Reviewer #1:

This study examines the sensitivity of model simulation of precipitation to five tuning parameters in the KF convective parameterization in an effort to quantify the uncertainty of the model to convective parameterization. The authors use the WRF model simulation of precipitation during a convectively active period in the US Southern Great Plains. The parameters they chose to examine are CAPE relaxation time scale, updraft mass flux entrainment rate, downdraft intensity, downdraft starting height and maximum TKE in sub-cloud layer. A statistical analysis method was used to quantify the model performance in terms of average bias and pattern correlation coefficient and to identify an optimal parameter set for the model. The authors found that the model simulated precipitation is sensitive to CAPE relaxation timescale, entrainment rate and downdraft intensity, but less so to maximum TKE and downdraft starting level. They further investigate the issue of transferability of the identified parameter set to higher model resolution and other precipitation regimes, and found that with the optimal parameter set, both simulations at higher resolution in the SGP region and in the North American Monsoon region are improved relative to the model default setup.

This is a very useful paper for uncertainty quantification of a widely used regional atmospheric model. It demonstrates that with proper choices of model parameters, the performance of the model can be significantly improved, and such improvement is transferable to other model resolutions and climate regimes. The manuscript is well written and is publishable with minor revision.

Specific comments:

1. Sec. 3.3. The authors suggest that optimization in precipitation simulation also improves the simulation of other model fields such as 2-m mean temperature and 10-m wind speed. From Fig. 11, it is difficult to gauge how significant these improvements are in terms of physical quantities. Additional information, e.g. geographic maps similar to Fig. 5 (or observations, plus difference from observations) for temperature and wind speed, would be helpful. Also, a concise comparison, such as a Taylor diagram, comparing model simulations (using both default and optimized parameter sets) with observations would be more indicative of the improvement of model performance.

We have replaced Fig. 11 by geographic maps of bias difference (|Optimal-Observation|-|Default-Observation|) for temperatures and wind speed. Negative values mean positive impact or reduced absolute bias by using the optimized parameters. It can be seen that model biases are generally reduced for Tmean, Tmin and wind speed.

2. The results presented are for two-month averages. Diurnal cycle of precipitation is very difficult to simulate in this region. How well is it simulated in the WRF model, and is it improved using the optimal set of the parameters identified? In particular, increased entrainment in updrafts can act to delay the initiation of deep convection, thereby preventing convection from occurring too early in the model. Is this seen in the WRF simulation?

As this reviewer pointed out, the diurnal cycle of precipitation for this flooding month is poorly simulated by WRF in this region. Observation shows a distinct nocturnal maximum precipitation associated with the larger nighttime moisture transport by the Great Plain Low-Level Jet and/or eastward propagating mesoscale systems from the Rockies that reach the Great Plains during nighttime. Our simulations show that the increased entrainment in updraft can slightly delay the timing of precipitation peak. However, using the optimal set of the parameters didn’t improve the diurnal cycle simulation noticeably in this study. Further study is needed to assess the relative roles of nocturnal moisture transport by the Low-Level Jet versus eastward propagating disturbance in enhancing nighttime convection and precipitation in the Great Plains and which aspects the model failed to capture. The latter (eastward propagating systems) will present particular challenge for model tuning based on optimization of model parameters that influence mainly local convection.

3. p. 31783, lines 17-18. How is kernel density estimation performed? A brief description would be helpful to readers.

We have added the following sentence.

In statistics, kernel density estimation, a non-parametric way of estimating the probability density function of a random variable, is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample.

4. p. 31785, last para: discussions on Fig. 8. An important effect of increased downdraft is enhanced cooling and drying of the boundary layer, where cold and dry air from downdrafts is dumped. I suggest the authors to include 2-m temperature and PBL moisture in Fig. 8 and include some discussions of this effect. The sensible heat flux variation is probably partly due to this enhanced cooling of PBL air: colder 2-m temperature leads to more sensible heat flux from the surface. Also, the authors interpret the increase of lower troposphere (800-900 hPa) air humidity with increasing downdrafts as a result of increased rain evaporation (supposedly within downdrafts). I suspect the reduced adiabatic drying in the convection environment is probably more important. As downdraft mass flux increases, the net upward mass flux inside convection (up minus down) is reduced, therefore requiring less compensating subsidence in the convection environment. This subsequently leads to less subsidence-induced adiabatic drying. The authors could easily check on this by comparing the relative importance of the moisture source and sink terms.

