Accounting for Land Model Uncertainty in Numerical Weather Prediction Ensemble Systems: Toward Ensemble-Based Coupled Land–Atmosphere Data Assimilation

CLARA SOPHIE DRAPER[a]

[a] *NOAA/Earth Science Research Laboratories/Physical Sciences Laboratory, Boulder, Colorado*

(Manuscript received 29 January 2021, in final form 19 May 2021)

ABSTRACT: The ensembles used in the NOAA National Centers for Environmental Prediction (NCEP) global data assimilation and numerical weather prediction (NWP) system are underdispersed at and near the land surface, preventing their use in ensemble-based land data assimilation. Comparison to offline (land-only) data assimilation ensemble systems suggests that while the relevant atmospheric fields are underdispersed in NCEP's system, this alone cannot explain the underdispersed land component, and an additional scheme is required to explicitly account for land model uncertainty. This study then investigates several schemes for perturbing the soil (moisture and temperature) states in NCEP's system, qualitatively comparing the induced ensemble spread to independent estimates of the forecast error standard deviation in soil moisture, soil temperature, 2-m temperature, and 2-m humidity. Directly adding perturbations to the soil states, as is commonly done in offline systems, generated unrealistic spatial patterns in the soil moisture ensemble spread. Application of a stochastically perturbed physics tendencies scheme to the soil states is inherently limited in the amount of soil moisture spread that it can induce. Perturbing the land model parameters, in this case vegetation fraction, generated a realistic distribution in the ensemble spread, while also inducing perturbations in the land (soil states) and atmosphere (2-m states) that are consistent with errors in the land–atmosphere fluxes. The parameter perturbation method is then recommended for NCEP's ensemble system, and it is currently being refined within the development of an ensemble-based coupled land–atmosphere data assimilation for NCEP's NWP system.

KEYWORDS: Coupled models; Data assimilation; Land surface model; Ensembles; Numerical weather prediction/forecasting

1. Introduction

The atmospheric data assimilation used in the NOAA National Centers for Environmental Prediction (NCEP) numerical weather prediction (NWP) system uses forecast uncertainty estimates that are based, in part, on an ensemble of model forecasts. NCEP's NWP ensemble is underdispersed (or overly confident of the forecast skill) at and near the land surface. This lack of ensemble spread in the land surface is common to the ensemble systems at international NWP centers (Lavaysse et al. 2013; Leutbecher et al. 2017; Gehne et al. 2019). This is not unexpected, since none of the schemes used to induce spread in these ensemble systems explicitly account for land surface model uncertainty. The present study investigates different mechanisms for accounting for error in the land model component of NCEP's NWP model when generating the NWP ensemble used in the global data assimilation system. Improved representation of forecast uncertainty for the land will improve ensemble-based forecast products of boundary layer processes (e.g., probabilistic precipitation forecasts), as well as providing more realistic forecast error estimates for use in ensemble-based data assimilation of land-sensitive variables. In particular, this study is motivated by the development of an ensemble Kalman filter (EnKF) to constrain the soil moisture and temperature states in NCEP's NWP system by assimilating 2-m atmospheric temperature (T2m) and specific humidity (Q2m) observations.

We cannot simply extend the methods used to introduce uncertainty in the atmospheric component of an NWP model to the land component, since the dynamics of the land and atmosphere are fundamentally different in ways that affect the behavior and growth of errors in each system. The atmosphere is chaotic and exhibits sensitive dependence on initial conditions, so that an initial condition error will grow over time. This error growth occurs more rapidly along instabilities in the atmospheric flow, providing information on flow-dependent uncertainties. In contrast, the land surface is strongly forced, in that the long-term state of an offline (land-only) model forecast will be entirely determined by its external atmospheric forcing. That is, over a long enough forecast length, forecasts from the same land model forced with the same atmosphere will eventually converge to the same state, regardless of their initial conditions (Cosgrove et al. 2003). Land surface flows at NWP scales are also predominantly vertical, and the land surface models used in NWP systems simulate the land independently at each grid cell, with no horizontal communication between adjacent land model grid cells. Consequently, and in contrast to an atmospheric model, a land model will not in general grow errors through a forecast, nor will it propagate errors horizontally. Instead, the spatial and temporal distribution of the spread in an ensemble of land model forecasts will much more directly reflect the perturbations (or errors) introduced into that ensemble by the ensemble generation scheme.

Several authors have investigated the impact on ensembles of atmospheric forecasts of perturbing land parameters and states. For short-range forecasts, Lavaysse et al. (2013), Bouttier et al. (2016), and Gehne et al. (2019) all achieved a

Corresponding author: Clara Sophie Draper, clara.draper@noaa.gov

DOI: 10.1175/JHM-D-21-0016.1

For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

modest increase in spread in the boundary layer over land by perturbing the land initial conditions (usually soil moisture and temperature), and/or a selection of the land model parameters used in each ensemble member, while Lavaysse et al. (2013) and Bouttier et al. (2016) also achieved improved ensemble forecast skill. These studies all evaluated forecasts of less than 5 days, hence they retained some impact of the perturbed initial land states. At longer (seasonal-plus) time scales, MacLeod et al. (2016), Orth et al. (2016), and Strommen et al. (2019) achieved improved ensemble spread in the boundary layer and improved forecast ensemble means and/or forecast skill by perturbing selected land parameters. MacLeod et al. (2016) also tested perturbing the soil states using a stochastically perturbed physics tendencies (SPPT) scheme, but found this to be less successful than perturbing the model parameters. There are also several established systems that generate and maintain land surface model forecast ensembles for use in offline ensemble-based land data assimilation. These systems typically perturb the model states at regular time intervals (Kumar et al. 2014; De Lannoy and Reichle 2016), and/or apply different atmospheric forcing to each ensemble member (Kumar et al. 2014; De Lannoy and Reichle 2016; Fox et al. 2018). In contrast to the atmospheric ensemble forecast studies above, offline land surface model ensemble systems usually do not perturb the model parameters, with the exception of data assimilation systems that are being used to update model parameters (Kumar et al. 2012; Pinnington et al. 2021). Another difference between offline land data assimilation systems and atmospheric forecast ensembles is that the atmospheric studies cited above all apply perturbations just once at the start of the ensemble of forecasts, while the land data assimilation systems (which must cycle continuously through data assimilation updates and ensemble forecasts) usually apply temporally varying perturbations at regular time intervals.

This study investigates the introduction of schemes to explicitly account for land model uncertainty in the ensemble forecasts from NCEP's NWP model, the Global Forecast System (GFS), that are used in the global NWP data assimilation system. First, the earlier assertion that the ensembles used in NCEP's global data assimilation system are underdispersive at and near the land surface is confirmed. In section 3a, the ensemble spread in archived operational GFS output is evaluated by comparison to target forecast uncertainty estimates, derived by evaluating time series of GFS output against independent data. Note that "forecast uncertainty" is used here to refer to the random error in the model forecasts, as measured by the forecast error standard deviation, while "ensemble spread" refers to the ensemble standard deviation, itself used to infer the forecast uncertainty within an EnKF. The spread in the key atmospheric variables that force the land surface in the operational GFS ensemble is also examined, by comparison to the spread typically applied in offline land data assimilation systems. In section 3b, three methods of introducing land model error into ensemble forecasts are investigated, drawing from the methods applied in both offline land-only and atmospheric ensemble systems, as appropriate. Each method is tested in a simplified version of NCEP's NWP system over a boreal summer, when soil–atmosphere coupling is strongest. The ensemble spread induced by each method is compared against the target forecast uncertainty estimates established earlier. This study is focused on understanding how different ensemble generation methods affect the forecast uncertainty estimates obtained from the resulting ensembles, and in particular the land–atmosphere error covariance terms that are central to coupled land–atmosphere data assimilation. While there is strong evidence that improved representation of land model forecast uncertainty can improve the atmospheric ensemble means and forecast skill (Lavaysse et al. 2013; Bouttier et al. 2016; MacLeod et al. 2016; Orth et al. 2016; Strommen et al. 2019), this is left for later work.

2. Background and methodology

a. *The GFS/GDAS NWP modeling and data assimilation system*

NCEP's NWP modeling and data assimilation system is the GFS/Global Data Assimilation System (GDAS), which has been running version 15 (GFSv15) since June 2019. At the time of writing, GFSv15 was running operationally on a cube sphere grid, at approximately 12-km resolution, with 64 vertical layers. The model consists of the FV3 dynamical core (Putman and Lin 2007), and the GFS physics package. The physics includes the Noah land surface model (Ek et al. 2003), which models the soil moisture and temperature in four layers. The GDAS atmospheric data assimilation is the GSI Hybrid 4DEnVar (Kleist et al. 2009; Kleist and Ide 2015), which is run every 6 h, and makes use of a 70-member ensemble of GFS forecasts, at a reduced resolution of 25 km. The land data assimilation is a once daily update to the snow depth. The soil moisture in layers 2–4 is also nudged toward climatology every 6 h, using a 60-day relaxation time scale.

b. *Target forecast uncertainty estimates*

This section describes the methods used to generate independent estimates of forecast uncertainty in the operational GFS system over boreal summer, for variables relevant to the EnKF assimilation of 2-m observations. These independent estimates will be used as the target uncertainty estimates, against which the ensemble spread from the operational GFS archives and each perturbation experiment will be measured. The target estimates are each obtained by comparing time series of archived GFS operational output to independent observations of each variable. For both soil moisture and soil temperature, only the first layer (0–10 cm) will be investigated, since it is the most well observed, and is also the most strongly coupled to the atmosphere. Summary details of the methodology and datasets used for each target estimate are included in Table 1.

1) NEAR-SURFACE SOIL MOISTURE

It is extremely difficult to evaluate large-scale soil moisture estimates (Draper et al. 2013; Gruber et al. 2020) since there is no agreed upon ground truth against which to perform an evaluation (Reichle et al. 2004). However, triple collocation

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

TABLE 1. Details of target estimates of forecast uncertainty for each variable.

