio and grid resolution "calibration factor". In general, the manuscript is well written and logically presented, but a number of clarifications are needed in the text. I recommend that it be published after addressing the specific comments listed below:

p. 3494, Line 24: also mention the first paper addressing the role of ozone as a greenhouse gas in the upper troposphere which is Lacis et al. (1990, JGR).

Citation added

pages 3495-3496: The manuscript provides a summary of the various flash rate parameterization schemes used in models of all scales. Please separate the discussion into separate paragraphs for those used in global or regional models with parameterized convection and those used in cloud-resolved models. Add mention of the scheme by McCaul et al. (2009, Weather and Forecasting) which is used in the operational NOAA HRRR model at 3-km resolution.

Thanks for the suggestion. We have separated the text into two paragraphs according to quantitative vs. qualitative indices and added the McCaul et. al. citation.

Page 3502: Some further clarification needed concerning the NLDN data. Weak positive polarity flashes detected by NLDN are thought to be IC flashes. NLDN data from 2011 are segregated into IC and CG flashes and the user probably does not need to be concerned with this issue. However, the original 2006 data did not have these designations, and the weak positive flashes had to be filtered out of the NLDN data set by the user. Were the 2006 data you used in this original form, or were they a reprocessed data set given in the same format as the 2011 data? If they were the original data, the filtering needs to be done before they are used in this analysis.

The reprocessed data were used. We added "Weak positive flashes with <15 kA have been filtered from all data."

page 3503, line 5: The authors have used a constant value of 65% detection efficiency for IC flashes from ENTLN, while the values may range from 50 to 85%. Please provide an estimate of how much error the estimate of 65% could induce in the IC/CG ratio.

We have added to section 3 more discussion on the range of possible flash counts, IC:CG ratios, and biases (p. 3305, line 4; p. 3506, lines 3-6). In section 2.3, we added: "Data from neither network were provided with the mapping of the corresponding DEs. To account for the spatial variability in DE, the range of possible flash counts, IC:CG ratios, and biases will be provided when appropriate within the discussion in Section 3."

page 3503, line 25: add the following phrase at the end of the sentence: "...except at the high end of the distribution."

We have added "…except at the high end of the distribution where limits of model grid resolution induce significant noise."

page 3505, line 4: Is this spatial variation of bias within the analysis region? Or, is this the range when the spatial variation of DE is taken into consideration? I don't understand why a range of the median bias is given here. It would seem to me that the variations in the DE fro CG flashes would have been taken into account BEFORE the median bias was calculated (ie., in determining the NLDN CG flash count).

This is the range when the range of possible DE is taken into account according to the values cited by Orville et al 2010 (Vaisala: >90%, Biagi et al 2007: 90-95%). No precise mapping of DE was provided with the data. See response to the previous comment.

page 3506, lines 3-6: ....for ENTLN CG and IC flashes... If you know the spatial distribution of DE, why not apply them rather than using constant values?

The range of spatial distribution is inferred from a map provided by Earth Networks in Fig 3 of Liu and Heckman 2011. The precise mapping of DE is not provided with the data, and that the equations used to derive the map may not be valid due to continuous network improvement. See also response to comment re: 3503 line 5.

page 3507, line 14: factor of ~10 bias. Is it a high or low bias?

We have clarified to: "… a factor of ~10 high bias" page 3508, line 1: double the 3-hourly OBSERVED lightning? But, JJA 2011 showed only 13% median daily bias for 36 km (p. 3505, line 3). Why is the 3-hour bias so much greater?

We have clarified this information in section 3.2 that the differences in the 3-hourly lightning used in that particular comparison is due to subsampling JJA 2011.

"It should be noted that these values were obtained by sampling all 3 months. Sampling one month would produce varying results. For instance, while JJA 2011 produces an overall median bias of 13%, July 2011 alone produces about twice as much lightning as observed but is offset by under-predictions in June and August."

page 3508, line 7: identical frequency distribution. For both 12-km and 36-km?