This is a good suggestion and we have added the PBL temperature and moisture in Fig. 8 and included some discussions of this effect. The sensible heat flux variation is probably partly due to the enhanced cooling of PBL air. However, in the WRF, the 2-m air temperature is a diagnostic variable and the response of 2-m air temperature is very similar to that of skin temperature, so we added the mean air temperature for 1000-900 hPa (representing PBL) instead of 2m air temperature in the new Figure 8.

The grid-scale subsidence can induce strong adiabatic drying at the lower atmospheric layer. However, downdraft flux itself can also have the adiabatic drying effect, together with the evaporation of precipitation that cannot be neglected. We also found that the PBL moisture is less affected by downdraft intensity compared to its upper layer (900-800hPa), which is probably because the PBL are also affected by the surface evapotranspiration.

Reviewer #2:

Summary:
This study applies a novel approach to understand the sensitivity of regional climate simulations to input parameters of the model parameterizations schemes. Specifically, simulations with the Weather Research and Forecasting (WRF) model over the Southern Great Plains region are examined in their sensitivity to five key input parameters to the Kain-Fritsch convection scheme. One product of this sensitivity study is identification of an optimal combination of input parameters. This combination optimal for one region is then shown to also improve simulations in another climate regime and at another model grid spacing, thereby identifying a more robust set of input parameters than currently used in the default version.

The technique to efficiently sample input parameters is both novel, valid and will be of interest to the wider weather and climate modeling community as a tool that allows for a comprehensive exploration of uncertainty. This study provides new insights into the important topic of understanding of uncertainty and given the popularity of the WRF model for regional climate modeling, this paper provides an important contribution. The subject matter is appropriate for Atmospheric Chemistry and Physics and is worth being published.

The paper is well written. The abstract can be understood without reading the paper first, and summarizes the main results. The introduction includes a comprehensive review of previously published work that provides motivation for this study. The methodology is sound. The final section includes some discussion of the wider implications of the results, in particular for multi-scale modeling. I recommend this paper for publication after some comments (detailed below) have been addressed. These comments can by addressed by including brief discussions on likely outcomes, rather than performing additional experiments which are beyond the scope of the paper.

Specific Comments
1) Make it clear to the reader whether the goal of the study is to produce the ‘best’ simulation or to understand uncertainty. Perhaps the latter is the goal and the former is a product of the technique?

The primary goal of the study is to quantify the uncertainty and sensitivity of modeled precipitation to the input parameters and how they are transferred across processes, spatial scales, and climatic regimes. This study can also provide useful information to improve model performance by optimizing parameter values for a specific model configuration. We have clarified this in the Abstract and introduction.

2) As stated in the manuscript, it will be interesting to apply the MVFSA to other climate regimes. It is perhaps a little premature to generally conclude that MVFSA results are transferable across processes, given that the two climate regimes considered were both convection-dominated regimes. In addition, the default parameter set resulted in positive precipitation bias in both these regimes. It will be useful to include a brief discussion on the likelihood of the MVFSA result performing similarly well in a regime with a low precipitation bias when using the default parameter set.

This is a good suggestion and we have added a few sentences in Section 4.

The two regions (SGP and NAM) selected in this study are both convection-dominated climate regimes and precipitation are overestimated using the default model parameters in both regions. It is not clear whether optimization performed for one region is also transferable to another region if model biases with the default parameters are of opposite sign in the two regions. The issue of transferability of the benefits of optimization across different climate regimes and different spatial resolutions is being investigated further along with optimization of other physical parameterization schemes, which will be reported in a follow on paper.

3) Mention briefly the expected impact of observational uncertainty on your technique and results.

If uncertainties of the observations are also considered, the estimated posterior PDFs for the input parameters may show a broader or flatter shape; consequently, this may impact the model simulations including extreme events.

We add a brief discussion to section 4 as follows:

In addition, uncertainties of the observations are not considered in this study, which may impact the shape of the posterior PDF of the input parameters and the model outputs including extreme events (Jackson et al. 2003).

4) Five key input parameters to the Kain-Fritsch convection scheme are considered. Is there any objective guidance on the a priori choice of input parameters to examine? A related question is how to decide the range of values to examine. Figure 4 (top) suggests that there may be even greater improvement for values of Pd and Pe beyond the selected range.

The key parameters and their ranges were chosen through careful review of the papers describing the Kain-Fritsch scheme and personal communication with Dr. Jack Kain, one of the developers of parameterization. In addition, some parameters must satisfy physical constraints. As an example, Pd must vary within a range for the net flux at the cloud base (upward-downward) to be positive, and the range of Pe was chosen based on possible cloud radius.