Variable	Evaluation method	Datasets/references	Time period
Soil moisture	Triple collocation	MERRA-2 and SMAP	1 Jun–31 Aug 2016–19
Soil temperature	Literature	Holmes et al. (2012)	—
T2m, Q2m	Standard deviation of differences	ERA5	15 Jun–31 Aug 2019

can be used to estimate the error standard deviation of three independent datasets, after accounting for differences in their mean and variance, by assuming that all three datasets are linearly related, and have independent and orthogonal errors (Scipal et al. 2008). Triple collocation relies on a using multi-year time series with stationary errors, yet the operational GFS NWP system underwent a major upgrade to version GFSv15 in June 2019. While GFSv15 NWP output was used for all other target estimates, for soil moisture the GEFSv12 Reanalysis (Hamill et al. 2021, manuscript submitted to *Mon. Wea. Rev.*) was used to obtain the required time series from the latest version of the GFS model. GEFSv12 uses a version of the GFSv15 model that is very similar to that use in the GFSv15 NWP system, with no major differences in the land model. The GFS soil moisture forecast error standard deviation (or forecast uncertainty) was estimated during the boreal warm season by applying triple collocation to global gridded estimates of surface layer soil moisture from (i) the Soil Moisture Active Passive (SMAP) level 3 enhanced soil moisture retrieval product (O'Neill et al. 2020), (ii) the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Global Modeling and Assimilation Office 2015), and (iii) and the GEFSv12 model. The GFS soil moisture was used as the reference dataset (to which the other two datasets were scaled) and the triple collocation estimates were applied to data from June–August for 2016–19 (368 days in total). The triple collocation was applied as described in Draper et al. (2013), with the exception that soil moisture anomalies were not used, since the use of only warm-season data effectively removed the seasonal cycle.

2) NEAR-SURFACE SOIL TEMPERATURE

As with soil moisture, there is no large-scale agreed upon ground truth against which soil temperature can be evaluated. Its evaluation is also made difficult by the need to account for differences in the sampling depth when comparing datasets (Holmes et al. 2008). To the author's knowledge, triple collocation has not yet been applied to soil temperature, and was not attempted here as it would entail considerable development effort. Instead, approximate values of the expected forecast uncertainty for the GFS layer 1 soil temperature have been taken from the literature. Specifically, Holmes et al. (2012) evaluated the surface layer soil temperature in various NWP models, including NCEP's GFS and the ECMWF model. By comparison to ground-based observations from 101 Oklahoma Mesonet sites, they estimated the mean error standard deviation in the model soil temperature to be around 1.5 K during the day and 2.0 K at night. These values have been used here, and they are supported by the similar, although slightly higher, values of 2.0 K (day) and 2.5 K (night) estimated by Albergel et al. (2015) for the ECMWF model. Both studies tested their error estimates for dependence on factors such as vegetation and soil characteristics, soil moisture, and orography, and found no obvious relationships.

3) 2-M TEMPERATURE AND SPECIFIC HUMIDITY

There are no global gridded subdaily observed temperature and relative humidity datasets. Hence, for T2m and Q2m, the target GFS forecast uncertainty estimates have been constructed by comparison to ERA5 (Hersbach et al. 2020) output at daily analysis times (0000, 0600, 1200, and 1800 UTC). At the analysis times, the 2-m fields from ERA5 are the output of an optimal interpolation analysis that merges station observations with the model background forecasts. While the ERA5 2-m analysis is not purely observation-based and includes model information, particularly in poorly observed regions, it is a well-understood and high-quality product and provides a reasonable and convenient benchmark against which to evaluate the GFS 2-m fields.

The target forecast uncertainty for the GFS 2-m variables has been estimated by calculating the standard deviation of the difference between archived operational GFS forecasts and the ERA5 analyzed fields from 15 June (after the operational upgrade to v15) to 31 August 2019. Note that, since errors in the ERA5 2-m analyses are not accounted for, this method will overestimate the GFS forecast uncertainty. Finally, since forecast uncertainty in the boundary layer is expected to vary through the diurnal cycle, to more easily understand and present the results, the target forecast uncertainty estimates have been converted from UTC to local time. Specifically, the original 6-hourly snapshots at the ERA5 analysis times have been converted to 6-h time windows centered at 0000, 0600, 1200, and 1800 local time, by calculating the local time for each UTC input time at each longitude, and then assigning the corresponding model fields to the appropriate 6-hourly local time interval.

c. *Ensemble perturbation experiments*

In this section several experiments used to test different land model uncertainty schemes in the GFS ensemble system are described. A summary of these experiments is provided in Table 2. The three land perturbation experiments were compared to a control experiment, referred to as "control," with no land perturbations applied, which is intended to be representative of the current operational GFS NWP ensemble system. All of the experiments used a simplified version of the October 2019 version of NCEP's operational NWP system. In each experiment, the atmospheric ensemble perturbation scheme, and the assimilated datasets were the same as used operationally, although the atmospheric assimilation was performed with a computationally cheaper Hybrid 3D-EnVar. A reduced

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

TABLE 2. Details of soil moisture (SM) and soil temperature (ST) perturbation ensemble experiments.

Experiment Name	Land perturbation method	Perturbation parameters
Control	None	—
State-pert	SM and ST perturbed	ST: $r_t \sim N(0, 0.1)$ K h^{-1} [Eq. (1)] SM: $s_t \sim N(0, 0.1)$ mm h^{-1} [Eq. (2)]
Soil-SPPT	SM and ST physics tendencies perturbed	$s_t \sim N(0, 1.0)$ [Eq. (4)]
Param-pert	Vegetation fraction perturbed	$s_t \sim N(0, 0.1)$ [Eq. (2)]

resolution of C192 (roughly 0.5°) was used for both the deterministic and ensemble forecasts, and the ensemble size was reduced to 30 members. The soil moisture nudging was also retained, since a test of this nudging showed it had very little impact on the ensemble soil moisture spread. Additionally, after each assimilation update the operational GFS ensemble is recentered around the deterministic forecast. This recentering was not applied in the land perturbation experiments, resulting in a systematic difference between the mean ensemble soil moisture and the deterministic soil moisture in the control (and all other) experiments. This was not investigated further, since the effect of the land perturbation schemes on the ensemble mean is outside the scope of the current work.

The initial conditions were regridded from NCEP's archived operational output on 1 July 2019, and then run forward to 14 July to allow any inconsistencies to equilibrate. The ensembles produced by each land model uncertainty scheme were then compared after 1 month, on 14 August 2019. Initial tests showed that the 1 month spinup period is long enough that the initial rapid growth in spread has slowed down, but it is not quite sufficient for the ensemble spread to be completely spun up (particularly for layer 4 soil moisture). Nonetheless 1 month was selected as a compromise between computational cost and obtaining a reasonable representation of the ensemble characteristics generated by each scheme.

1) SOIL MOISTURE AND TEMPERATURE STATE PERTURBATIONS

In the first land perturbation experiment, referred to as "state-pert," randomly generated Gaussian perturbations were directly added to the model soil moisture and soil temperature at each model time step. This method is commonly applied to soil moisture in offline land data assimilation and ensemble systems (Kumar et al. 2014; De Lannoy and Reichle 2016); however, here both the soil moisture and soil temperature have been perturbed. For the state $x(i)_t$ in soil layer i, the perturbed state $x(i)'_t$ is calculated by adding the perturbation $r(i)_t$ scaled by the soil layer number:

$$x(i)'_t = x(i)_t + r(i)_t 0.5^{i-1}. \tag{1}$$

For soil temperature, the $r(i)_t$ perturbations are identical in each layer, and are generated according to a zero-mean Gaussian distribution. For soil moisture, which is bounded, the standard approach is to also apply zero-mean Gaussian perturbations and then to prevent the perturbed soil moisture from exceeding its bounds, by simply correcting any values outside the physical soil moisture range back to the boundary value (Kumar et al. 2014; De Lannoy and Reichle 2016). Consequently, as the soil moisture nears its upper or lower bounds, the applied perturbations become biased, since they are effectively limited in one direction (Ryu et al. 2009). Here, for the soil moisture perturbations the standard approach has been amended by applying a flat-top filter, which (symmetrically) reduces the magnitude of the applied perturbation close to the upper and lower boundaries ($x_{\min}$ and $x_{\max}$, respectively). For a random s_t, generated according to a Gaussian distribution, the $r(i)_t$ for each layer in Eq. (1) are calculated:

$$r(i)_t = s_t(1 - |z(i)_t^p|), \tag{2}$$

$$z(i)_t = -1 + 2\frac{x(i)_t - x(i)_{\min}}{x(i)_{\max} - x(i)_{\min}},$$

where p is the flat-top parameter, set to 5 here.

In this study, for soil temperature the r_t in Eq. (1) are generated using a standard deviation of 0.1 K h^{-1}, while for soil moisture the s_t in Eq. (2) are generated using a standard deviation of 0.1 mm h^{-1}. Both are given a horizontal decorrelation length of 1000 km, and temporal decorrelation length of 24 h. There is likely a relationship between errors in soil moisture and errors in soil temperatures; however, given that this relationship is not well understood, and likely not consistent in time, independent random fields have been used to generate the soil moisture and soil temperature perturbations, rather than imposing a correlation between the two.

2) SPPT OF SOIL STATES

A significant disadvantage of the state-pert approach is that it requires an assumed error correlation between the applied soil moisture and soil temperature perturbations (or assumed independence, as was done here). Yet, as noted above, the relationship between soil moisture and temperature (and presumably their errors) is not well understood, and is likely variable. The second land perturbation experiment, referred to as "soil-SPPT," addresses this by extending the GFS's atmospheric SPPT scheme (Zhou et al. 2021, manuscript submitted to *Wea. Forecasting*) to also apply to the soil states, so that the model itself is used to determine the relative perturbations applied to the soil moisture and temperature. SPPT applies randomly generated perturbations r_t to the model state (X_t) tendencies at the end of each model time step, by setting the perturbed model state (X'_t) to

$$X'_t = X_{t-1} + (1 + r_t)(X_t - X_{t-1}). \tag{3}$$

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

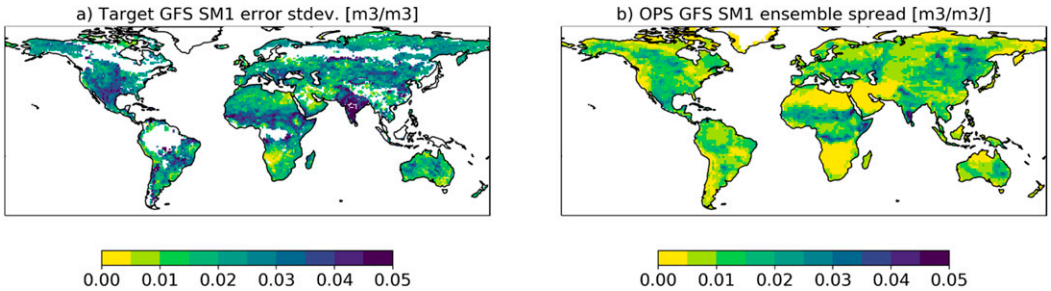

FIG. 1. Forecast uncertainty for the GFS layer 1 soil moisture (SM1) over boreal summer, estimated from (a) the target error standard deviation (triple collocation) and (b) the GFS NWP operational ensemble spread at 0000 UTC 15 Jul 2019. In (a) data gaps indicate unavailability of SMAP, usually due to dense vegetation.

