

Land Surface Parameter and State Perturbations in the Global Ensemble Forecast System

MARIA GEHNE

CIRES, University of Colorado Boulder, and NOAA/Earth System Research Laboratory, Boulder, Colorado

THOMAS M. HAMILL

Physical Sciences Division, NOAA/Earth System Research Laboratory, Boulder, Colorado

GARY T. BATES AND PHILIP PEGION

CIRES, University of Colorado Boulder, and NOAA/Earth System Research Laboratory, Boulder, Colorado

WALTER KOLCZYNSKI

I.M. Systems Group, and NOAA/NCEP/Environmental Modeling Center, College Park, Maryland

(Manuscript received 14 February 2018, in final form 2 February 2019)

ABSTRACT

The National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) is underdispersive near the surface, a common characteristic of ensemble prediction systems. Here, several methods for increasing the spread are tested, including perturbing soil initial conditions, soil tendencies, and surface parameters, with physically based perturbations. Perturbations are applied to the soil initial conditions based on empirical orthogonal functions (EOFs) of differences between normalized soil moisture states from two land surface models (LSMs). Perturbations to roughness lengths for heat and momentum, soil hydraulic conductivity, stomatal resistance, vegetation fraction, and albedo are applied, with the amplitude and perturbation scales based on previous research. Soil moisture and temperature tendencies are also perturbed using a stochastic perturbation scheme. The results show that surface perturbations, through their impact on 2-m temperature spread, have a modest positive impact on the skill of short-range ensemble forecasts. However, adjusting the forecasts using an estimate of the systematic bias shows that bias correction has a greater impact on the forecast reliability than surface perturbations, indicating that systematic bias in the model needs to be addressed as well.

1. Introduction

The current National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS; Zhou et al. 2017) is underdispersive, biased near the surface, or both. Like many other ensemble prediction systems (EPSs; Hamill and Colucci 1997; Buizza et al. 2000; Mullen and Buizza 2001; Hamill and Whitaker 2007) the GEFS produces an ensemble with too little spread and too high ensemble-mean error. Hamill and Colucci (1997) showed that the Eta-RSM 850-hPa temperature and precipitation rank histogram distributions are U shaped. Buizza et al. (2000) showed that

spread in the ECMWF EPS was too small compared to ensemble-mean error over Europe. Mullen and Buizza (2001) showed that rank histograms of precipitation forecasts in the ECMWF EPS are U shaped and Hamill and Whitaker (2007) showed similar behavior for 2-m temperature in the NCEP GFS. Ensembles tend to be more certain in their forecast than warranted, making decision support based on ensembles suboptimal. Ideally, ensemble forecasts should exhibit consistency between the spread of the ensemble and the root-mean-square error (RMSE) of the ensemble mean with respect to the truth (Fortin et al. 2014). Our hypothesis in this article is that one of the major causes of the insufficient spread is a lack of treatment of uncertainty in the soil state and in the associated land surface model (LSM) parameters,

Corresponding author: Maria Gehne, maria.gehne@noaa.gov

DOI: 10.1175/MWR-D-18-0057.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

and that spread can be increased through the introduction of realistic stochastic parameterizations of surface parameters and states. The LSM used in the GEFS is the Noah LSM (Ek et al. 2003).

What if this hypothesis cannot be confirmed? A possible explanation could be that we have not chosen to perturb the correct variables or parameters, or we have misestimated the magnitude of perturbations needed. In an ideal ensemble, the mean state is unbiased with respect to the verification data, which itself is commonly assumed to be free of error. If there are systematic biases, the RMSE will be enlarged, reflecting contributions from both random and systematic error, and it will thus be unrealistic to expect the introduction of stochastic methods related to land surface variables to result in consistency between spread and RMSE; the stochastic methods in general are designed to deal with model uncertainty, not ameliorate systematic error. For diagnostic purposes, then, it may be more appropriate to compare spread against the RMSE after systematic errors have been removed, such as through an ex post facto bias correction of the mean (Ciach and Krajewski 1999; Bowler 2008). Wang et al. (2018) show that it is necessary to bias correct an ensemble forecast before evaluating its skill, since the ensemble is expected to account for random error and not systematic error. Also, the assumption that the verification data is perfect may be particularly inappropriate for surface-related variables studied here, as land surface heterogeneity introduces large representativeness errors.