This sentence has been reworded to emphasize that the grid-flash frequency distribution and their similarity lie in the location of the drop-off:

"…able to produce the same drop-off in grid frequency distribution beyond 200 flashes per grid per 3-hour…"

Figure 2: Can a different color scale be used. It is difficult to differentiate between the values in the 100 to 1000 range (a factor of 10) to assess model performance because the colors are so close to the same.

The color scale has been modified to emphasize the 100-1000 range (Figure R1)

Figure 3: Figure needs improvement. I can't see solid versus dotted red lines on the time series plots.

We now use thicker lines that are red and blue instead of solid and dotted. (Figure R2)

Figure 9: I don't see the black line for "WRF online adjust". Is it coincident with the blue line?

Yes but only in the time series. Differences show up when comparing the frequency distributions. Added clarification in the discussion and figure description.

References

Liu, C. and Heckman, S.: Using total lightning data in severe storm prediction: global case study analysis from north America, Brazil and Australia, in: 2011 International Symposium on Lightning Protection (XI SIPDA), 20–24, 2011.

Orville, R. E., Huffines, G. R., Burrows, W. R., and Cummins, K. L.: The North American lightning detection network (NALDN) – analysis of flash data: 2001–09, Mon. Weather Rev., 139, 1305–1322, doi:http://dx.doi.org/10.1175/2010MWR3452.110.1175/2010MWR3452.1, 2010.

Comments from Referee #2

The paper "Evaluating a lightning parameterization based on cloud-top height for mesoscale numerical model simulations“ by Wong et al. deals with parameterized lightning using the commonly applied lightning scheme by Price & Rind in the regional chemistry weather model WRF-CHEM.

The paper is well written and points to interesting results in the applied system; however for me it is relatively difficult to assess to which degree the results from this study can be applied to similar modeling questions, e.g. to other regional chemistry models. Hence, the paper is missing a "discussion section", in which this question could be answered. After clarification of some minor aspects, the paper merits publication in GMD.

The answer to this question is largely addressed in the conclusion section, in which separate paragraphs reiterate and summarize the recommendations for how to apply PR92, PR93, and PR94 respectively. Detailed reasoning and discussion is included within the results section, accompanying the data results. It is now renamed "Results and Discussions." The main recommendations of how the results from our study should be applied to other regional-scale models are 1) adjust cloud top to match the 20 dBZ radar reflectivity top, 2) avoid PR93 unless it is not necessary to get information on the frequency distribution of flashes, and 3) scale by the area ratio for resolutions against that which produces approximately 1 storm per grid within the domain of interest (36 km grid spacing in this study).

1. The Grell parameterisation of convection should be described in a little more detail (e.g. how the convective microphysics are treated) as it is crucial for the input to the lightning scheme.

We added the following to the first paragraph of Section 2.2:

The implementation of the GD scheme employed in this study consists of 3x3x16=144 ensemble members comprised of interactions between different dynamic control closures and static control/feedback closures. The maximum moist static energy (MSE) is then used as input with entrainment to calculate the level of neutral buoyancy (LNB), or cloud top. For further information about the convective parameterization, readers are encouraged to refer to Grell (1993).

2. Precipitation is used as a measure how well the convective activity is reproduced by the model. This is a little critical as lightning activity depends (in reality) more on other aspects of convective cells than the precipitation, i.e. the dynamical part of the convective clouds such as the vertical motion. Of course, the convective precipitation provides a reasonable estimate of the spatial distribution of convective activity and hence lightning. However, analyzing the mean convective mass fluxes and their spectra would shed more light onto the substantial changes when other horizontal resolutions are applied.

The reviewer brings up an important comment of what is the best way to measure convective activity in the model and what parameters best connect to lightning. Lightning flash rates are mostly a result of graupel-ice collisions brought about by both dynamics (vertical motion) and microphysics (especially graupel formation). Modeling studies have shown that the production of graupel and snow are also correlated with precipitation amounts. Thus, lightning and precipitation depend on similar processes, but differ in that precipitation is also affected by “warm rain” production and evaporation. On the other hand, convective mass fluxes rely more on the dynamics part of the equation, although they may include the buoyancy term which is a product of the microphysics.