In these experiments, X_t contains the soil moisture and soil temperature in each model layer, and the r_t are randomly generated perturbations. Initial experiments using a Gaussian perturbation for r_t often resulted in model instabilities when $(1 + r_t)$ in Eq. (3) was negative. Following the approach used in the GFS atmospheric SPPT scheme, this sign reversal was avoided by limiting r_t in Eq. (3) to be between -1 and 1, by setting it to

$$r_t = \frac{2}{1 + e^{s_t}} - 1, \tag{4}$$

where s_t is a randomly generated zero-mean Gaussian perturbation. Here, s_t was given a standard deviation of 1.0, and the same horizontal and temporal scales as for the state-pert experiment. The resulting r_t distribution has a standard deviation of 0.4. Finally, the land SPPT perturbations were generated independently from the atmospheric SPPT perturbation. While using the same perturbation fields for both would in theory produce consistent land and atmosphere perturbations, the atmospheric SPPT scheme is tapered toward the surface to avoid generating boundary layer instabilities, reducing the potential value of this consistency.

3) PERTURBING MODEL PARAMETERS

The third land perturbation experiment, referred to as "param-pert," is also designed to apply perturbations that are consistent across model variables, in this case by perturbing the model parameters used in the parameterization of the fluxes between the soil states and the overlying atmosphere. To demonstrate this approach the vegetation fraction was perturbed, since in experiments perturbing several GFS land model parameters Gehne et al. (2019) found vegetation fraction to have the greatest impact on short-range forecasts. The param-pert approach has the advantage over soil-SPPT of creating consistent perturbations across more than just the soil states, and in particular it has the potential to apply consistent perturbations to both the land and atmospheric components of each ensemble member. Since vegetation fraction is bounded between 0 and 1, the perturbations are applied using the flat-top filter, analogous to Eqs. (1) and (2), with s_t in the latter equation again assumed to be zero-mean Gaussian, with a standard deviation 0.1, and the same horizontal and temporal scales as the state-pert experiment.

3. Results

a. Operational GFS ensemble spread

1) LAND SURFACE STATE VARIABLES

Figures comparing the target estimates of forecast uncertainty from Table 1 to ensemble estimates from NCEP's archived operational GFS NWP output are shown in Figs. 1–4. The GFS ensembles are the 6-h forecasts that the GDAS uses to estimate forecast uncertainty. Global maps are plotted for all variables, except soil temperature for which boxplots are shown, since the target estimates are approximate and do not include any spatial information. A preliminary investigation of the archived GFS output showed little subseasonal variation in the large-scale characteristics of the ensemble spread that are investigated here, hence the various ensemble-based forecast uncertainty estimates are calculated from ensemble output on a single day. The forecast uncertainty maps are also summarized in Figs. 5 and 6, by plotting the forecast uncertainty estimates binned by the GFS Soil Wetness Index (SWI), where

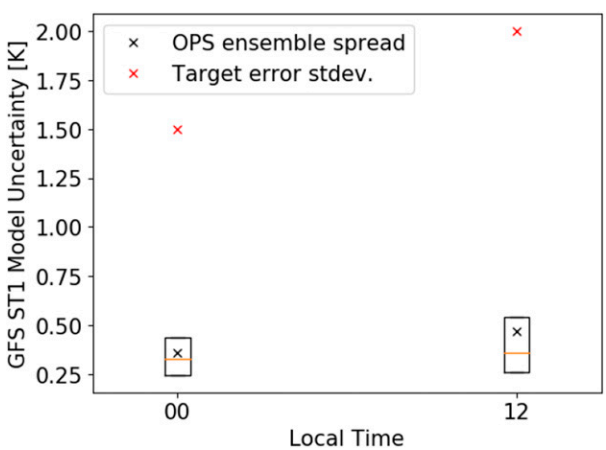

FIG. 2. Forecast uncertainty for the GFS layer 1 soil temperature (ST1) for 6-h time windows centered at 0000 and 1200 local time, estimated from the target error standard deviation (Holmes et al. 2012) and the GFS NWP operational ensemble spread on 15 Jul 2019. Crosses indicate the spatial mean, and boxplots for the ensemble estimate show the median and interquartile range.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

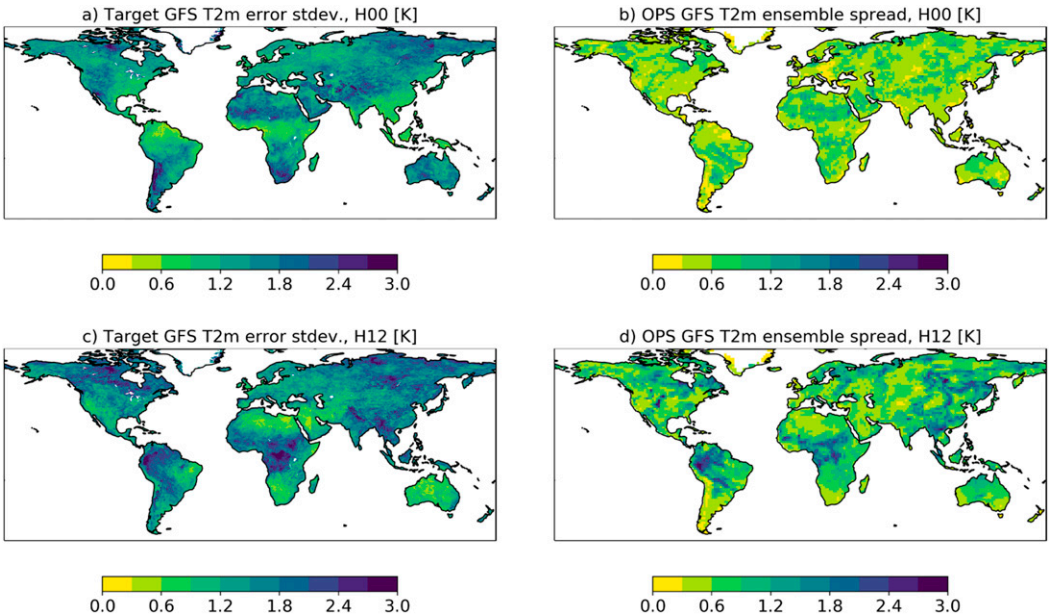

FIG. 3. Forecast uncertainty for GFS T2m over boreal summer for 6 h time windows centered at (top) 0000 and (bottom) 1200 local time estimated by the (left) target error standard deviation (from comparison to ERA5) and (right) GFS NWP operational ensemble spread on 15 Jul 2019.

the latter is defined as the soil moisture scaled between 0 (driest possible) and 1 (wettest possible) at each grid cell. To prepare Figs. 5 and 6 from global maps of forecast uncertainty, each grid cell is assigned to the appropriate SWI bin, and the root-mean-square of the forecast uncertainty is then plotted for each SWI bin, together with box plots, showing the interquartile range within each bin. The interquartile range is generally fairly narrow (compared to the variation of the mean with variation in SWI), confirming the appropriateness of plotting these fields against SWI. The exception is for the 2-m variables during the day, when the interquartile range is quite large for the operational GFS (Figs. 6c,d), and the mean results here should be interpreted with some caution. Note that these plots are not intended to imply a direct dependence between SWI and the forecast uncertainty. Instead they are introduced as a convenient method to summarize the maps in Figs. 1–4. The SWI in these plots is best interpreted as a proxy for the surface conditions, with dry values indicating dry, and warm and generally low vegetation conditions, and wet values generally indicating wet and cool and generally high vegetation conditions.

Starting with soil moisture, Figs. 1 and 5 demonstrate that the GFS ensemble underestimates the forecast soil moisture error, giving a mean uncertainty from the ensemble of $0.011\,\mathrm{m^3\,m^{-3}}$, compared to $0.025\,\mathrm{m^3\,m^{-3}}$ for the triple collocation estimate. However, in terms of regions of relatively large and small forecast uncertainty, the spatial patterns in the ensemble uncertainty appear to be reasonable, and Fig. 1 shows similar regions with larger uncertainty (tropics, central United States, northeast Asia, and regions of Australia) in the ensemble and target estimates. Likewise, in Fig. 5 both show the same pattern of relatively larger uncertainty in the center of the SWI distribution, and reduced values toward the wet and dry boundaries (more so toward the dry end). For soil temperature, Fig. 2 again shows that the ensemble is underdispersed, with a global mean forecast uncertainty of 0.36 K for the 6-h window centered on 0000 local time, and 0.47 K for the window centered on 1200 local time, compared to 1.5 and 2.0 K for the literature estimates. In both cases the interquartile range for the GFS mean is narrow, with upper limits still very far below the target literature values, indicating little spatial spread in the soil temperature ensemble standard deviation.