For improved near-surface predictions, we should consider initial condition uncertainty of both the atmospheric and the land state. Previous work found introducing perturbations to soil moisture initial conditions can increase the precipitation forecast spread (Sutton et al. 2006). Lavaysse et al. (2013) show improved skill in 2-m temperature (T_{2m}) and 10-m wind speed (U_{10m}) forecasts when perturbing the initial conditions for several surface parameters and variables using stochastic perturbations. They show soil moisture state, albedo, leaf area index, and SST perturbations had the largest impact on T_{2m} , though increase in spread was modest. Atmospheric initial states are typically defined through data assimilation (Hamill et al. 2011; Bonavita et al. 2012; Houtekamer et al. 2014; Houtekamer and Zhang 2016), combining two general types of information, a model forecast background state and newly available observations. Both data types have errors, and consequently the initial state is imperfectly estimated. Modern methods like the ensemble Kalman filter (EnKF; Mitchell and Houtekamer 2000; Hamill 2006; Evensen 2009; Wang et al. 2013) that underpin

atmospheric data assimilation in the current GEFS automatically estimate the initial state uncertainty.

For ensemble predictions, we seek an ensemble of state estimates drawn from the probability distribution of plausible analysis errors (Hamill et al. 2002). In contrast to the atmosphere, land-state uncertainty is more challenging to estimate and approximations are currently necessary for initializing an ensemble of soil states in the GEFS. An ensemble of initial state estimates for the land surface is generated through forcing of the land component of the Global Data Assimilation System (GDAS) with the GEFS member forecast model precipitation, surface solar radiation, and near-surface parameters: temperature, humidity, and wind speed (Campana and Caplan 2005). Since there is no formal soil state data assimilation in the GDAS the initial state estimate may drift significantly from truth. To mitigate model drift, soil moisture is nudged, with a 60-day relaxation, toward an externally supplied global soil moisture monthly climatology (Campana and Caplan 2005). These soil state estimates inherited from the GDAS may not represent a draw from the distribution of plausible analysis states and in the operational GEFS the surface initial conditions for all members are the same as the control.

However, it may be possible in future versions of the GEFS to simulate some uncertainty through a procedure of perturbing the soil state. Ensemble-based methods are commonly used for land-only data assimilation (Reichle et al. 2002), but are not yet part of NCEP's operational land-state estimation procedures. Here, we propose a different way to generate perturbed soil initial conditions by considering differences in soil states from two LSMs. Specifically, we will generate perturbations based on empirical orthogonal functions (EOFs) of differences between normalized soil moisture states from these LSMs.

Model uncertainty is also a factor in the growth of forecast error. This is a consequence of an imperfect model due to finite resolution, imperfect numerics of the dynamical core, and parameterization deficiencies. While there have been recent efforts to develop distribution-based parameterizations for some processes, most operational parameterizations are still deterministic (i.e., the response of a parameterization given the large-scale state in a grid column is always the same). In reality, the actual response can be sensitively dependent upon the unresolved scales. Further, in the context of the LSM there are several fixed parameters that are not well known, such as values for roughness lengths for heat and momentum, parameters related to soil hydraulic conductivity, stomatal resistance, and vegetation fraction (Clapp and Hornberger 1978;

Betts and Beljaars 1993; Hamill 1997; Sutton et al. 2006; Lavaysse et al. 2013; Tennant and Beare 2014). There are also potentially errors in the parameterizations that can lead to growth of biases. Further, subgrid-scale heterogeneity is not accounted for in the Noah LSM used with the GEFS. The land surface type (soil type and vegetation type) in a LSM grid cell is set to the dominant type within that grid cell instead of allowing multiple types in proportion to their actual gridcell fraction. The vertical structure of the soil is approximated by distinct layers of increasing thickness with depth and the coarse vertical discretization may be a source of model error. Empirical formulations for snow cover, snowmelt, runoff, and interception of precipitation by the vegetation cover are all necessarily approximations.

In recent years, several methods have been proposed to address atmospheric model uncertainty in reasonably simple ways: stochastically perturbed parameterization tendencies (SPPT; Buizza et al. 1999; Palmer et al. 2009, 2887–2908), the stochastic kinetic-energy backscatter scheme (SKEB; Shutts 2005; Berner et al. 2009) and the stochastic humidity perturbations in the boundary layer scheme (SHUM; Tompkins and Berner 2008) are, for example, planned for the upcoming version of the GEFS. More details on these can be found in section 2.