We also noticed that greater improvement can be obtained for values of Pd and Pe beyond the selected range. However, such values are physically unrealistic and represent what is needed from the convective parameterization to compensate the errors introduced by other parameterizations schemes that also influence the diabatic heating.

5) An important result of the study is that the optimal combination of input parameters at one horizontal grid spacing also improves the simulation at another horizontal grid spacing. Given that aspects of the convection scheme are sensitive to the vertical profile, what is the likelihood that the optimal combination will also work well across different vertical resolutions?

This is a good question. While we are testing the optimization at different horizontal resolutions, it will be interesting to do more simulations to test the impact of different vertical resolutions. We will report the results in another separated paper.

6) It is clear that the optimal parameter set will depend on the variable used to assess skill. This study uses precipitation, which is inherently noisy and provides a hard test of the method. Please discuss the likelihood that using wind or geopotential height would result in more robust parameter sets.

We chose precipitation in this study because it is an important variable of the water cycle and it is strongly influenced by the convection scheme (most precipitation in this case is contributed by convective rain). We could add other variables such as wind and geopotential height in the skill metric, but these large scale fields are influenced by many other processes besides convection. More importantly, winds and geopotential height in a regional model are constrained by the lateral boundary conditions (particularly in our experimental design with frequent reinitialization), so they may not be very responsive to changes in parameters associated with the convection scheme. However, it would be interesting to consider these variables in the skill metrics for global simulations or regional simulations with a large domain.

7) An interesting result of the paper is that reduction of precipitation intensity biases also improves the spatial pattern of precipitation. This should be emphasized.

Yes we have emphasized this result. This may be partly because the region is not so big.

8) Figures 8 and 9 show the impact of the optimal parameter set on other variables but are the changes in the right direction i.e. closer to the observations?

These two figures are mainly used to show the sensitivity and response of other variables to the input parameters we selected, so we didn't compare each of those variables with the corresponding observation. For some variables, only limited observations are available from flux tower sites. We have, however, revised Fig. 11 to show geographic maps of bias difference (|Optimal-Observation|-|Default-Observation|) for temperatures and wind speed. Negative values mean positive impact or reduced absolute bias by using the optimized parameters. It can be seen that model biases are generally reduced for Tmean, Tmin and wind speed.

9) I recommend citing a similar study that looks at efficient sampling of input parameters. How does the MVFSA compare to the approach used in this study?
Hacker, J. P., S.-Y. Ha, C. Snyder, J. Berner, F. A. Eckel, M. Pocernich, J. Schramm and X. Wang, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. Tellus , 63 , 625-641. doi: 10.1111/j.1600870.2010.00497.

Good suggestion. Hacker et al. (2011) evaluated several techniques to account for mesoscale initial-condition (IC) and model uncertainty in a short-range ensemble prediction system based on the WRF model. They chose 10 parameter sets by applying the Latin Hypercube Sampling approach, which can sample throughout the parameter spaces for a limited number of parameters. The approach used in our study (i.e. MVFSA) is different from the Latin Hypercube Sampling. MVFSA can progressively move toward regions of the parameters space that minimize the model errors based on a skill score. Therefore, MVFSA features high converging efficiency for optimization study, while the Latin Hypercube Sampling is more suitable for sensitivity study.

We have added a brief discussion and cited the paper in Introduction.

Hacker at al. (2011) evaluated the impacts of initial-condition and model parameterization uncertainties on a WRF-based ensemble prediction system and found that different combination of parameterization schemes associated with the perturbed parameters could result in most skillful ensemble prediction.

10) In Eq. (4), model biases are assumed to be spatially and temporally uncorrelated. Is the likely violation of this assumption acceptable?

Following Jackson et al. (2003), the assumption that model biases are uncorrelated spatially and temporally simplify the optimization. This is also a typical assumption used in data assimilation. Assessing its validity or impacts is beyond the scope of this study.

11) Is the impact of going from 25 km to 12 km grid spacing (both using the default input parameters) a bigger improvement than going from 25 km with default parameters to 25 km using optimal parameters? It is beyond the scope of the current paper but it would be interesting to run MVFSA on both resolutions to look at the differences in parameter PDFs.

The 12-km simulation with default input parameters still significantly overestimates precipitation (see Figure 12) so our initial assessment is that larger improvements can be gained from using optimized vs default parameters at 25 km compared to using 25 km vs 12 km grid spacing with the default parameters. To quantify this assessment, we are investigating this issue using the MVFSA on the 12km resolution and comparing the results with 25 km grid spacing. The results will be reported in a separate paper.