For T2m and Q2m, the GFS ensemble is again underdispersed, as shown in Figs. 3 and 4. For T2m, the GFS mean ensemble spread over land is 0.58 K at 0000 and 0.92 K at 1200 local time, much less than the target values of 1.43 and 1.58 K, respectively. Likewise, for Q2m, the GFS ensemble mean spread is $0.51\,\mathrm{g\,kg^{-1}}$ at 0000 local time and $0.59\,\mathrm{g\,kg^{-1}}$ at 1200 local time, compared to 1.00 and $1.10\,\mathrm{g\,kg^{-1}}$ for the target estimates. Despite the low bias, the spatial patterns in the ensemble spread agree qualitatively with those of the target estimates. For T2m, in Figs. 3a and 3b and 6a the night time forecast uncertainty is highest in dryer, warmer and less vegetated regions (Saharan and southern Africa, Australia, central Asia, the western United States and Chile) for both the target and Unified Forecast System (UFS) estimates. In Figs. 3c and 3d and 6c the daytime T2m forecast uncertainties for both estimates are highest in wetter, more vegetated regions (the tropics, south Asia, high latitudes). Similar patterns are observed for Q2m during the day (Figs. 4c,d and 6c). During the night, the Q2m forecast uncertainties (Figs. 4a,b) differ from those for T2m (Figs. 3a,b), with the target estimate showing a region of relatively large uncertainty loosely aligned with the intertropical convergence zone, which is less evident in the GFS ensemble estimate.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

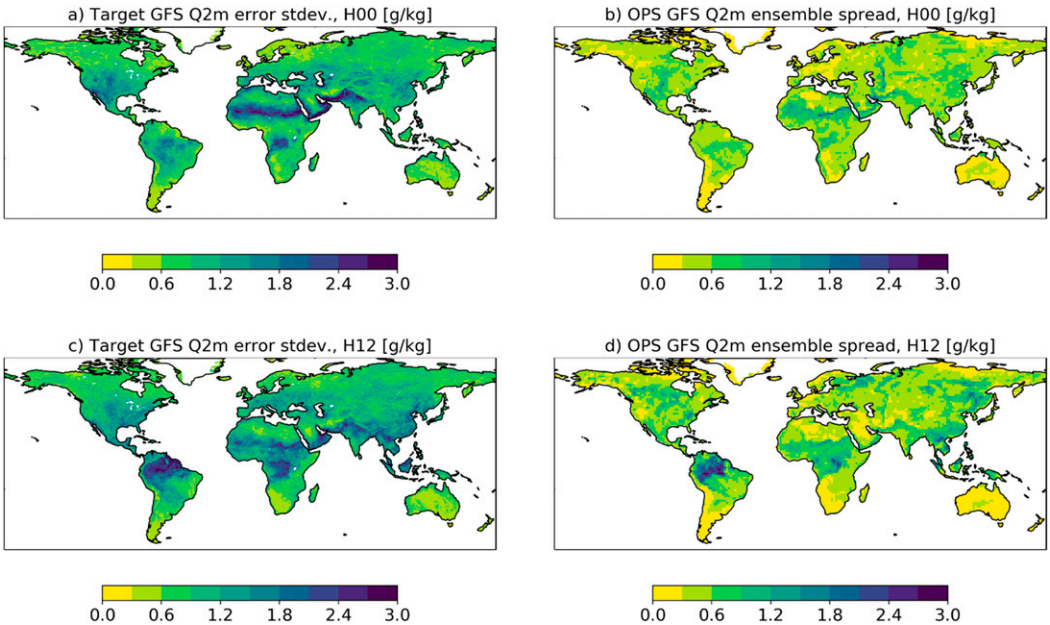

FIG. 4. Forecast uncertainty for GFS Q2m over boreal summer for 6-h time windows centered at (top) 0000 and (bottom) 1200 local time estimated by the (left) target error standard deviation (from comparison to ERA5) and (right) GFS NWP operational ensemble spread on 15 Jul 2019.

2) LAND SURFACE FORCING VARIABLES

Having confirmed that the operational GFS ensemble is underdispersed at and near the land surface, the role of the spread in the atmospheric fields that force the land surface is now investigated. This is done by comparison to the atmospheric ensembles typically applied in offline land data assimilation systems, using NASA's SMAP Level 4 (De Lannoy and Reichle 2016; Reichle et al. 2017) land data assimilation system as a representative offline system [NASA's Land Information System (Kumar et al. 2014) uses very similar atmospheric perturbations to SMAP Level 4]. These offline systems are based on ensembles of offline (land-only) land model forecasts. A single set of atmospheric fields is used to force all ensemble members, with randomly generated perturbations added to select fields in each ensemble member. For the SMAP Level 4 system, precipitation and downwelling surface shortwave radiation are both multiplicatively perturbed (with lognormally distributed perturbations, of standard deviations of 0.5 and 0.3, respectively), and the downwelling surface longwave radiation is additively perturbed (with normally distributed perturbations of standard deviation 20 $\mathrm{W\,m^{-2}}$).

Figure 7 compares the ensemble spread in the atmospheric forcing from the operational GFS ensemble to that from the SMAP Level 4 system. The SMAP Level 4 estimate was constructed by calculating the ensemble standard deviation that would result from applying the above listed perturbations to a single (arbitrarily chosen) member of the operational GFS ensemble. There are clear differences in both the spatial patterns and magnitude of the ensemble spread from the coupled and offline systems (here "coupled" refers to the land model being coupled to the atmosphere). Since the land ensemble members within the coupled ensemble are each forced by a different realization of the atmosphere (rather than a randomly perturbed copy of a single realization, as occurs offline), the ensemble can account for uncertainty in the location of precipitation events. Hence, it has a larger area on nonnegligible ensemble spread in precipitation (more than 2.25 times) than the offline ensemble (Figs. 7a,b). Similarly, for the radiation terms, spatial variation in the

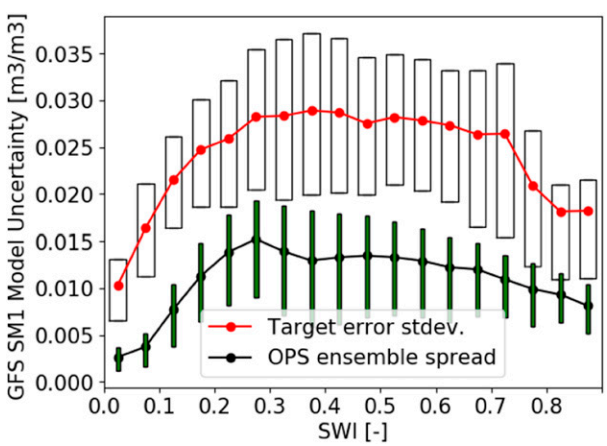

FIG. 5. Forecast uncertainty in GFS SM1 from the target estimate (triple collocation) and the GFS NWP operational ensemble spread on 15 Jul 2019 (both from data plotted in Fig. 1), binned by soil wetness index (SWI). Plotted circles show the mean for each bin, and the boxplots show the interquartile range.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

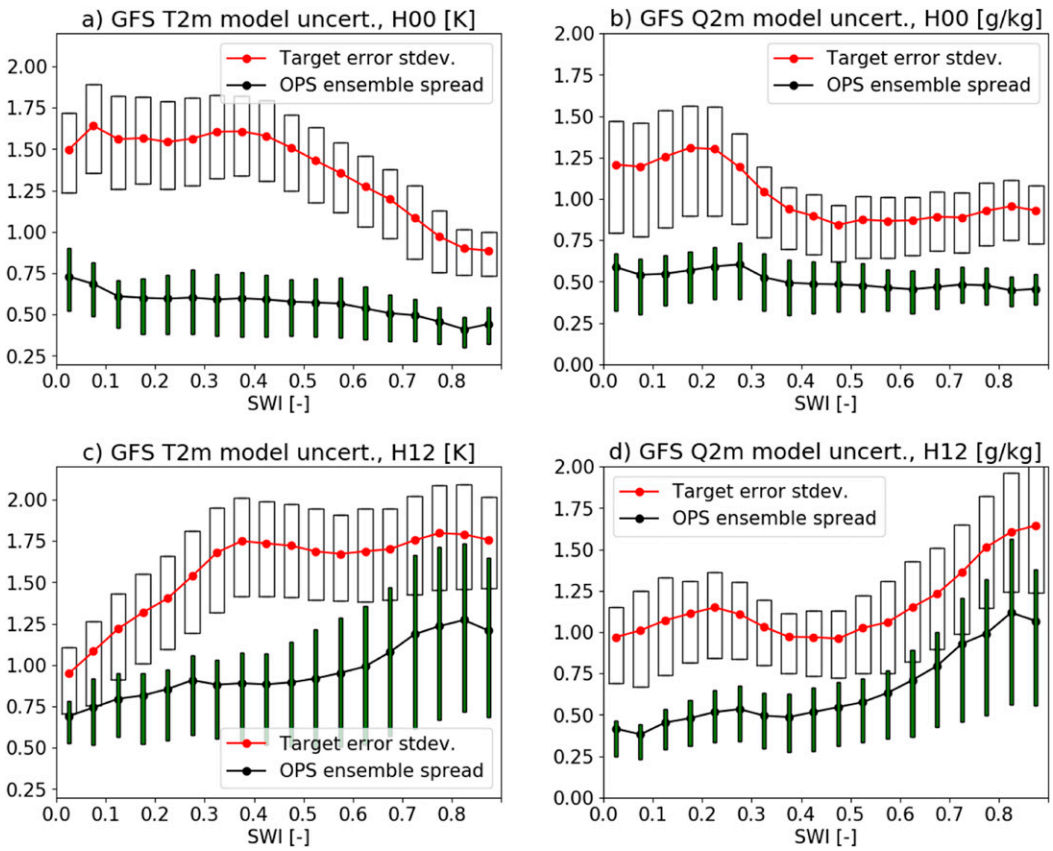

FIG. 6. Forecast uncertainty in GFS (left) T2m and (right) Q2m, for 6-h time windows centered at (top) 0000 and (bottom) 1200 local time from the target estimates (comparison to ERA5) and from the GFS NWP operational ensemble spread on 15 Jul 2019 (both from data plotted in Figs. 3 and 4), binned by SWI. Plotted circles show the mean for each bin, and boxplots show the interquartile range.

coupled estimates reflects uncertainty in the forecast cloudiness (Figs. 7d,f), producing more realistic spatial patterns than the offline ensemble, which is globally uniform for longwave (Fig. 7c) and has the greatest shortwave uncertainty (Fig. 7e) where shortwave is itself greatest (i.e., under clear-sky conditions, which can generally be forecast more confidently than cloudy skies). However, while the magnitude of the ensemble spread in the precipitation is similar (the mean across regions with nonnegligible precipitation spread is 3.2 mm day^{-1} coupled, 3.9 mm day^{-1} offline), the spread in the radiation fields in the coupled ensemble is well below the offline estimates. For shortwave, the mean ensemble spread in regions of nonnegligible spread was just 50 W m^{-2} for the coupled ensemble, compared to 116 W m^{-2} for the offline ensemble. For longwave, the difference is less marked, with a mean of 8.4 W m^{-2} for the coupled ensemble, compared to 20 W m^{-2} for the offline ensemble. While the underdispersed spread in radiation in the GFS ensemble will contribute to the underdispersed land surface spread, the developers of offline land ensemble systems have found that perturbing the forcing alone is insufficient to achieve reasonable ensemble spread in the land. Hence, increasing the radiation spread in the GFS ensemble alone is unlikely to address the underdispersal of the land component.