Perturbing physical constants within a parameterization is an alternative approach to using multiple parameterizations for different members. Common atmospheric constants or “parameters” to be perturbed are related to the convection, boundary layer, microphysics, or radiative transfer scheme (Bowler et al. 2008; Hacker et al. 2011; Reynolds et al. 2011; Leutbecher et al. 2017). Lavaysse et al. (2013) perturb vegetation fraction, leaf area index, albedo, and roughness length using multiplicative random perturbations and find a small but positive impact on near-surface spread. Ries et al. (2010) find a reduction in wind speed bias by reducing sea ice surface roughness length to an unrealistically small value. In a single column setting Pitman (1994) shows model sensitivity of latent and sensible heat fluxes, canopy temperature, soil moisture and soil temperature to perturbations of albedo, vegetation fraction, roughness length, and saturated hydraulic conductivity among other LSM parameters. Section 3 contains details on the perturbation strategies explored here. Section 4 describes the experiments.

The impacts of the surface perturbations are discussed in section 5, with focus primarily on T_{2m} . The focus on T_{2m} is due to the fact that there are reliable observations available and that other variables in the GEFS show

limited sensitivity to the surface perturbations consistent with previous studies. Impact of the perturbations on other variables is mentioned where appropriate.

2. The Global Ensemble Forecast System

The version of GEFS used in this study is V11.0.0, augmented by the stochastic physics schemes discussed below, with semi-Lagrangian horizontal advection run at T574 (roughly 27-km grid spacing) with 64 hybrid vertical levels (Zhou et al. 2017). The output is available at 6-h intervals and interpolated to a 1° grid. Model error in the GEFS version used here is represented by using three experimental stochastic physics schemes in development for GEFS v12: SPPT, SHUM, and SKEB. These replace the stochastic total tendency perturbations (STTP) used in the operational GEFS v11.0.0. Initial atmospheric perturbations are generated using 6-h EnKF background forecasts. For more details on the GEFS V11.0.0 configuration, see Zhou et al. (2017).

SPPT represents uncertainty within physical parameterizations. In SPPT, random spatial patterns are multiplied by the spatial patterns of the physical tendencies of model variables (Buizza et al. 1999; Palmer et al. 2009). The random spatial patterns have a specified decay time and spatial decorrelation scale, but no vertical variability, except that the amplitude is typically reduced near the surface and tapers to zero above 100 hPa for numerical stability (Palmer et al. 2009). The random pattern for each level uses a length scale of 500 km and a time scale of 6 h.

The SHUM scheme is based on the idea that the actual triggering of deep convection will happen from plumes below the scale of the model grid. There is a stochastic aspect to the subgrid variability of temperature and moisture. The stochastic effect of this subgrid variability within the parameterization of deep convection is estimated by perturbing the near-surface grid-scale humidity field directly, multiplying that field by a random pattern with mean 1.0 and variance that decays exponentially with height. Tompkins and Berner (2008) have implemented a stochastic convection scheme based on this idea, but cautioned that this can decrease probabilistic skill for some parameters in the medium range. Here a single random pattern is used, with a length scale of 500 km and a time scale of 6 h.

SKEB was developed to model the upscale propagation of small-scale variability that is commonly lost through numerical diffusion (Shutts 2005; Berner et al. 2009). SKEB introduces random perturbations to the streamfunction with a prescribed power spectrum and amplitude dependent on the local dissipation rate to counteract excessive kinetic energy loss in regions with

large dissipation. Unlike other implementations of SKEB, the GFS implementation only considers numerical dissipation, which is estimated by the magnitude of the vorticity gradient following [Palmer et al. \(2009\)](#). The omission of the mountain/gravity wave drag and convection scheme's contribution to dissipation is to avoid the double counting of physics tendencies since SKEB is intended to run concurrently with SPPT. The stochastic patterns for SKEB in the GEFS are correlated in the vertical by smoothing the patterns in the vertical by approximately 30 passes of a 1–2–1 filter. The random pattern for each level uses a length scale of 1000 km and a time scale of 6 h.

