Review of GMD-2023-28 titled "Optimized Stochastic Representation of Soil States Model Uncertainty of WRF (v4.2) in the Ensemble Data Assimilation System" by S. Lim, S. K. Park and C. Cassardo

This study used a micro-genetic algorithm to perturb soil moisture and temperature in Noah-LSM within the WRF model (V4.2) to improve ensemble spread that can lead to the short-term prediction in the planetary boundary layer (PBL). Tuning parameters such as the amplitude and the horizontal decorrelation length/time scale of random forcing applied to soil states are evaluated based on 6-h forecasts in temperature and moisture in the boundary layer. While authors claimed that they improved ensemble forecasts by tuning parameters for random perturbations of soil states, this study only worked on an initial ensemble spread, which is different from the actual ensemble spread that grows with cycles. Also, it is not clear how the random perturbations added to soil temperature and moisture can represent the uncertainty of Noah LSM since the actual model uncertainty or error is never examined. The concept of ensemble cycling presented in this draft is either incorrect or vague to make the experiment design and figures not supportive of authors' main points. Numerous fundamental or structural issues are found throughout the manuscript. Unfortunately, authors might need to advance their understanding of the ensemble data assimilation system, as specified in my major comments below. I would recommend authors to spend enough time revisiting the issues, performing ensemble cycling experiments to examine the effect of perturbing land states on the weather prediction, with a clear target time in mind (as six hours is not the characteristic time scale where soil states are expected to significantly affect atmospheric forecasts). For that, I would recommend "Rejection" of the manuscript for now.

→ We appreciate the valuable and constructive comments, which helped us improve the quality of the manuscript. We carried out three major tasks in this study: i) developing the stochastic perturbations to soil states (e.g., soil temperature and soil moisture) scheme (SPSS); ii) optimizing the random forcing (RF) tuning parameters of SPSS; and iii) applying SPSS to an ensemble data assimilation (EDA) system. As a result, our newly developed SPSS inflates the ensemble spread during the forecast time, not the initial ensemble spread in the EDA system. If our major ideas were not well described, we revised the manuscript, and we respond to your valuable comments below.

Major comments:

1. Lack of innovation: Various stochastic perturbation techniques have been already introduced in the WRF model and widely used to increase ensemble forecast skills (See the references in https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/v4.2/users_guide_chap5.html#stochastic). Authors should recognize all those related efforts specific to the WRF model (including the version 4.2 used in this study) and should justify why we need another perturbation algorithm despite all the corresponding options having been fully supported in the WRF system for a decade, already. For example, users can easily create a 3-D Gaussian random perturbation by simply turning on a namelist parameter (e.g., rand_perturb=1), and the capability of stochasticallyperturbing parameters also exists for RUC LSM, which can be easily expanded or applicable to Noah LSM. Because these are basically doing the same thing as what authors try to do here, with almost the same tuning parameters, it is mandatory to clarify the need of another algorithm for thesame system.

   Response: Thank you for your practical and specific comments. As you introduced, the WRF model contains various stochastic perturbations such as stochastically perturbed parameterization tendency scheme (SPPT), stochastic kinetic energy backscatter scheme (SKEB), and stochastically perturbed parameterizations scheme (SPP), as well as stochastic perturbations to the boundary conditions. Ollinaho et al. (2017) suggested that future representations of probabilistic model error should address coupled processes such as the surface and ocean. In particular, land surface is commonly underestimated in ensemble systems (Lavaysse et al. 2013; Leutbecher et al. 2017; Gehne et al. 2019). Our WRF-GSI/EnKF system also suffered from an underestimated ensemble spread in soil states and near-surface atmospheric variables, so we introduced stochastic perturbation on soil states (e.g., soil moisture and soil temperature). Since soil moisture and soil temperature can directly affect near-surface temperature and humidity forecasts through the heat flux response (Kim and Hong, 2007; Lin and Pu, 2020), previous studies attempted to perturb land surface problems using soil states physics tendencies (Gehne et al., 2019; Draper, 2021), initial soil states (Sutton et al., 2006; Gehne et al., 2019), and direct soil states (Draper, 2021). Although SPP, already implemented in RUC LSM of WRF, also perturbs the land surface parameters (i.e., constant values) to improve the boundary layers, it unfortunately does not perturb variables (i.e., model states). To investigate whether the soil states perturbation can inflate the atmospheric variables in the planetary boundary layer in the WRF-GSI/EnKF system, we developed the stochastic perturbations to soil states scheme (SPSS) in WRF. Note that the RF tuning parameters to determine the perturbation features are still incompletely understood (Lupo et al., 2019), and land and atmosphere have different behaviors in dynamics and error growth (Draper, 2021). Therefore, we propose an optimization strategy to find the optimal RF tuning parameters of soil states to figure out the necessity of special scales of RF tuning parameters for land surface perturbations. Finally, we identified whether SPSS with potentially optimal RF tuning parameters can help inflate the ensemble BEC in the EDA system. To clarify motivation of this work, we revised the paragraph in L49-58 as below:

"These stochastic representations of model uncertainty can address a coupled process (e.g., atmosphere-land surface) where a lack of spread exists in the near-surface variables (Leutbecher et al., 2017). In particular, the land surface model (LSM) interacts with the lower atmosphere as boundary conditions. Because it is strongly coupled to the atmospheric state at certain times and in certain places, the reduced land-surface uncertainty may lead to better atmospheric forecasts (MacLeod et al., 2016). In particular, the soil states directly affect the near-surface temperature and humidity forecasts through the sensible and latent heat flux responses (Kim and Hong, 2007; Deng et al., 2016; Lin and Pu, 2020; Sutton et al., 2006; Wang et al., 2010b). If the surface is heated during the daytime, sensible energy transfers to the atmosphere and moisture evaporates from the soil; thus, exact soil states are important within the planetary boundary layer (PBL), influencing convection and precipitation (Sutton et al., 2006). Previous studies perturbed land surface problems using soil states physics tendencies, initial soil state, direct soil states, and surface parameters to examine the impact on atmospheric ensembles (Sutton et al., 2006; MacLeod et al., 2016; Orth et al., 2016; Gehne et al., 2019; Draper, 2021); however, direct soil state perturbation during the forecasts is not yet implemented in the Weather Research and Forecast (WRF) model."

Reference:

Sutton, C., Hamill, T. M., and Warner, T. T.: Will perturbing soil moisture improve warm-season ensemble forecasts? A proof of concept. Mon. Weather Rev., 134, 3174–3189, https://doi.org/10.1175/MWR3248.1, 2006.

Wang, Y., Kann, A., Bellus, M., Pailleux, J. and Wittmann, C.: A Strategy for perturbing surface initial conditions in LAMEPS. Atmos. Sci. Lett., 11, 108–113, https://doi.org/10.1002/asl.260, 2010b.

Gehne, M., Hamill, T. M., Bates, G. T., Pegion, P., and Kolczynski, W.: Land surface parameter and state perturbations in the global ensemble forecast system. Mon. Weather Rev., 147(4), 1319-1340, 2019.

2. Inappropriate references: Along the line, the micro-genetic algorithm should be introduced and understood as an alternative to the existing options available in the WRF model, but in the Introduction section, authors did not include any of the previous work specific to the WRF implementation. It is not convincing if the algorithm introduced in this study or the study *per se*could make any meaningful contributions to improving our understanding or ensemble forecastskills. One should not ignore others' decade-long efforts on the same system for the same problem (e.g., inflating ensemble spread).

   Response: Thank you for your constructive comments. To supplement the absence of an introduction to the micro-genetic algorithm (μ-GA), we added the following paragraph after L58:

   "Parameterizations contain uncertain parameters that can lead to sensitive results; hence, optimal parameter estimation is important to enhance the accuracy of the NWP model. As one of the optimization algorithms, the genetic algorithm (GA) is a global optimization based on the Darwinian principles of natural selection (Holland, 1975; Goldberg, 1989). Standard GA and micro-GA, which is efficient GA with a small population, have been successfully used for parameter optimization of a cumulus parameterization scheme in the fifth-generation Pennsylvania State University (PSU)/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) (Lee et al., 2006), a convective parameterization scheme in WRF (Yu et al., 2013), and snow-relaxation parameters in the offline Noah Land Surface Model (Noah LSM) (Lim et al., 2022)."