b. *Land perturbation experiments*

Figures 8–10 show the forecast uncertainty estimated from the ensemble spread for each of the experiments listed in Table 2, compared to the target forecast uncertainty estimates in Table 1. These experiments have not been tuned to optimize the magnitude of the induced ensemble spread, and the focus is on investigating how the model responds to the different perturbation approaches, in terms of aspects such as spatial distribution in the induced spread, and the impact on ensemble correlations. Starting with soil moisture in Fig. 8, note that the distribution of spread in the control experiment is very similar to that from the archived operational output in Fig. 5, confirming that the experimental setup for the control (and other experiments) is representative of the operational GFS's ensemble spread in the land. Also, in Fig. 8, the state-pert experiment has the greatest increase in spread, and also has a distribution that is quite different from the control and target estimates, showing a narrow maxima at a SWI around 0.2, rather than the broader maxima between SWI of 0.2 and 0.7. Examination of the model response to the applied perturbations

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

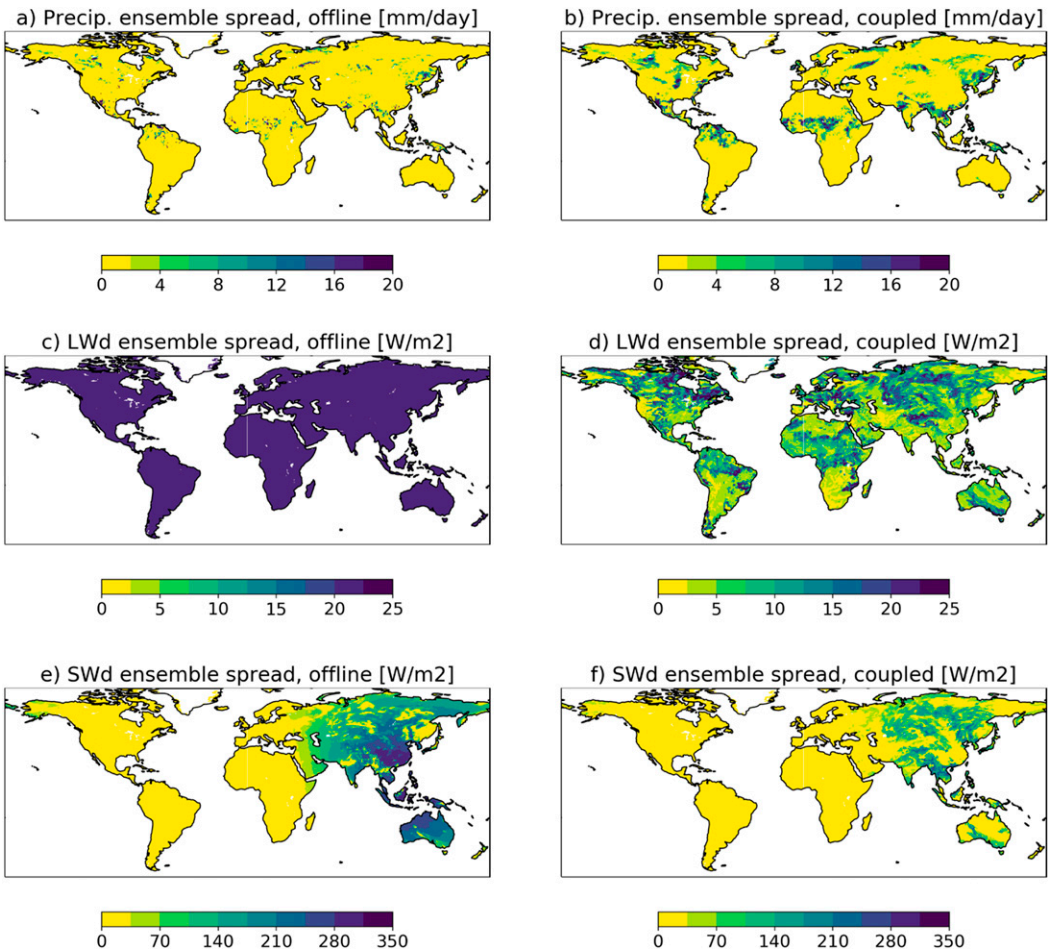

FIG. 7. Example of the ensemble spread in selected land forcing fluxes for (top) precipitation (prec.), (middle) downwelling longwave (LWd), and (bottom) downwelling shortwave (SWd) for (left) typical values used in an offline land data assimilation systems and (right) the land forecast coupled to the atmosphere in the operational GFS NWP ensemble. All plots are for 16 Aug 2019 for the 4-h forecast valid at 0400 UTC.

shows that this is caused by the model retaining the applied perturbations for longer under drier conditions, due to the reduced evaporation rates. Over time, as the (temporally correlated) perturbations are added to the model states, they accumulate more rapidly under drier conditions, enhancing the ensemble spread in those regions. There is an additional experiment plotted in Fig. 8, labeled state-pert-noflattop, which is the same as the state-pert experiment, but without the flat-top filter [Eq. (2)] applied. In this experiment, the overestimation of the ensemble spread is even worse in dry regions (the introduction of the flat-top filter was an attempt to fix this issue, but it made only a slight improvement).

The remaining two experiments in Fig. 8, soil-SPPT and param-pert, added less spread than the soil-pert experiment, while maintaining the pattern of the target estimate of relatively large spread in the center of the SWI distribution. The soil-SPPT experiment increased the mean spread only slightly, to a mean of $0.013\,\mathrm{m}^3\,\mathrm{m}^{-3}$, compared to a mean of $0.010\,\mathrm{m}^3\,\mathrm{m}^{-3}$ for the control experiment. While the magnitude of the spread has not been tuned in these experiments, and is not being explicitly evaluated, for the soil-SPPT experiment additional tests attempting to increase the soil moisture spread were unsuccessful (while for param-pert, it was possible to increase the spread by increasing the magnitude of the applied perturbations). For soil-SPPT, doubling the standard deviation of the distribution for the s parameter in Eq. (4) increased the mean ensemble spread by just $0.001\,\mathrm{m}^3\,\mathrm{m}^{-3}$. The soil moisture spread induced by soil-SPPT is limited by two factors. First, the need to restrict $(1 + r_t)$ in Eq. (3) to be nonnegative, combined with the use of symmetric perturbations, limits the applied $(1 + r_t)$ to between 0 and 2, effectively limiting the maximum standard deviation of the applied perturbations. Second, the change in soil moisture over a time step is itself very small, unless precipitation has recently occurred. Combined, the above two factors limit the magnitude of perturbations that can be applied when applying SPPT to soil moisture, since both terms in Eq. (3) are small. This is an inherent limitation of the method itself, rather than in the tuning of the experiment presented here.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

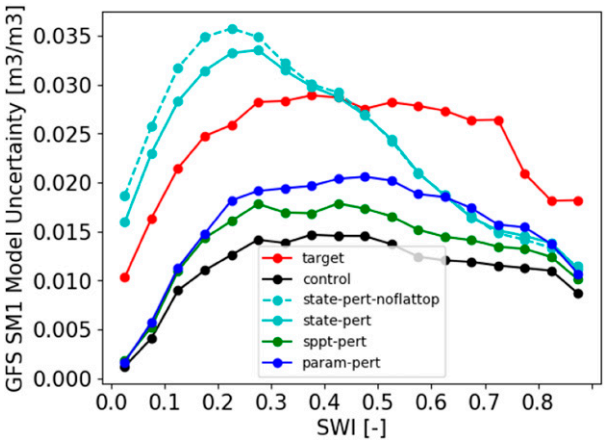

FIG. 8. Forecast uncertainty in the GFS layer 1 soil moisture (SM1) from the target error standard deviation estimate (triple collocation) and from the ensemble spread in each of the land perturbation experiments at 0000 UTC 14 Aug 2019, binned by SWI. Plotted circles indicate the mean for each bin.

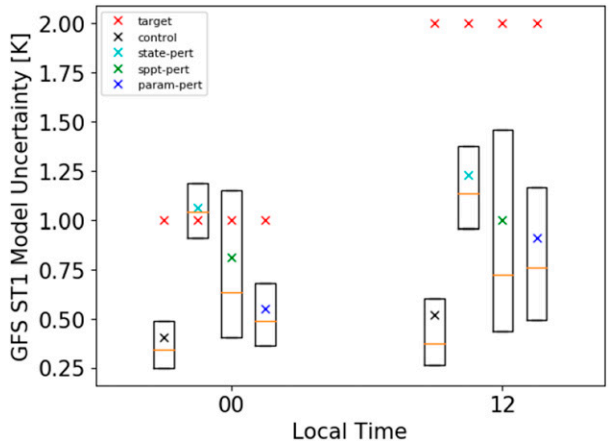

FIG. 9. Forecast uncertainty in GFS layer 1 soil temperature (ST1) from the target error standard deviation estimate (Holmes et al. 2012) and from the ensemble spread in each of the land perturbation experiments on 14 Aug 2019 for the 6-h time windows centered on 0000 and 1200 local time. Boxplots for the ensemble spread show the median and interquartile range.

Figure 9 shows boxplots of the global gridded ensemble spread in soil temperature for each experiment, compared to the literature estimates of global mean forecast uncertainty. Each experiment has increased the spread, although in most instances it remains well below the target literature estimate, particularly during the day.

The results are similar for the 2-m variables in Fig. 10, in that the ensemble spread was increased in all of the experiments, while remaining well below the target estimates. This was the case even for the state-pert experiment, which added a relatively large spread to the soil moisture and temperature (particularly under dry conditions). This is not unexpected, as the forecast errors in the 2-m variables are not exclusively due to errors in the land surface processes. For T2m, all three perturbation experiments added a similar amount of spread during the day (with some deviation in the exact distributions), while during the night the relative amount of spread added by each experiment follows the same pattern for the relative amount of spread added to soil temperature in Fig. 9 (in order of increasing spread: param-pert, soil-SPPT, state-pert). Note, however, that the parameter perturbation experiment has barely changed the T2m spread at night. For Q2m, the perturbation experiments all added a similar small amount of spread, while still remaining far below the target estimate. The exception is the vegetation parameter perturbation experiment, for which the additional Q2m spread is gradually reduced as SWI increases, while during the day the amount of spread is actually decreased, compared to the control experiment.