3. Stochastic surface perturbation methods and verification procedures

Here, we seek ways of increasing near-surface spread that are physically based. We hypothesize that a combination of state, parameter, and tendency perturbations will increase the spread. We choose to perturb initial soil moisture and temperature states ([Sutton et al. 2006](#); [Lavaysse et al. 2013](#); [Tennant and Beare 2014](#)), roughness length of heat and momentum ([Lavaysse et al. 2013](#)), leaf area index, soil hydraulic conductivity and albedo ([Hamill 1997](#)). We also extend the SPPT to perturb tendencies of soil moisture and temperature, achieved by multiplying the model tendencies with a random number, keeping the sign of the tendency but changing its magnitude. Both soil moisture and temperature tendencies are perturbed using the same random pattern and amplitude to ensure consistency between the soil moisture and temperature. However, it is not clear that soil SPPT perturbations are physically defensible, as will be discussed.

a. Initial state uncertainty

SOIL TEMPERATURE AND MOISTURE INITIAL CONDITIONS

Soil temperature (T) and soil moisture (θ) initial condition perturbations were generated using EOFs of normalized differences between soil state estimates from two different LSMs. This approach allows the perturbations to describe aspects of uncertainty in the soil state associated with the choice of LSM. The patterns identified through the EOF analysis describe the patterns associated with the most variance of the difference between the two LSMs.

The Global Land Data Assimilation System (GLDAS) drives several different LSMs to generate optimal estimates of the soil state using observed and analyzed atmospheric forcings ([Rodell et al. 2004](#)). Here we use

TABLE 1. Vertical layer depths for the Noah and CLM LSMs.

Noah	CLM
0–0.1 m	0–0.018 m 0.018–0.045 m 0.045–0.091 m
0.1–0.4 m	0.091–0.166 m 0.166–0.289 m 0.289–0.493 m
0.4–1.0 m	0.493–0.829 m 0.829–1.383 m
1.0–2.0 m	1.383–2.294 m 2.294–3.433 m

Noah (v2.7; [Ek et al. 2003](#)) and the Community Land Model (CLM v2.0; [Bonan et al. 2002](#)) GLDAS output from 1985 to 2010 at 1° and 3-hourly grid spacing, provided by the Goddard Earth Sciences Data and Information Services Center ([Rodell and Beaudoin 2007, 2015](#)). For the following analysis, diurnal variability was not considered and daily means of the data were used.

To find the perturbation patterns we focused on the average soil moisture content (%) and soil temperature (K) in the root zone, which corresponds to the top ~1 m of soil ([Guo and Dirmeyer 2006](#)). Because Noah and CLM have different vertical resolution, the top 1 m for Noah and 1.383 m for CLM were used. See [Table 1](#) for details on the vertical resolution of both LSMs. One other LSM (MOSAIC; [Koster and Suarez 1996](#)) was also considered, but due to MOSAIC having only three vertical layers it was decided to use CLM instead. The root-zone soil moisture (temperature) was computed by adding the (layer depth weighted) soil moisture (temperature) from the top layers down to the relevant depth (top three layers for Noah, top eight for CLM). The 1° GLDAS data was interpolated to the higher-resolution GEFS grid using conservative regridding.

Comparison of soil moisture from different LSMs is not straightforward. Soil moisture is model specific in its mean and variability, due to the fact that each LSM has different and model-specific parameters such as porosity, hydraulic conductivity, wilting point, and layer depth, as well as soil moisture being driven by the specific evaporation and runoff formulation of the model. Noah and CLM also use different soil textures. [Schaake et al. \(2004\)](#) show that the total water storage and its range in four LSMs is in fact very different. Consequently, [Koster et al. \(2009\)](#) argue that soil moisture should not be compared indiscriminately among LSMs. Rather, soil moisture should be interpreted as an index of the moisture state, and soil moisture should be standardized before comparison. Standardization here