We have added the following to the beginning of section 3.1 to explain why precipitation is used rather than other measures e.g. convective mass flux.

"While lightning does not directly depend on precipitation, they are both the result of the same processes that promote ice-graupel collisions. Further, precipitation is observed robustly and continuously, thus giving us a high quality measurement for validating model results. On the other hand, while convective mass flux may produce a more consistent correlation with lightning, the lack of well-controlled direct observations and the large uncertainty in model calculations make it an inferior proxy for convective strength in this context."

3. In Figure 2, the strong precipitation regions are difficult to distinguish, however, they are the most interesting ones for lightning activity. Even though the precipitation spectra match the observations well, the strongest extreme events are not captured, which might have some implications on the lightning results.

We have improved color scale in figure 2.

4. In Figure 3, the dotted line appears to be more or less on the solid one; hence all the precipitation is subgrid-scale. Even though convection will substantially contribute to total precipitation, it is a little surprising that none of the precipitation events are on the scales of the grid cells and therefore handled by the grid-scale cloud scheme, especially in the eastern part of the analysis region.

Figure 3 has been improved to show that there are instances where a non-minuscule portion of the precipitation happens at the resolved grid-scale.

5. The interplay of the convection parameterization and the grid scale clouds is crucial for the determination of lightning. Even though precipitation seems to be dominated by subgrid-scale rainfall, the properties of the associated clouds on the grid scale can be of relevance for the lightning production. Starting at 36km this becomes even more prominent for the smaller scale grid cells with improved resolution. This interplay should be analysed in more detail.

Even though the LNB comes from the convective parameterization, it is representative of the storm dimension permitted on the grid-scale (keep in mind we are not parameterizing against precipitation but storm dimension as proxied by LNB). The calculation of LNB and the convective parameterization in general, uses grid-scale environmental variables as input, and thus the interplay of grid-scale and subgrid-scale convection is not completely excluded from a lightning parameterization using this LNB.

Moreover, even when precipitation and area wherein flashes are triggered occur on the grid-scale, electrification often occurs within or near the convective cores (e.g. Wiens et al 2005 JAS analysis of STEPS 2000 storms, or from explicit electrification modeling studies e.g. Mansell et al JGR 2002), which are much smaller and thus are more appropriately captured by the convective parameterization.

6. The model overestimates the lightning substantially along the Eastern coast of the US, similar to the precipitation; is this due to too much moisture input from the Gulf of Mexico or caused by other meteorological features?

It is unlikely that the input or large scale meteorology are the causes as meteorology (U, V, T, water vapor) is nudged to NCEP GFS reanalysis. That said, during nudging, there is a possibility that convection may be removing excessive moisture in comparison to GFS and thus additional moisture may be introduced. However, the goal of this paper is not to evaluate the convective parameterization nor specific model set-up, but rather to evaluate the performance characteristics of the lightning parameterization when implemented into a regional model with all the expected (and unexpected) defects.

To address the imperfect simulation of convection and thus precipitation over an entire summer, model-model and obs-obs flash rate vs precipitation are compared and illustrated in Figure 6 and discussed in the 3rd paragraph of section 3.2. In spite of the bias in simulating convection, the comparison does show that the model produces the appropriate order of magnitude of lightning and a realistic physical relationship with another proxy for convective strength (in this case precipitation).

6 (cont'd) On the other hand, can you provide a (physical) reason why the lightning activity in the central part of your analysis region is underestimated? According to the precipitation it is not obvious where this underestimation originates from? This potentially supports the aspect of point #5.

It is not clear where the reviewer is referring to. We find, in Figure 4, that lightning activity is similar or overestimated in the analysis region instead of underestimated.

7. How well do the total simulated flashes match between WRF-CHEM and LIS/OTD data. Maybe the differences originate from the CG:IC ratio and its parameterization and not only from the convective activity.