This reduction in spread for the parameter perturbation experiment is curious, and is associated with the control experiments having relatively high Q2m (and T2m) ensemble spread in several small regions, all characterized by densely vegetated tropics. These regions of elevated spread are seen in all of the other experiments, except param-pert, and are also evident in the operational ensemble. Figure 11 plots T2m and Q2m versus precipitation for each ensemble member for the param-pert and control experiments at a grid cell in one of these regions. The higher ensemble spread in the control experiment is associated with a larger difference in T2m and Q2m between the precipitating and nonprecipitating ensemble members than occurs in the param-pert experiment. The nonprecipitating members are warmer and drier in the control experiments (mean T2m of 303.1, mean Q2m of 14.0 g kg^{-1}) than those in the param-pert experiment (means 301.2 K, 16.2 g kg^{-1}). The ensemble members with precipitation then have higher precipitation amounts in the control (mean 20.5 mm day^{-1}) than in the param-pert (mean: 7.1 mm day^{-1}), leading to cooler and wetter T2m and Q2m (means: 299.4 K, 19.7 k kg^{-1}) than in param-pert (300.5 K, 16.6 g kg^{-1}). Across this region the param-pert experiment tends to be cooler in the absence of precipitation, and the hypothesis is that the cooler conditions lead to reduced precipitation amounts, giving a lesser response to that precipitation, reducing the ensemble spread between the precipitating and nonprecipitation ensemble members. Regardless of the exact mechanism, this highlights that the nonlinear response of land models to applied perturbations can alter the ensemble mean behavior, which can then have significant follow-on effects.

Finally, Figs. 12 and 13 show the ensemble correlations between the soil states (moisture and temperature), and each of T2m and Q2m, for each of the experiments, again binned by SWI. A few interesting features stand out from these figures. Starting with the control experiment, the patterns of each correlation term appear generally reasonable, with the exception of the positive nighttime soil moisture–T2m correlation under dry conditions, which is the opposite sign to observed and modeled estimates of the correlation (Dai et al. 1999; Drusch et al. 2009). This is present in all of the experiments as well as in the archived operational ensembles (and so it is not associated with a particular perturbation approach),

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

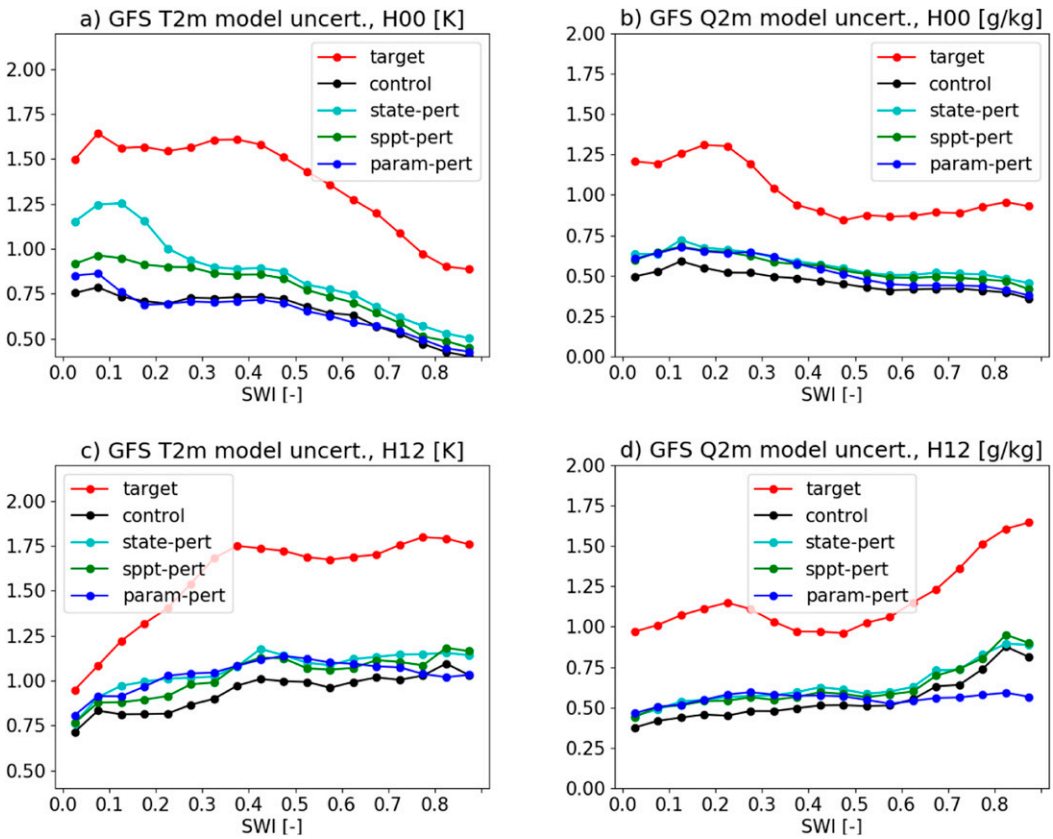

FIG. 10. Forecast uncertainty in GFS (left) T2m and (right) Q2m for the 6-h time windows centered at (top) 0000 and (bottom) 1200 local time from the target error standard deviation estimate (comparison to ERA5) and from the ensemble spread in each of the land perturbation experiments on 14 Aug 2019, binned by SWI. Plotted circles show the mean for each bin.

and indicates an error in the forecast model. This model error will be problematic when performing assimilation updates to the model soil moisture (since an assimilation update improving the soil moisture will generate a change in the T2m of the incorrect sign). The cause of this model error is unknown, and is currently being investigated by NCEP's model development team.

Next, looking at the perturbation experiments, there is a clear distinction between the two experiments that add perturbations directly to the soil states (state-pert and soil-SPPT), and the parameter perturbation experiment, which perturbed the state variables indirectly by perturbing the calculated water and energy fluxes. The state-pert experiment generally had the greatest change in the magnitude of the ensemble correlations, particularly for drier soil moisture conditions, when it also induced the largest ensemble spread in soil moisture. However, the magnitude of the ensemble correlations is sometimes increased and sometimes decreased by the state-pert experiment. In a coupled system, like a land–atmosphere model, this is not an unexpected response to applying perturbations to one component of that system. For example, land–atmosphere interactions, such as are quantified by the correlations between the soil states and 2-m variables plotted here, consist of both the atmosphere influencing the land, and the land influencing the atmosphere, with the balance between these two varying according to the local conditions (Seneviratne et al. 2010; Dong and Crow 2019). Under dry soil conditions, the soil moisture available for evaporation will exert a strong control over the partition of incoming surface radiation into latent and sensible heating, and so will control the T2m and Q2m. When the state-pert experiment then adds perturbations to the soil moisture, in dry conditions this is translated into perturbations in the T2m and Q2m, enhancing the preexisting soil moisture–2-m correlation in Figs. 12a and 12c. On the other hand, in conditions where the correlation between the land and atmosphere is driven by atmospheric processes controlling the soil states, adding perturbations to the soil states will reduce the 2-m atmosphere–soil state correlations, by adding noise to the soil states used in the correlation calculation. The reduced daytime correlations between the soil temperature and T2m in Figs. 13a and 13b suggest that this occurred in these locations.

The soil-SPPT experiment also added perturbations directly to the soil states, resulting in a similar response to state-pert in the ensemble correlations for soil temperature. By contrast, the soil-SPPT induced little change in the soil moisture correlations, likely because it induced such a small change in the soil moisture itself.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

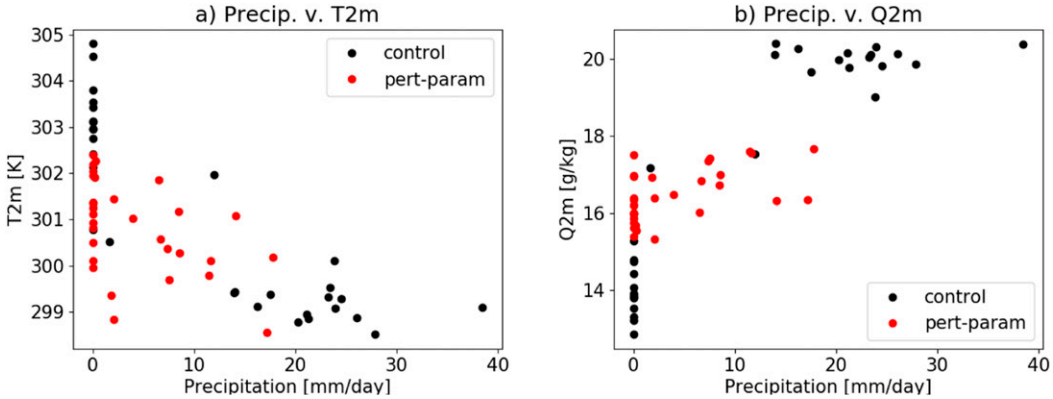

FIG. 11. Precipitation vs (a) T2m and (b) Q2m for each ensemble member at a grid cell in the Amazon (lon = 294.9, lat = 3.2) at 1800 UTC (~1400 local time) 10 Aug for the control and param-pert experiments.

In contrast to the other experiments, the parameter perturbation experiment generally increased the magnitude of the preexisting land–atmosphere correlations in Figs. 12 and 13. Again this is expected, since perturbing the parameters induces perturbations in the land–atmosphere fluxes that define the land–atmosphere coupling. Perturbing these fluxes generates consistent perturbations in the land and atmosphere that emphasize the preexisting land–atmosphere coupling, regardless of which component is driving the land–atmosphere controls in that instance.

4. Discussion

This study examined the ensemble spread in the soil moisture and temperature and the atmospheric 2-m temperature and humidity in NCEP's operational GFS NWP ensemble system. As is the case in other NWP ensemble systems, the GFS ensemble is underdispersed at and near the land surface. This lack of spread prevents the use of land data assimilation based on the GFS ensemble estimates of forecast uncertainty. While the ensemble-based forecast uncertainty estimates could be inflated to obtain more realistic estimates, in many regions the ensemble cross covariances between the soil and 2-m variables are noisy (not shown), suggesting that the forecast spread is small enough to induce sampling errors in the cross covariances. The ensemble spread in the surface radiation fields are likely also underestimated over land, and could be improved by enhancing the uncertainty of cloud processes in the ensemble members. However, based on the experience of offline land data assimilation systems, enhancing the forcing spread alone would likely be insufficient to obtain a more realistic spread in the land states. Hence a series of experiments was conducted to test different methods of explicitly introducing perturbations to account for the uncertainty in the soil moisture and temperature forecasts in each ensemble member.