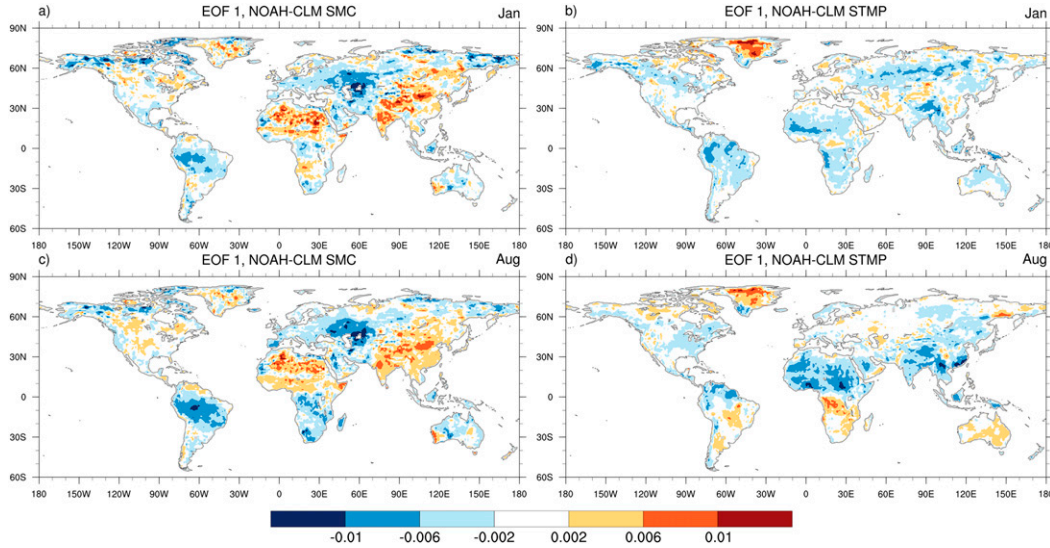


FIG. 1. Patterns of the leading joint EOF of the Noah and CLM difference for (a),(b) January and (c),(d) August. EOF patterns for (left) soil moisture content difference and (right) soil temperature difference. EOF patterns are unitless and are normalized so that the sum of squares of each pattern equals one.

refers to subtracting the climatological mean and dividing by the climatological standard deviation.

To represent the uncertainty in the initial soil state, the dominant modes of variability of the difference of the normalized soil moisture and temperature anomalies from both LSMs were estimated. Anomalies for each land model were computed using the daily climatology of the 26 years of daily GLDAS output. These anomalies were then standardized at each grid point by scaling by the full time series temporal standard deviation. The initial condition perturbations were computed as follows: let $\tilde{\theta}'_N$ and $\tilde{\theta}'_C$ be the standardized anomalies from the Noah and CLM LSMs, respectively, and similarly for T . Define the difference between the normalized Noah and CLM anomalies as

$$\Delta\theta(t, x_i, y_j) = \tilde{\theta}'_N(t, x_i, y_j) - \tilde{\theta}'_C(t, x_i, y_j) \quad (1)$$

$$\Delta T(t, x_i, y_j) = \tilde{T}'_N(t, x_i, y_j) - \tilde{T}'_C(t, x_i, y_j), \quad (2)$$

where t is time and x_i, y_j describe the longitude and latitude of the grid points, respectively. We then computed the first 20 joint EOFs (von Storch and Zwiers 1999) of the combined array of moisture and temperature differences $[\Delta\theta, \Delta T]$ for each month of the year after weighting by the cosine of latitude. To produce more smoothly varying EOF patterns throughout the year, a 10 day overlap with the prior and following month was used, so for each month there were roughly 26 years \times 50 days data points. The monthly EOFs were computed to permit seasonal differences in standardized

soil-state difference patterns. The eigendecomposition produced 20 spatial patterns for $\Delta\theta$ and ΔT each (E_m^l and F_m^l , $l = 1, \dots, 20$, $m = 1, \dots, 12$ for each month of the year m) and associated eigenvalues (λ_m^l). The first 20 EOFs capture about 70% of the total variance of the differences. The leading EOF patterns for January and August are shown in Fig. 1. Considering all 12 monthly EOF patterns, the leading soil moisture EOF patterns show only a weak seasonal cycle, in contrast to the higher-order EOFs. This indicates that the leading mode of variability of the difference between Noah and CLM soil moisture anomalies does not depend much on the time of year. The Noah LSM has more variance than the CLM LSM, but both are normalized before computing the difference. Based on the associated principal component time series the first EOF is most likely a representation of the interannual variability of the normalized differences.

For a given initial date in month m , following Houtekamer (1993), the perturbation patterns are as follows:

$$P_\theta^p(x_i, y_j) = \sigma_\theta \sum_{l=1}^{20} r_p^l \lambda_m^l E_m^l(x_i, y_j), \quad p = 1, \dots, N_p, \quad (3)$$

$$P_T^p(x_i, y_j) = \sigma_T \sum_{l=1}^{20} r_p^l \lambda_m^l F_m^l(x_i, y_j), \quad p = 1, \dots, N_p, \quad (4)$$

where the r_p^l are randomly drawn from a standard normal distribution, N_p is the number of ensemble members, and σ_θ and σ_T are the global mean of the temporal