An attempt to compare LIS data with WRF cross-track coincident pixel flash count has been made. It should be noted that a large number of assumptions and filters are used to accomplish this comparison due to several limitations of the instrument. It is assumed that LIS has a 0.40 - 0.65 geographical-invariant, diurnal-invariant detection efficiency, and that IC and CG flashes have the same DEs. A single WRF grid is sampled and scaled by the ratio of 36 km to 0.5 deg pixel dimensions for each LIS footprint. There is also a latitudinal sampling bias, for which some latitudes were sampled (up to ~2.5x) more frequently than others. Only pixels with more than 60 seconds of effective view time by LIS are used. Finally, there is a shift in diurnal sampling bias throughout JJA 2006, for which different hours were sampled differently for different months.

The following results are for continental flash counts east of 105 W and north of 25 N:

	total	max[1]
Adjusted LIS (0.65):	37655	170
Adjusted LIS (0.40):	61190	277
Sampled[2] WRF:	60290	21
Area-scaled WRF[3]:	143294	51

1. maximum cross track 1-hour flash count
2. sampled single cross-track WRF grid at the center of individual LIS footprint, representing minimum flash count sampled
3. instantaneous predicted flash rate x LIS pixel effective view time, representing high-end sampling

The total lightning bias after areal scaling, which is ~2-3x too high, is consistent with CG comparison in section 3.2. If no areal scaling is invoked the WRF flash rate prediction is within LIS range if not scaled by footprint area. The underestimation of the maximum permitted flash count per pixel is demonstrated in the analysis against NLDN, which can be found in the last paragraph of section 3.2.

8. Please provide a reason, why you use the (modified) LNB as a proxy and not the convective cloud top height, or the level of maximum detrainment. If the convection is parameterized, then this would be more consistent than neglecting all entrainment modifications of the buoyant parcel, and assuming simple or complex correction terms.

The LNB used in the parameterization does take into account convective entrainment and detrainment functions within the parameterization. We added the following to clarify this point: "…, with convective entrainment and detrainment accounted for within the calculation of cloud moist static energy, …" to the 3rd paragraph of section 2.2.

Another reason why "LNB" is used is that it is readily available within the model when the Grell option is used. This point is stated in the same paragraph. Other proxies such as level of maximum detrainment may be available in other models, and they will require different correction terms to reconcile their differences with radar reflectivity height, which is the cloud top used in deriving PR92. This is discussed in the appendix.

9. Analysing the resolution dependency you find a substantial overestimation with smaller grid size, such that you need an additional correction factor. You provide an argument that this is needed as the data from which the parameterization has been derived has been compiled on much larger scales. For me, this is not an obvious explanation, as already the 36km are substantially smaller than the original grid size, and hence a reduction of 1/3 of the grid size, does not substantially change the physically resolved processes, i.e. individual convective cells are still parameterized, whereas organized convective systems will be (partly) resolved in both WRF-CHEM simulations whereas they are still subgrid scale with respect to the original 8x10○.

There are two aspects of the arguments: (1) The extrapolation argument serves as the most plausible explanation for why PR94 does not work when it has been shown to work for global models. (2) The argument for why scaling by areal ratio at the tested resolutions should be performed is re-summarized as follow and is elaborated at similar length in the 1st paragraph of section 4.1:

PR92 gives flash per storm, thus when approaching almost convection-resolving resolutions, where the expected storm size is comparable to the grid size, the appropriate scaling should be done according to number of storms per grid. Since Δx=36 km gives a reasonable result (within 1 order of magnitude over 3 months), we assumed 1 storm per grid at this grid size (base case), and scale the flash counts from the Δx=12 km simulation accordingly. The ambiguity of using Δx=36 km as the base case is also discussed in the third paragraph in the conclusion.

Since PR94 evaluates to ~1 for both 12 km (c=0.973) and 36 km (c=0.977) (given at the end of section 2.1), "additional" may not be the appropriate word to describe the use of areal ratio. It is, rather, a proposed replacement for the PR94 calibration factor at these resolutions when one resolution is known to produce good results.