3. Inappropriate title: Authors called it optimized representation, but I would expect a much more generic approach to optimize ensemble configurations in the context of a coupled system, not based on a single case study. To me, the presented study rather seems to be one of many ad-hoctuning practices.

   Response: Thank you for your comments. We agree that generalizing optimized tuning parameters is risky because we only used a single case for daytime and nighttime. As a result, we revised the title as below:

   "Stochastic Perturbation of Soil Sates Model Uncertainty of WRF (v.2) in the Ensemble Data Assimilation System: Preliminary Design for Optimization of Random Forcing Tuning Parameters"

4. Clarity issues: The manuscript needs an extensive work on clarifications. It took me a while to figure out that this study dealt with initial ensemble spread, not the spread during cycling because authors mixed up the two different things throughout the manuscript and from the very beginning. The first statement in the Introduction, for instance, is incorrect: "The ensemble data assimilation (EDA) describes both initial conditions (ICs) and model uncertainties represented by the flow- dependent background error covariance (BEC)." => In fact, EDA only "requires" initial ensembles to start cycling, and the initial ensembles are not described by EDA because they can be generated separately as this study shows. Also, EDA does not need to describe model uncertainties since there are different ways to construct ensembles without taking model uncertainties into account (e.g., perturbing observations). On the other hand, the general description of EnKF in GSI (e.g., Eqs. (9)-(15)) may not be necessary unless authors made any changes for this application.

Response: Thank you for your comments. As step by step, we answered your comments.

i) This study demonstrated the introduction of SPSS to account for land surface model uncertainty in ensemble forecasts from the WRF-GSI/EnKF system. SPSS inflated the ensemble spread of soil temperature or soil moisture in ensemble forecasts during the DA cycling, not the initial ensemble spread. If we carefully explain our experimental design again, first, optimization experiment finds an optimal RF tuning parameters of the SPSS using a single case on each daytime and nighttime. Second, we turned on the SPSS scheme with optimized tuning parameters in EDA system. As the other stochastic perturbation scheme, SPSS perturbed soil temperature or soil moisture during the forecast time at every time step. In other words, ensembles used to calculate the ensemble BEC further perturb soil temperature or soil moisture to increase near-surface uncertainty. Schematic diagram about EDA system is described as below:

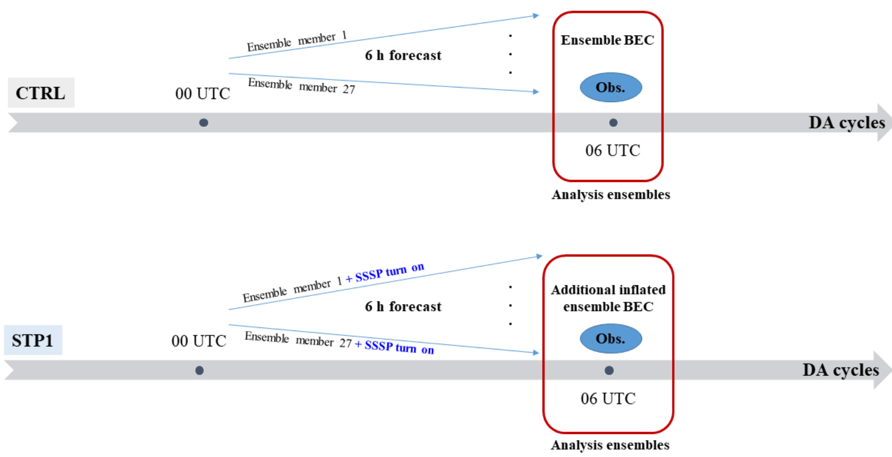


Figure S1. Schematic diagram of SPSS application in the WRF-GSI/EnKF system with examples of CTRL and STP1 experiments.

ii) At the first cycle, the initial ensemble members are generated by the lagged forecast, but the initial conditions (i.e., analysis) in every cycle assimilated by available observation and previous 6-hour forecast information; thus we revised the first sentence as below:

"Ensemble data assimilation (EDA) estimates the flow-dependent background error covariance (BEC) starting from a set of initial conditions (ICs) with available observations (Hamill and Whitaker, 2005)."