The results of these experiments highlighted the ways in which the dynamics of the land, and hence land models, differ from those of the atmosphere. As noted earlier, the land is not a chaotic system and the land surface schemes used in atmospheric models do not simulate horizontal flow between grid cells. Consequently, land models do not have dramatic error growth along instabilities, or horizontal communication of forecast errors between neighboring grid cells. Instead, the spatial patterns of forecast uncertainty in an ensemble of forecasts will be largely determined by the spatial patterns of perturbations applied to the ensemble, combined with the localized persistence of these perturbations in the model. This was clearly demonstrated by the state-pert experiment, in which soil moisture and soil temperature perturbations were applied globally, assuming the same Gaussian perturbation distribution everywhere. At the regional level, this resulted in soil moisture ensemble spread that reflects the local persistence of soil moisture perturbations, resulting in larger values in drier regions, which does not agree with the target soil moisture forecast uncertainty estimates (Fig. 8).

This result raises interesting questions about the role of errors of the day in land data assimilation. If forecast uncertainty is not driven by chaotic processes, and instead reflects localized errors in the modeled physical processes and the model memory of those errors, then the forecast errors at any given point in space and time will depend strongly on the local conditions. One might then be tempted to approximate the errors using a climatological forecast uncertainty estimate. However, atmospheric forcing is a major source of land model forecast error, and so superimposed on the climatological error covariance field are forecast errors associated with transient atmospheric events, most notably precipitation. The latter cannot be reasonably neglected, since errors in these events will induce large errors in the forecast land surface state, while also potentially shifting the land surface regime (say to a wetter state after precipitation), which can dramatically affect the error covariances between pairs of variables (say, between soil moisture and T2m).

Following the state-pert experiment, the soil-SPPT experiment was an attempt to apply soil moisture and temperature perturbations with a more realistic spatial pattern, by making use of the model tendencies to determine the magnitude of the perturbations applied to the soil states. However, the results showed that when applying SPPT to the soil states, the amount

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

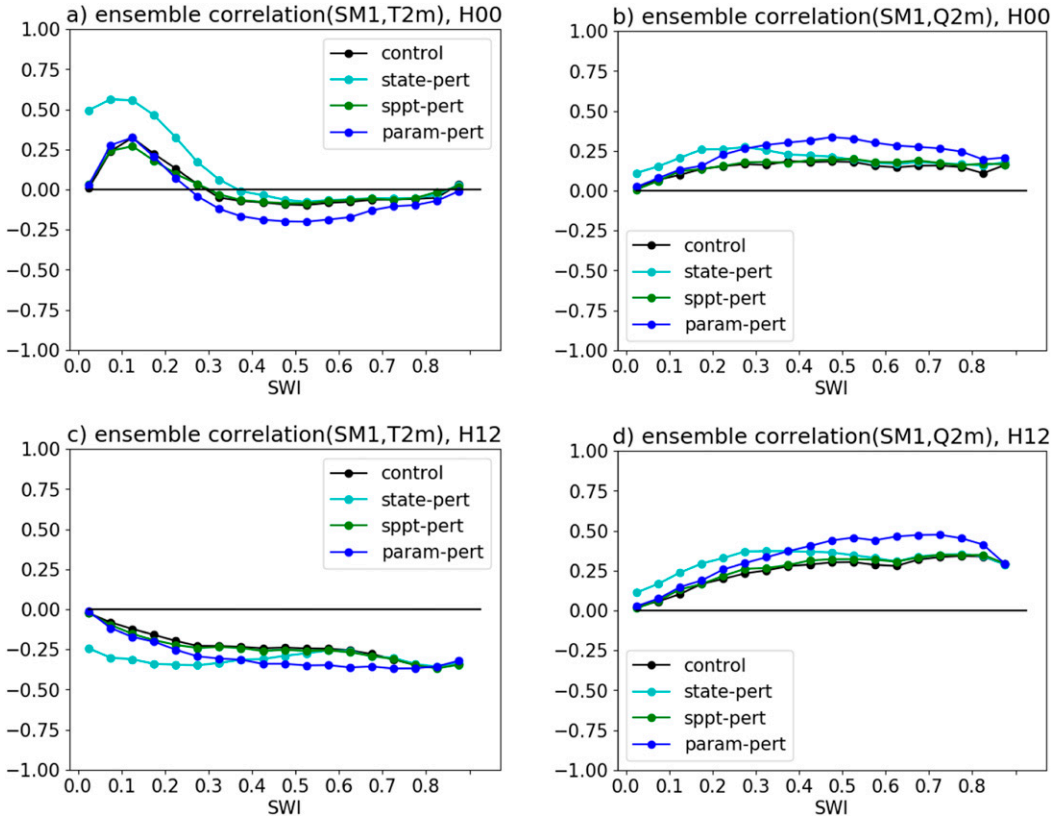

FIG. 12. Ensemble correlation between layer 1 soil moisture (SM1), and each of (left) T2m and (right) Q2m for the 6-h time windows centered at (top) 0000 and (bottom) 1200 local time binned by SWI. Plotted circles show the mean for each bin.

of soil moisture spread that could be induced is limited, due to the need to maintain perturbations that are both symmetric about unity and positive, and due to the typically very small change in soil moisture in an individual time step. Hence, the SPPT approach is not recommended for soil states. The third experiment, param-pert, was also designed to make use of the model physics to introduce consistent perturbations across a range of model variables, and this experiment was more successful. Perturbing the model parameters (in this case, vegetation fraction) led to a distribution in the soil moisture ensemble spread that agreed reasonably well with that from the target estimate in Fig. 8. More importantly, perturbing the parameters effectively perturbs the fluxes across the land–atmosphere interface, resulting in land and atmosphere perturbations representative of errors in those fluxes. Hence, the preexisting land–atmosphere ensemble error correlations from the control experiment were enhanced in the param-pert experiment. By contrast, both the state-pert and soil-SPPT experiments applied perturbations to only the land states within the ensemble, which resulted in enhanced land–atmosphere error correlations in locations where the land states were driving the land–atmosphere fluxes (i.e., moisture-limited locations during the day), and reduced land–atmosphere error correlations in locations where the atmospheric states were driving the land–atmosphere fluxes (soil temperature and T2m during the night). This is an important result, since the land–atmosphere error covariances are central to coupled land–atmosphere data assimilation. Within the context of an EnKF assimilation designed to constrain the model soil states by assimilating 2-m observations, applying a perturbation approach that only perturbs the soil states is expected to over/undercorrect the soil states where the ensemble has over/underestimated the strength of the land–atmosphere error correlations, depending on the whether the land–atmosphere interactions are locally controlled by the land or the atmosphere. The more representative land–atmosphere error covariances obtained from the parameter perturbation approach are expected to yield improved ensemble-based data assimilation results; however, data assimilation experiments are required to confirm this result.

The objective of this study was to investigate how different methods of accounting for land model error affected ensemble-based estimates of the forecast error covariances that are central to coupled land–atmosphere data assimilation. Investigation of how the different land model uncertainty methods may have affected the ensemble mean states and forecast skill is left for later work. However, Fig. 11 highlighted that the nonlinear response of land models to symmetrically applied perturbations can affect the ensemble mean states, and also downstream model processes. If the applied ensemble perturbations

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

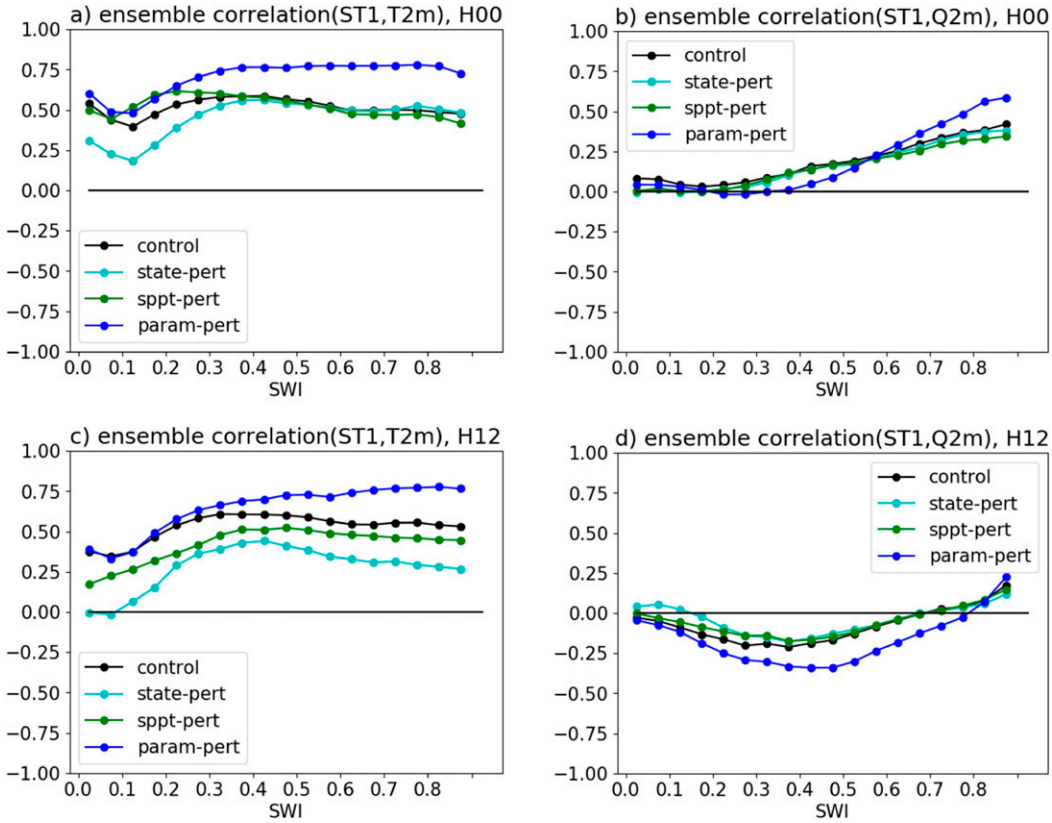

FIG. 13. Ensemble correlation between layer 1 soil temperature (ST1), and each of (left) T2m and (right) Q2m for the 6-h time windows centered at (top) 0000 and (bottom) 1200 local time, binned by SWI. Plotted circles show the mean for each bin.

reflect the true model uncertainty (say, by representing subgrid-scale heterogeneity), we would expect the perturbations to improve the ensemble forecast mean and skill, as occurred in MacLeod et al. (2016), Orth et al. (2016), and Strommen et al. (2019). However, as noted by Strommen et al. (2019), if the perturbation scheme significantly adjusts the mean land states, this may require a time-consuming model retuning to obtain these forecast benefits, since NWP boundary layer schemes tend to be heavily tuned to their current (deterministic) land model climatology.