9 (cont'd) Is the convection getting more intense with the smaller grid size? To which degree do the changes in resolved versus subgrid scale clouds contribute to this? How much does the vertical grid scale velocity change, i.e. what is the w_max in both simulations?

Δx=12 km simulation is showing higher total precipitation (inferring convective activities) largely due to significant over-prediction between July 15 and 19, but otherwise within the variability of NWS and 36 km (Figure R3). What is unexpected, however, is that the convective rainfall fraction at 12 km is consistently larger than at 36 km. This is likely due to the same problem with the model settings that caused the diurnal cycle anomalous bias as elaborated in the 2nd paragraph of section 3.1. This problem does not have any bearings on the result that lightning was predicted to be an order of magnitude higher prior to areal adjustment.

10. Going to even smaller sizes, the skill of the precipitation prediction is substantially decreased overestimating rainfall, but the w_max approach now underestimates lightning activity. How does this fit into the context of point #9?

We assume this comment is referring to the 4 km simulation. Since no convective parameterization is used in this simulation, it almost warrants an entirely separate study, but it is included for completeness as PR92 provided both $f(z_{top})$ and $f(w_{max})$. That said, the differences in precipitation-lightning flash rate relation from previous comparisons is something unexpected, and the conclusion drawn from this analysis is that additional caution is required when switching between $f(z_{top})$ and $f(w_{max})$ due to the uncharacterized behavior of formulation dependency. This is brought up in the third paragraph of section 4.2.

11. Using the Boccippio data, can you derive an estimate how an alternative fitting function for the IC:CG ratio would look like? This could be used as an alternative, if it matches the spectra better.

The Boccippio data is "better" because it provides an assumed climatological constant, and that a constant multiple of the total lightning spectra does not produce the spectral drop-off from using PR92+PR93. However, a fitting function cannot be derived based on that data alone. The inter-annual variability of flash is large, and the use of a single season of ENTLN total lightning will not be sufficient in producing an alternative fitting function. Another factor is the need for a parameter with sufficiently large spatial and temporal coverage, continuity in observation, and a well-defined counterpart within numerical models. While precipitation satisfies both requirements, again, the differences between 2006 and 2011 correlation (fig 6) between precipitation and lightning flash rate shows the insufficiency of using precipitation alone in deriving a new fitting function. The need for continuous monitoring and archiving of storm properties with lightning flash rate measurement is suggested in the conclusion for this very reason.

12. The last paragraph of the conclusion discussing LNOX would fit better as a motivation into the introduction and is not part of the conclusions, as none of the statements are discussed in the manuscript. Consequently, this does not belong into a conclusion section of a paper dealing with flash parameterization. This paragraph should be eliminated completely.

Removed parts of this paragraph related to LNOx. The remaining sentences are relevant in response to #11, which provides suggestions for the future direction of lightning monitoring and parameterization.

References

Grell, G. A.: Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations, Monthly Weather Review, 121, 764–787, 1993.

Mansell, E. R., MacGorman, D. R., Ziegler C. L., and Straka, J. M.: Simulated three-dimensional branched lightning in a numerical thunderstorm model, J. Geophys. Res., 107(D9), doi:10.1029/2000JD000244, 2002.

Wiens, K. C., Rutledge, S. A., and Tessendorf, S. A.: The 29 June 2000 Supercell Observed during STEPS. Part II: Lightning and Charge Structure, J. Atmos. Sci., 62, 4151--4177, doi:10.1175/JAS3615.1, 2005.