Reference:

Hamill, T. M., and J. S. Whitaker: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. Mon. Weather. Rev., 133, 3132–3147, https://doi.org/10.1175/MWR3020.1, 2005.

iii) Representing model uncertainty in the EDA may affect the estimated ensemble BEC based on ensemble forecasts (Leutbecher et al., 2017).

iv) As you suggested, we removed the Eqs. (9)-(15) and revised L181-206 as below:

"Ensemble Kalman filter (EnKF; Evensen, 1994; Whitaker and Hamill, 2002; Houtekamer and Zhang, 2016) uses an ensemble of forecasts to estimate the BEC in the Kalman filter. Based on the Monte Carlo approach, it produces a set of random samples for the analysis and background state probability distributions (Buehner, 2005). We used EnKF (v1.3) provided in the Gridpoint Statistical Interpolation (GSI) community (v3.7) composed of two parts, GSI observer and EnKF (Liu et al., 2018): the GSI observer computes the observation innovations (i.e., observation − background) using the observation operator, and EnKF generates the analysis of each ensemble member. The GSI/EnKF provides two algorithms (i.e., a serial ensemble square root filter (EnSRF) (Whitaker and Hamill, 2002) and a local ensemble Kalman filter (LETKF) (Hunt et al., 2007)) to calculate the analysis increment. The current implemented algorithm is EnSRF, which avoids sampling errors by perturbing observations (Whitaker and Hamill, 2002)."

5. Improper goal setting with poor experiment design and methodology: From my view, this is oneof the most fundamental problems this work has.
   a. The motivation of this study seems to tackle the insufficient ensemble spread that can lead to poor forecast skills. When it comes to under-dispersive ensemble systems, however, the actual problem lies on the reduction of ensemble spread with cycles, not the spread in the initial ensemble. As it takes at least dozens of cycles to saturate the ensemble spread in the regional cycling system, the initial ensemble construction certainly matters to the efficiency (e.g., how quickly the spread grows), but that is a fundamentally different problem from the filter divergence issue where observations are gradually rejected due to the lack of spread.

      Response: Thanks for your comments. We agree that the under-dispersive ensemble spread that causes filter divergence should be addressed by the DA period rather than the initial ensemble. Therefore, we implemented SPSS to add continuously evolving random patterns to soil temperature and soil moisture, allowing the ensemble spread to inflate continuously across consecutive assimilation windows. Some unclear descriptions of the experimental design contributed to misunderstanding our main work, so we modified "3.2. experimental design" as follows.

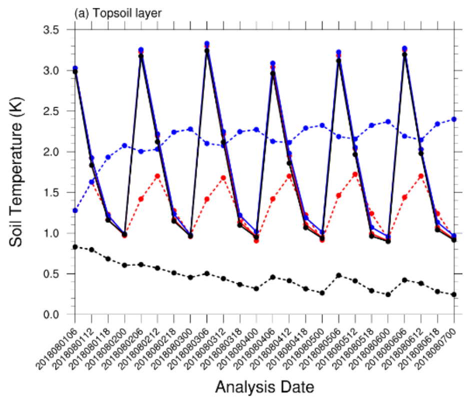
      "In this study, we carried out two main experiments in this study: i) optimizing the RF tuning parameters of SPSS; and ii) applying SPSS to ensemble forecasts can inflate ensemble BECs during DA cycles. First, we conducted the following optimization experiments in a coupled system of μ-GA and SPSS: (1) Optimization of the RF tuning parameters for Soil Temperature Perturbation at Daytime (OSTP-D); and (2) the corresponding one at Nighttime (OSTP-N); (3) Optimization of the RF tuning parameters for Soil Moisture Perturbation at Daytime (OSMP-D); and (4) the corresponding one at Nighttime (OSMP-N). Experiments were conducted in August, when soil-atmospheric coupling is strongest (Draper, 2021). We ran the 6 hour forecast for the daytime starting at 00 UTC (09 KST) 1 August 2018 and the nighttime starting at 12 UTC (21 KST) 1 August 2018. As for the optimization configuration in μ-GA, we followed recommended settings (Carroll, 1996; Yu et al., 2013; Yoon et al., 2021), i.e., 5 population size, uniform crossover, and 100 generations. A coupled system of μ-GA and SPSS found a potential solution of RF tuning parameters within the assigned ranges