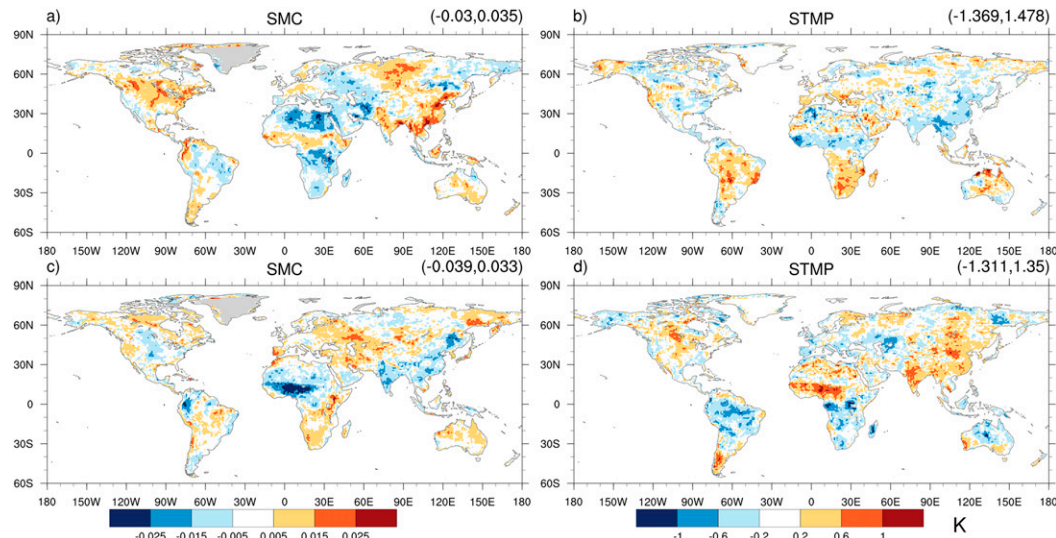


FIG. 2. Sample perturbation patterns as computed using Eqs. (3) and (4) for 1 Aug 2014 initial date. (a),(c) Sample soil moisture perturbations and (b),(d) sample temperature perturbations. Values in the top right corner give the minimum and maximum values.

standard deviation of the differences $\Delta\theta$ and ΔT , respectively. The random seed for the r_p^l depends on the initial date, but for a given initial date both T and θ perturbations are given the same random numbers. This approach allows generation of an arbitrary number of perturbation patterns and the model differences provide the relationship between $\Delta\theta$ and ΔT through the joint EOF analysis and the perturbation patterns preserve that relationship.

To ensure that the initial condition perturbations do not add energy or mass to the global system the (area weighted) global mean of the perturbations was normalized to zero for each ensemble member. At each grid cell the mean across ensemble members was also set to zero. The second normalization ensures that the ensemble mean is equal to the control mean. These two normalizations interact with each other, and the first one is more important. In the experiments presented here the perturbations are small and changes in the global mean from the second normalization are negligible. For experiments with larger perturbations it may be necessary to relax or omit the second normalization. The perturbed initial condition for all layers is given by the sum of the unperturbed value and the perturbations computed in Eqs. (3) and (4). The same perturbation is applied to all soil layers. By not scaling the soil moisture perturbations by the layer depth we add the same percentage of soil moisture, but not the same amount of moisture to each layer. An example of the perturbation patterns for two ensemble members is shown in Fig. 2. The perturbation patterns show distinct large-scale coherence for both soil moisture and temperature.

b. Model uncertainty

We now consider perturbations to LSM parameters. The two main considerations are amplitude and spatial pattern of the perturbations. Ideally, the spatial patterns of the perturbations should be based on empirical knowledge of the parameter uncertainties. However, observations necessary to estimate these parameter uncertainties are commonly unavailable, especially for estimating spatial decorrelation length scales. Therefore, as a first pass to gauge the impact of parameter perturbations, random spatial structures were used: for each ensemble member and case, the spatial pattern for the perturbation was generated identically to the atmospheric SPPT patterns (Palmer et al. 2009), described in section 2. These spatial patterns and the parameter perturbations associated with them were fixed for the duration of the forecast, but differed among the perturbed quantities. The amplitudes of the perturbations were based on existing literature estimating the uncertainty in these parameters, moderated by physical/ model constraints, where necessary. Table 2 presents a summary of the uncertainty or range of values associated with the parameters and variables that are considered in this section.