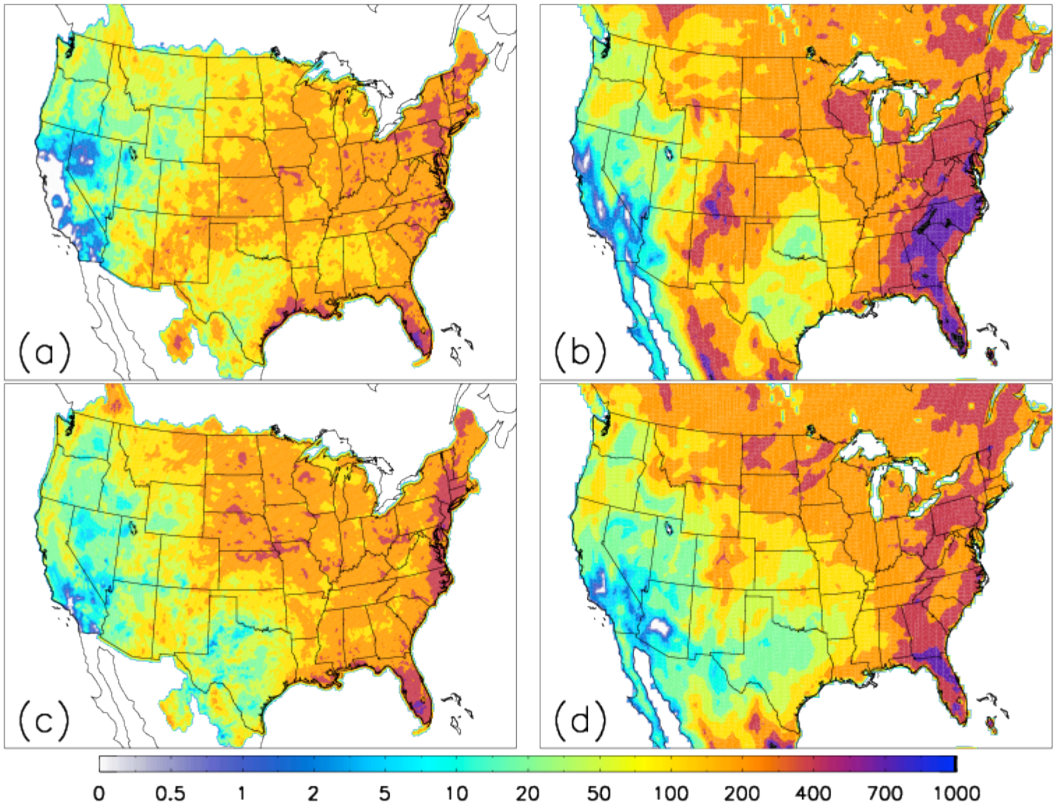

Figure R1. Modified color scale for Figure 2 of the manuscript.