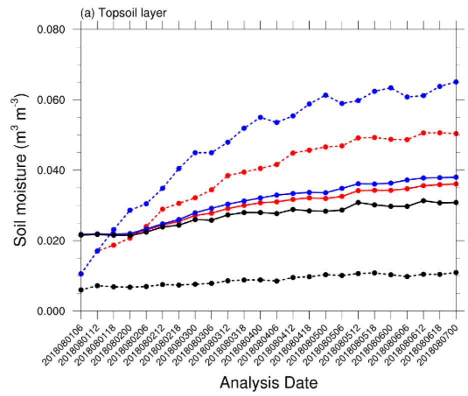
(Table 2) by randomly choosing the candidate value among 64, 64, and 16 cases for amplitude, decorrelation length, and time scale, respectively. The ensemble ICs (i.e., five ensemble members describing the ensemble system) were produced by the random control variables (CV) method, implemented in the WRF Data Assimilation system (WRFDA). It generated the ensemble ICs by adding the random noise to analysis in the control variable space (Gao et al., 2018); thus, the general perturbation patterns followed the background error. We used the basic option, CV option 3, composed of the following control variables: stream function (Φ), unbalanced velocity potential (χ_u), unbalanced temperature (T_u), pseudo relative humidity (q), and unbalanced surface pressure ($P_{s,u}$).

Second, we used SPSS with optimized RF tuning parameters in DA cycles to add continually evolving random patterns to soil temperature and soil moisture, allowing the ensemble spread to inflate across consecutive assimilation periods. To prepare the WRF-GSI/EnKF system, we used 27 ensemble members, which were known to be the best ensemble size in terms of accuracy and computational costs (Kunii and Miyoshi, 2012), generated by the random CV option 3. The control variables were u-component wind, v-component wind, surface pressure, virtual temperature, and specific humidity. To prevent filter divergence, we used the multiplicative inflation method with a 0.9 inflation parameter to inflate the analysis ensemble spread back to the background and the covariance localization with a horizontal length scale of 500 km and a vertical length scale of 0.4 scale height, based on distance from the observation. We investigated whether SPSS for soil temperature and soil moisture can alter the ensemble BECs for temperature and water vapor mixing ratio in PBL and the effectiveness of diurnally-varying RF tuning parameters in the DA cycling experiments as follows: (1) Soil Temperature Perturbation 1 (STP1) perturbs soil temperature using the daytime tuning parameters obtained from OSTP-D, and (2) STP2 perturbs soil temperature using the diurnally-varying tuning parameters obtained from OSTP-D and OSTP-N; (3) Soil Moisture Perturbation 1 (SMP1) perturbs soil moisture using the daytime tuning parameters obtained from OSMP-D, and (4) SMP2 perturbs soil moisture using the diurnally-varying tuning parameters obtained from OSMP-D and OSMP-N; (5) These were compared to the control experiment (CTRL), representing the current WRF-GSI/EnKF system. All experiments were cycled from 06 UTC 1 August 2018 to 00 UTC 7 August 2018, and the spin-up period was the first 3 days of the total period."

b. Moreover, this study focuses on the uncertainties of soil states, it is thus critical to have the land states that are well spun up before the initial time. Considering the imperfect land surface model with no land data assimilation, the characteristic time and spatial scale of soil states, and the initialization from the NCEP-FNL at coarse resolution (1°x1°), a three-day spin-up used in this study is not even close to the very minimum requirement (say, a month).

   Response: Thank you for your comments. As you mentioned, land surface models require sufficient initialization, such as a month. Due to enormous computational time, this preliminary study investigated the impacts of soil temperature and soil moisture perturbations using SPSS for 1 week. In the time series of ensemble spread and ensemble errors for soil temperature and soil moisture (Figures 7-8(a)), the rapidly increased or decreased ensemble errors (solid line) and ensemble spreads (dashed line) were relatively saturated after 3 days, such as from 00 UTC on August 4. Accordingly, we assumed that the first 3 days were the spin-up period. In the future study, we will extend the experimental period by at least 1 month based of this a proof of concept study.