Typing Errors:
1) Section 3.4, line 7. Change 'compute intensive' to 'computationally intensive'.

Done.

2) Page 31792, line 20. Change ‘has investigated’ to ‘has been investigated’.

Done.

Reviewer #3:

General Comments. This is a valuable contribution, because I am not aware of this method being applied in detail to a physics parameterization in a meteorological model before. The authors have chosen a good range of variables, and introduced a method of exploring the large parameter space efficiently to optimize a chosen aspect, in this case precipitation. The authors have shown good familiarity with the physics parameterization scheme chosen and have explained the mechanisms behind the parameters well. While I would have liked to have seen an independent period or year chosen to back up their findings, I feel this is a good initial presentation of the method that can lead to future related work.

Specific Comments. The caveat is that in a different year, with perhaps a drier or moister soil, or in a different region of precipitation, the results may have optimized towards different values, which would have been quite instructive. It would be dangerous to take these results at face value to apply to this scheme in all situations.

Yes we agree with this comment and have been cautious about the implications of our study. We agree that the optimal parameters obtained in this study may not yield the best result in other cases or other model configurations. We emphasize that our objectives are to investigate some issues of uncertainty quantification such as the transferability of optimal parameters and the impacts of high resolution data rather than to provide an optimal parameter set for the KF scheme in the WRF model.

Technical Comments:
1. Equation 8. Is RND the same in both uses in this equation or are they different random numbers?

RND used in Eq. (8) are different random numbers (we didn’t use different names to separate them in this equation).

2. p31780. I assume that 50 experiments means that K=50, but this is not quite clear.

The K represents the time series in Eq. (11) (i.e. 30 days). To be clearer, we change it to N.

3. p31780, line 26. The word may be "constraint".

Done.

4. p31782. The method of doing the overlapping simulations was presented at the end of Section 2, after the description of some physics tests in Figure 3. Were these physics tests carried out with the same simulation technique? And if so, it might be beneficial to put this description before these tests. 5. p31784 and earlier description of EC on p.31780. Since E and C appear to have such different magnitudes, it is not clear that EC=E-C is sufficiently normalized to make sense.

Yes, those physics tests were carried out with the same simulation technique (i.e. overlapping simulations).
We have added a few words in the last paragraph of Section 2.3 to clarify this.
We normalized E and C first and then added them together. Therefore EC can be considered as a metric equally weighting E and C.

Reviewer #4:

In general, this manuscript is very well written and organized. I have following specific comments:

Content

- Is that possible to describe the physical meanings of the parameters used for adjustment?

We have described those parameters in details in Section 2.1 and summarized in Table 1.

- They find wet biases and determine that most precipitation is convective; thus, assumes that the error within WRF is due to cumulus parameterization. Through the adjustments of those parameters in the KF scheme, the simulated precipitation is significantly improved. Can similar adjustments be used in the other cumulus schemes?

The methodology demonstrated in this study can be applied to other cumulus schemes with different input parameters and ranges.

- Why were the particular ranges of values for the 5 parameters selected (section 2.1, table 1)?

The key parameters and their ranges were chosen through careful review of the papers describing the Kain-Fritsch scheme and personal communication with Dr. Jack Kain, one of the developers of parameterization. In addition, some parameters must satisfy physical constraints. As an example, Pd must vary within a range for the net flux at the cloud base (upward-downward) to be positive, and the range of Pe was chosen based on possible cloud radius.

- No mention of going from 25km to 12km in the methods section.

We have added a sentence in the methods section.

- Not clearly stated that UW data used for MVFSA technique (bottom of page 31780).

Done.

- Not clear as to whether output is 25km or 12km (Figures and Tables).

All simulations for the optimization used a 25 km grid spacing over SGP. Two additional sets of simulation were conducted with a different resolution (i.e. 12km) and region (i.e. NAM), respectively. We have added the related information in the Captions of Figures and Tables.

Grammar

- Line 13 of page 31774 should be revised. Not clear. Should read something like “magnitude and intensity of precipitation.”

Done.

- Paragraph that starts on line 8 on page 31774 contains very long run-on sentence. Consider breaking apart.

Done.

- End of aforementioned paragraph contains sentence fragments.

Done.

- I would remove the first sentence on line 6 of page 31775 (not absolutely necessary).

We believe keeping this sentence reads better.

- Paragraph at the end of page 31776/beginning of page 31777 could be clarified.

Done