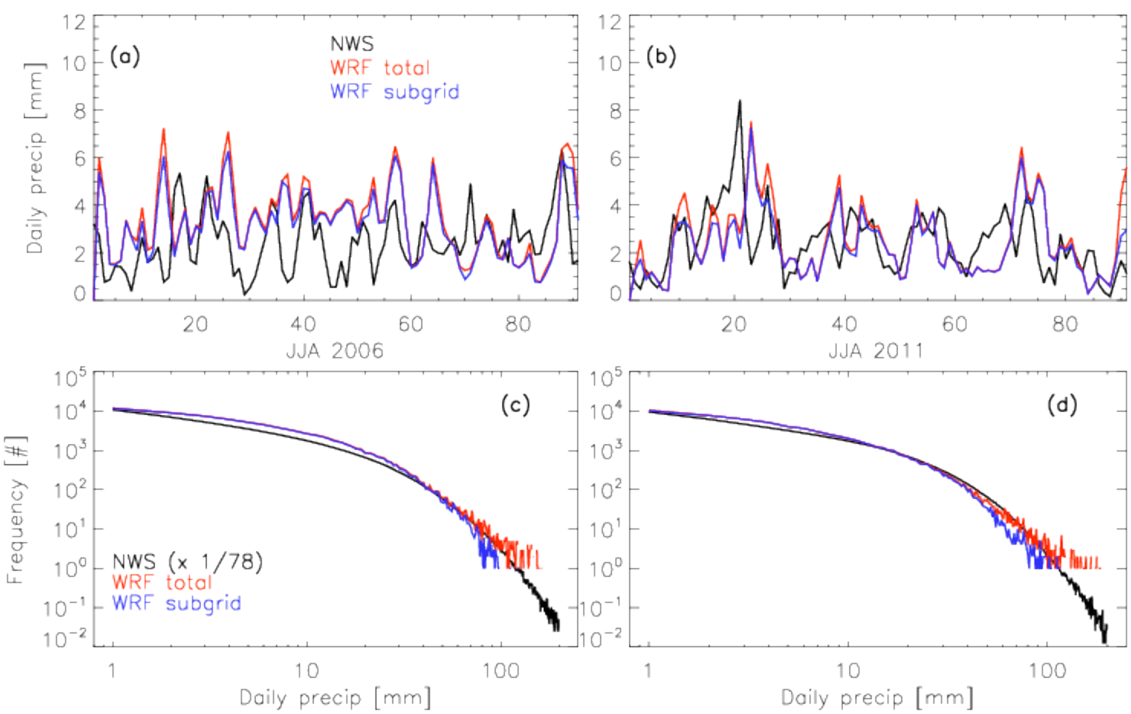

Figure R2. Modified line colors for FIgure 3 of the manuscript.

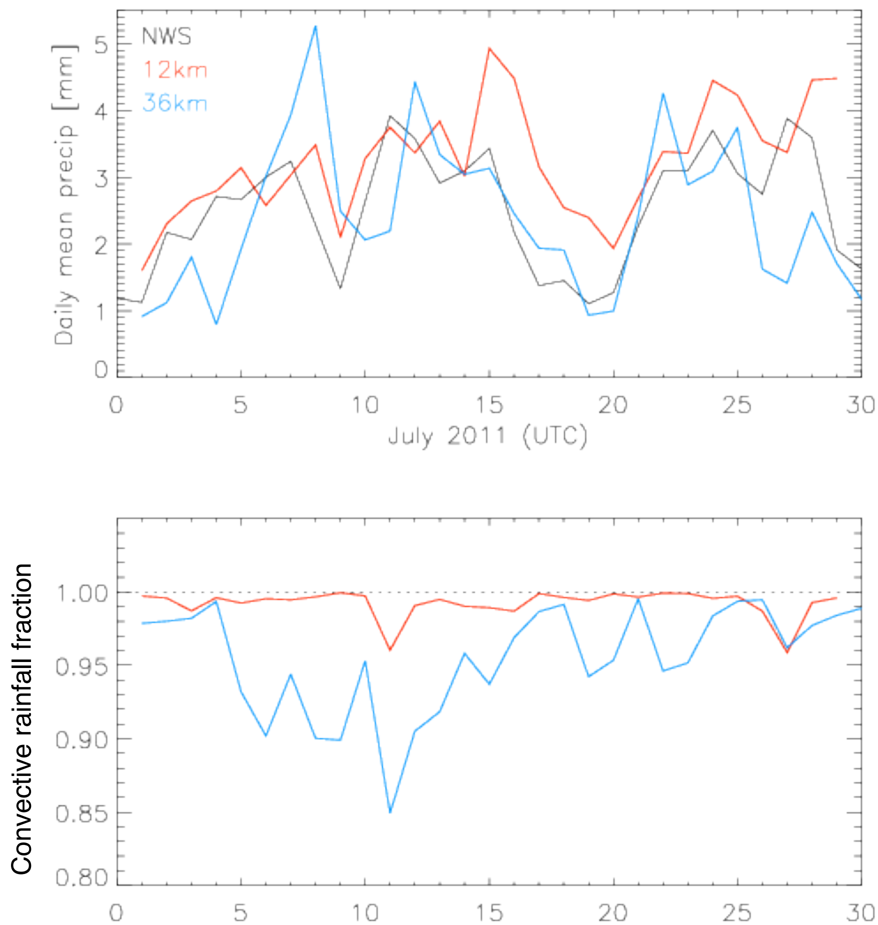

Figure R3. Daily mean precipitation and Convective rainfall fraction as a function of day in July 2011 from the NWS observations, the 12 km simulation and the 36 km simulation.