(Left) **Figure 7(a).** Time series of ensemble mean error (solid line) and ensemble spread (dotted line) in CTRL (black), STP1 (blue), and STP2 (red) for background during experimental periods: (a) soil temperature (K) at the topsoil layer over the land.

(Right) **Figure 8(a).** Time series of ensemble mean error (solid line) and ensemble spread (dotted line) in CTRL (black), SMP1 (blue), and SMP2 (red) for background during experimental periods: soil moisture (m^3 m^{-3}) at the topsoil layer over the land.

c. The experiment design is not well described, but if only 27 ensemble members were used, covariance localization must have been used for such a small ensemble size. In that case, the localization would certainly affect ensemble spread and additive perturbations, but are not found anywhere in this script.

Response: Thank you for pointing this out. Since we didn't mention the configuration of covariance localization, we included it in L241-242:

"To prevent filter divergence, we used the multiplicative inflation method with a 0.9 inflation parameter to inflate the analysis ensemble spread back to the background and the covariance localization with a horizontal length scale of 500 km and a vertical length scale of 0.4 scale height, based on distance from the observation."

d. The construction of an initial ensemble needs clarification: Was the random CV option inWRFDA used to perturb atmospheric variables for a 5-member ensemble (L232-235) while the micro-genetic algorithm was used only for soil perturbations in a 27-member ensemble (L239)? How were the two different ensembles combined in your experiment, then? The choice of ensemble size is critical to ensemble spread, but the description of the ensemble system is unclear.

Response: Thank you for pointing this out. The random CV option is used to generate initial ensemble members. The 5 ensembles are used to optimize the RF tuning parameters, while the 27 ensembles are used to describe the EDA system. To clarify the ensemble system, we revised L238-240 as below:

"To prepare the WRF-GSI/EnKF system, we used 27 ensemble members, which were known to be the best ensemble size in terms of accuracy and computational costs (Kunii and Miyoshi, 2012), generated by the random CV option 3."

e. Table 1: How did you decide the optimized ranges for soil moisture and temperature?

That range used in this study seems to be ad-hoc, not necessarily representing either model or observation uncertainty.

Response: Thank you for your comments. We defined an optimization range by widening from the typical three scales of RF tuning parameters for decorrelation length scale and standard deviation (e.g., de-correlation length scales of 500, 1000, and 2000 km; and standard deviations of 0.52, 0.18, and 0.06, respectively) in order to increase the probability that the micro-GA will find the optimal solution. We refined the time scale to smaller values because it is rather sensitive to SPSS. The detailed descriptions are included in L161-162, as below:

"First, μ-GA randomly initializes RF tuning parameters from the assigned ranges. We assumed potential tuning parameter ranges for the length scale and standard deviation based on three general scales of tuning parameters (Leutbecher et al., 2017). As for the time scale, however, it was redefined for the SPSS because typical ranges (e.g., 6 hours, 3 days, and 30 days) caused excessive perturbations."

f. Authors defined a fitness function in Eq. (8) to determine the best parameter values among several candidates. Although it is true that meteorological variables in the boundary layer are closely tied to land states, atmospheric fields are characterized at different time and spatial scale from that of soil states. Considering the response time of atmospheric variables to soil perturbations as well as many other potential contributors to the boundary layer structure (such as advection, convection, radiation, and clouds), it is questionable if the fitness function based on 6-h forecasts can fully capture the actual impact of soil perturbations on boundary layer forecasts or can be used as a proxy to the optimal parameter settings.