5. Conclusions and recommendations

A number of conclusions and recommendations can be drawn from this study, starting with those for offline land data assimilation systems. First, the standard method of adding soil moisture perturbations to each ensemble member at regular time intervals produces unrealistic spatial patterns in the ensemble-based forecast soil moisture uncertainty. Specifically the forecast uncertainty is overestimated (relative to other locations) in dry regions. At a minimum, applying a flat-top filter to the soil moisture perturbations is recommended. By reducing the magnitude of the perturbations close to the soil moisture limits, the flat-top filter ameliorates, but does not solve, the overestimated forecast soil moisture uncertainty in dry regions. While not investigated here, the flat-top filter is also expected to improve the tendency for offline land model ensembles to be biased wet in dry conditions, compared to a nonperturbed simulation. Second, now that most atmospheric reanalysis and NWP systems are ensemble-based, offline land data assimilation systems would benefit from using these ensembles as atmospheric forcing (currently, only the Data Assimilation Research Testbed uses an ensemble of atmospheric realizations; Fox et al. 2018). In addition to providing atmospheric forcing with internal consistency between variables in each ensemble member, using an ensemble of atmospheric forecasts will provide more accurate spatial variation in the model forcing uncertainty, through better representing the impact of cloudiness uncertainty on radiation uncertainty, and the positional uncertainty in clouds and precipitation.

In terms of coupled land–atmosphere data assimilation, as well as other coupled data assimilation systems (e.g., ocean/atmosphere), ensemble perturbation approaches that directly target the fluxes between the coupled components are recommended. This approach creates physically consistent perturbations in each component (in this case the land and the atmosphere), leading to error cross covariance that reflect the coupling generated by the perturbed fluxes. By contrast, perturbation methods that target only one component will lead to overestimated ensemble error covariances where that component

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

is driving the coupling between the components, and underestimated covariances where the other component is driving the coupling.

For NCEP's global NWP system specifically, based on these results the parameter perturbation approach will be applied to enhance the ensemble spread at and near the land, as the first step to developing an EnKF-based coupled land–atmosphere data assimilation system. This approach both satisfies the above requirement of directly perturbing the land–atmosphere fluxes and leads to a realistic distribution of the forecast uncertainties. Additional testing is now being carried out to refine the parameter perturbation approach to also perturb additional parameters, and to test the impact throughout the year. For the EnKF assimilation of 2-m variables, if perturbing the model parameters cannot induce sufficient ensemble spread to approximate the target error estimates in Table 1, the additional forecast error could be applied within the data assimilation using the Hybrid 4DEnVar approach already used for atmospheric data assimilation. Due to the incorrect nighttime soil moisture–T2m relationship observed in the GFS ensembles, initial development of the EnKF will focus on updating soil temperature from T2m, until this modeling issue is resolved. Finally, the code used to generate the state and parameter perturbations are available through the UFS Community (UFS Community Weather, and Stochastic Physics) GitHub repositories.

Acknowledgments. Philip Pegion, of NOAA/Earth Sciences Research Laboratories/Physical Sciences Laboratory, is acknowledged for assistance with calling and modifying the UFS stochastic physics package.

Data availability statement. Output from all ensemble experiments is currently archived at NOAA's National Environmental Security Computing Center, and can be made available by contacting the author.

REFERENCES

Albergel, C., and Coauthors, 2015: Soil temperature at ECMWF: An assessment using ground-based observations. *J. Geophys. Res. Atmos.*, **120**, 1361–1373, https://doi.org/10.1002/2014JD022505.

Bouttier, F., L. Raynaud, O. Nuissier, and B. Menetrier, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quart. J. Roy. Meteor. Soc.*, **142**, 390–403, https://doi.org/10.1002/qj.2622.

Cosgrove, B., and Coauthors, 2003: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). *J. Geophys. Res.*, **108**, 8845, https://doi.org/10.1029/2002JD003316.

Dai, A., K. Trenberth, and T. Karl, 1999: Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. *J. Climate*, **12**, 2451–2473, https://doi.org/10.1175/1520-0442(1999)012<2451:EOCSMP>2.0.CO;2.

De Lannoy, G., and R. Reichle, 2016: Global assimilation of multiangle and multipolarization smos brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *J. Hydrometeor.*, **17**, 669–691, https://doi.org/10.1175/JHM-D-15-0037.1.

Dong, J., and W. Crow, 2019: L-band remote-sensing increases sampled levels of global soil moisture-air temperature coupling strength. *Remote Sens. Environ.*, **220**, 51–58, https://doi.org/10.1016/j.rse.2018.10.024.

Draper, C., R. Reichle, R. de Jeu, V. Naeimi, R. Parinussa, and W. Wagner, 2013: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens. Environ.*, **137**, 288–298, https://doi.org/10.1016/j.rse.2013.06.013.

Drusch, M., K. Scipal, P. de Rosnay, G. Balsamo, E. Andersson, P. Bougeault, and P. Viterbo, 2009: Towards a Kalman filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System. *Geophys. Res. Lett.*, **36**, L10401, https://doi.org/10.1029/2009GL037716.

Ek, M., K. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, https://doi.org/10.1029/2002JD003296.

Fox, A., and Coauthors, 2018: Evaluation of a data assimilation system for land surface models using CLM4.5. *J. Adv. Model. Earth Syst.*, **10**, 2471–2494, https://doi.org/10.1029/2018MS001362.

Gehne, M., T. Hamill, G. Bates, P. Pegion, and W. Kolczynski, 2019: Land surface parameter and state perturbations in the Global Ensemble Forecast System. *Mon. Wea. Rev.*, **147**, 1319–1340, https://doi.org/10.1175/MWR-D-18-0057.1.

Global Modeling and Assimilation Office, 2015: MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged, Single-Level, Assimilation, Land Surface Diagnostics V5.12.4. Goddard Earth Sciences Data and Information Services Center, accessed 11 January 2020, https://disc.gsfc.nasa.gov/datasets/M2T1NXLND_5.12.4/summary.

Gruber, A., and Coauthors, 2020: Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sens. Environ.*, **244**, 111806, https://doi.org/10.1016/j.rse.2020.111806.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Holmes, T., M. Owe, R. De Jeu, and H. Kooi, 2008: Estimating the soil temperature profile from a single depth observation: A simple empirical heatflow solution. *Water Resour. Res.*, **44**, W02412, https://doi.org/10.1029/2007WR005994.

——, T. Jackson, R. Reichle, and J. Basara, 2012: An assessment of surface soil temperature products from numerical weather prediction models using ground-based measurements. *Water Resour. Res.*, **48**, W02531, https://doi.org/10.1029/2011WR010538.

Kleist, D., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational/ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, https://doi.org/10.1175/MWR-D-13-00350.1.

——, D. Parrish, J. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Kumar, S., and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeor.*, **15**, 2446–2469, https://doi.org/10.1175/JHM-D-13-0132.1.

——, R. Reichle, K. Harrison, C. Peters-Lidard, S. Yatheendradas, and J. Santanello, 2012: A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resour. Res.*, **48**, W03515, https://doi.org/10.1029/2010WR010261.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC

Lavaysse, C., M. Carrera, S. Belair, N. Gagnon, R. Frenette, M. Charron, and M. K. Yau, 2013: Impact of surface parameter uncertainties within the Canadian Regional Ensemble Prediction System. *Mon. Wea. Rev.*, **141**, 1506–1526, https://doi.org/10.1175/MWR-D-11-00354.1.

Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, https://doi.org/10.1002/qj.3094.

MacLeod, D., H. Cloke, F. Pappenberger, and A. Weisheimer, 2016: Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Quart. J. Roy. Meteor. Soc.*, **142**, 79–90, https://doi.org/10.1002/qj.2631.

O'Neill, P., S. Chan, E. Njoku, T. Jackson, R. Bindlish, and J. Chaubell, 2020: SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 4. NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 11 January 2020, https://doi.org/10.5067/NJ34TQ2LFE90.

Orth, R., E. Dutra, and F. Pappenberger, 2016: Improving weather predictability by including land surface model parameter uncertainty. *Mon. Wea. Rev.*, **144**, 1551–1569, https://doi.org/10.1175/MWR-D-15-0283.1.

Pinnington, E., J. Amezcua, E. Cooper, S. Dadson, R. Ellis, J. Peng, E. Robinson, and T. Quaife, 2021: Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data. *Hydrol. Earth Syst. Sci.*, **25**, 1617–1641, https://doi.org/10.5194/hess-25-1617-2021.

Putman, W., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, https://doi.org/10.1016/j.jcp.2007.07.022.

Reichle, R., R. Koster, J. Dong, and A. Berg, 2004: Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeor.*, **5**, 430–442, https://doi.org/10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2.

——, and Coauthors, 2017: Assessment of the SMAP Level-4 surface and root-zone soil moisture product using in situ measurements. *J. Hydrometeor.*, **18**, 2621–2645, https://doi.org/10.1175/JHM-D-17-0063.1.

Ryu, D., W. Crow, X. Zhan, and T. Jackson, 2009: Correcting unintended perturbation biases in hydrologic data assimilation. *J. Hydrometeor.*, **10**, 734–750, https://doi.org/10.1175/2008JHM1038.1.

Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner, 2008: A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophys. Res. Lett.*, **35**, L24403, https://doi.org/10.1029/2008GL035599.

Seneviratne, S., T. Corti, E. Davin, M. Hirschi, E. Jaeger, I. Lehner, B. Orlowsky, and A. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.*, **99**, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004.

Strommen, K., H. Christensen, D. MacLeod, S. Juricke, and T. Palmer, 2019: Progress towards a probabilistic Earth system model: Examining the impact of stochasticity in the atmosphere and land component of EC-Earth v3.2. *Geosci. Model Dev.*, **12**, 3099–3118, https://doi.org/10.5194/gmd-12-3099-2019.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 07/20/22 05:12 PM UTC