Response: Thank you for your valuable comments. First of all, I would like to introduce why we defined our fitness function (Eq. (8)). For example, when soil temperature changes, ground heat flux (G_0) is affected by Equation (6). To satisfy the surface energy balance (Eq. (3)), the changed G_0 can do repartitioning to sensible (H) and latent heat fluxes. As a result, the changed heat fluxes affect the atmospheric temperature. If we see Eq. (4), at least the potential temperature at the surface (T_{sfc}) and the atmospheric temperature at the lowest model level (T_{air}) can be adjusted. Finally, the perturbed soil temperature was propagated to the planetary boundary layers even in the 6 hour forecast. Since our interest was in how perturbed soil states can change the ensemble BEC composed of 6 hour forecasts in DA cycles, we only evaluated 6 hour forecasts in our fitness function. In a future study, we will evaluate with a longer lead time to consider the response time of atmospheric variables to soil perturbations as well as many other potential contributors to the boundary layer structure since the user can define a fitness function depending on the objective of optimization; this may improve ensemble forecasts over a longer period.

This is indeed one of the most complex issues to disentangle clearly, but it is my concern thatthe approach used in this study seems overly simple to resolve such a challenging issue. Anyhow, given that the WRF system already provides various perturbation techniques, another perturbation algorithm cannot guarantee a noble work, and the only way I can see tomake this type of work meaningful is to examine how the initial ensemble affects the ensemble forecast skills for a long period of time in a statistical manner (e.g., not in a single case).

Response: Thank you for your comments. This study is noteworthy in that it provides a proof-of-concept for a RF tuning parameter optimization applied to the soil state perturbation. Although we only optimized a single case for daytime and nighttime, the RF tuning parameters for soil temperature and soil moisture suggested different values reflecting each physical characteristic. Furthermore, although the RF tuning parameters were ad-hoc, the SPSS was effective in improving ensemble spread in soil states and atmospheric variables in PBL. Based on this study, in a future study, we will include more cases and elaborate a fitness function during optimization to suggest general RF tuning parameters for soil state perturbations.

6. Figures not supportive of main points:
   - Figure 3: If this study is all about inflating spread, it is expected to show ensemble spread, nota single member of soil states.

     Response: Thank you for your comments. Figure 3 and its caption were updated from a single ensemble member to an ensemble spread as below:

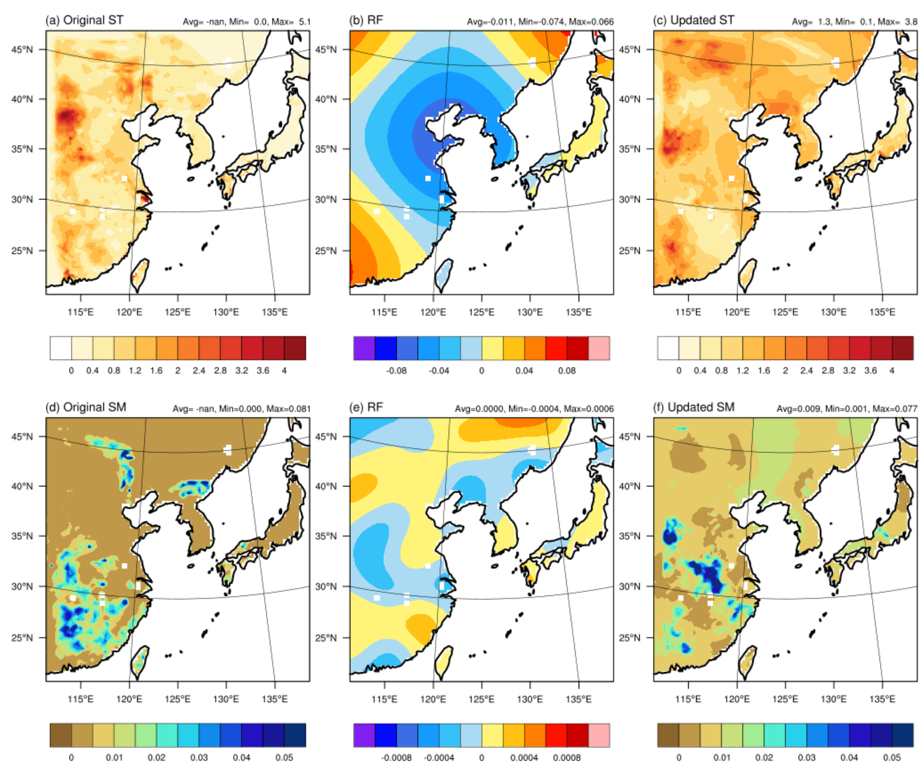


Figure 3. Ensemble spread of soil temperature (ST in K; upper panels) and soil moisture (SM in m^3 m^{-3}; lower panels) and ensemble mean of RF at 06 UTC on 1 August 2021: (a) original ST, (b) RF applied to ST (with σ = 0.13 K, L = 2900 km, and τ = 120 s), and (c) updated ST (i.e., original + RF); (d) original SM, (b) RF applied to SM (with σ = 0.0003 m^3 m^{-3}, L = 250 km, and τ = 900 s), and (f) updated SM.

   - Note that an initial ensemble is only a start of cycling and is not supposed to represent the saturated ensemble spread. Hence, Fig. 6 is not needed.

     Response: Thank you for your comments; however, this is not an initial ensemble spread inflation study, as we responded to comment #4. Figure 6 is a composite of the 25 cycles' background from 06 UTC on 1 August 2018 to 06 UTC on 7 August 2018. Therefore, Figure 6 is significant to distinguish the under-(or over-)estimated ensemble spread (right panels) compared to ensemble errors (left panels) in the CTRL experiment. As a result, we can determine which inflation methods are required to inflate where an underestimated ensemble spread was reported. To clarify Figure 6, we revised the figure's caption as below:

"**Figure 6.** The zonal mean ensemble mean error (left panels) and ensemble spread (right panels) for temperature (K; top panels) and water vapor mixing ratio (g kg^{-1}; bottom panels) as for the 6 hour forecasts of CTRL over the land. Results are averaged from 06 UTC on 1 August 2018 to 06 UTC on 7 August 2018 with a composite of the 25 cycles' background fields (i.e., 1 cycle per 6 hour)."

- Figures 7, 8, and 10 only show the sensitivity to the initial ensemble, not the optimal ensemble spread.
  Response: Thank you for your comments; however, this is also not an initial ensemble spread inflation study, as we responded to comment #4. Figures 7, 8, and 10 show the sensitivity to the SPSS using the optimized RF tuning parameters in ensemble forecasts in every DA cycles. In other words, we optimized the RF tuning parameters, not the ensemble spread. Although the optimized RF tuning parameters are ad-hoc, they were effective in SPSS to perturb soil temperature or soil moisture. In DA cycles, SPSS perturbs soil temperature or soil moisture at every time step for every ensemble member during the 6-hour forecast (i.e., background).

- Figure 9: Again, we do not expect the saturation of ensemble spread at the initial time.
  Response: Thank you for your comments, however, we inflated the ensemble spread in ensemble forecasts in every DA cycles as responded to comment #4.

- Figure 11: Analysis increments at the initial time are not indicative of the system performance, and the GFS analysis is not quite trustworthy near the surface.
  Response: These are the composites of analysis increments during the DA cycles. Figure 11 shows that SPSS helps to modify the analysis increments to reduce the background errors by inflating the ensemble BECs during the DA cycles. Because the GFS analysis uses the same number of soil layers and depth as Noah LSM, it is able to evaluate on the same grid-point verification without interpolation uncertainties. In future work, we can evaluate our performance with other reanalysis data or in-situ observations.

7. Section 2.1 is named WRF-Noah LSM Coupled System. As far as I know, Noah-LSM is just oneof the physics options available in WRF. Did you develop/change anything to enhance the coupling part either in the model or in your analysis step? As all the physics parameterization schemes are interacting with each other within the WRF framework, it is not clear why authors emphasized the "coupled" system here. How does your Noah LSM work differently from all other studies using the same option, again from the modeling or DA aspect?

   Response: Thank you for pointing this out. We included our stochastic perturbations scheme (i.e., SPSS) in the Noah LSM code to produce the perturbed soil temperature or soil moisture, but we did not change the other physical parts in Noah LSM. Although we intended to emphasize the interactions between atmospheric and land surfaces, we revised the title of subsection 2.1 and 2.2 (L71 and L112) to prevent misleading the coupled system, as follows:

   "2.1. WRF Configurations"
   "2.2. Stochastic Perturbations to Soil States scheme (SPSS)"