1) ROUGHNESS LENGTHS FOR HEAT AND MOMENTUM

The momentum roughness length z_0 was perturbed with logarithmic scaling of the perturbations

$$z_0' = z_0 10^a \quad (5)$$

TABLE 2. Uncertainty or range of values identified for the perturbed parameters.

Parameter or variable	Estimated uncertainty or range	Reference
Albedo α	2%–12% of α	Grant et al. (2000); Qu and Hall (2005)
Vegetation fraction σ	20%–30% of σ	Computed from MODIS vegetation fraction data
Momentum roughness length z_0	Factor of 0.316 – 3.16	Hamill (1997)
Ratio of heat to momentum roughness length (z_t/z_0)	Range: 10^{-4} – 10^{-1}	Beljaars and Holtslag (1991)
Soil hydraulic conductivity (SHC), exponent b	40% of b	Clapp and Hornberger (1978)
Leaf area index (LAI)	Range: 0–8	Bréda (2003)

with the spatial pattern a drawn from a normal distribution with a standard deviation of 0.14. Previous studies (Zhang and Anthes 1982; Diak et al. 1986; Betts and Beljaars 1993; Hamill 1997) suggested a plausible multiplicative uncertainty of z_0 by a factor of about 0.316 – 3.16, which corresponds to $a \in (-0.5, 0.5)$. Since a is drawn from a normal distribution the value chosen here means that about 95% of the values of a are within the interval $(-0.28, 0.28)$. This interval is smaller than the $(-0.5, 0.5)$ applied in Hamill (1997), because values that are too large lead to numerical instabilities in the GEFS boundary layer winds with the current parameterization.

The roughness length for heat (z_t) was not perturbed directly, but through perturbations of the momentum roughness length and perturbations of the ratio of the heat and momentum roughness lengths z_t/z_0 . This ratio may range from 10^{-4} – 10^{-1} (Beljaars and Holtslag 1991) and was perturbed similarly to z_0 , but with a standard deviation of 0.08 for the spatial pattern. Again, for reasons of numerical stability, the perturbation interval was chosen smaller than the uncertainty would suggest.

2) LEAF AREA INDEX

The value for leaf area index (LAI) in the GEFS V11.0.0 was set to a fixed value of 3.0 for all vegetation types. Perturbations were applied using a linear scaling

$$\text{LAI}' = \text{LAI}(1 + p), \quad (6)$$

where the spatial pattern p is drawn from a normal distribution with standard deviation of 0.25. Measurements of LAI give values from close to 0 to 8, with a large variability depending on the vegetation type, time of year, and measuring technique (Bréda 2003). Choosing a standard deviation of 0.25 means that about 95% of the perturbed LAI values are between 1.5 and 4.5. Using 1-km LAI data from the ECOCLIMAP database (Masson et al. 2003), the global average standard

deviation of LAI within a GEFS T574 cell is 0.5 and the GEFS T574 gridcell internal LAI standard deviation can be as high as 2.4 (not shown). The chosen value is conservative compared to these estimates of uncertainty and could potentially be increased in future studies.

3) SOIL HYDRAULIC CONDUCTIVITY

Soil hydraulic conductivity (SHC) was estimated from a measure of the soil wetness and an empirical exponent b :

$$\text{SHC} = \text{SHC}_{\text{sat}} \left(\frac{\theta}{\theta_{\text{sat}}} \right)^{2b+3}. \quad (7)$$

Here θ is the volumetric soil moisture, θ_{sat} is the saturation value of the soil moisture, and SHC_{sat} is the saturation soil hydraulic conductivity. Both θ_{sat} and SHC_{sat} depend on the soil type.

Clapp and Hornberger (1978) showed that for measurements of SHC based on soil samples of different soil types the standard deviation for estimates of this exponent b is approximately 40% of b . Since the goal of the perturbations was to represent the uncertainty in these estimates the exponent was perturbed using

$$b' = b(1 + p), \quad (8)$$

where the spatial pattern p was drawn from a normal distribution with standard deviation 0.4.

4) ALBEDO

Albedo α is a bounded quantity and the perturbed values need to be constrained between 0 and 1. To achieve this, albedo at a grid point was perturbed by a quantile mapping from the normal distribution the spatial patterns are based on to a beta distribution (Krzyštofowicz 1997; Wilks 2011). A beta distribution was estimated based on the unperturbed albedo value

and a perturbation size. The beta distribution probability density function is

$$\text{beta}(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\int_0^1 t^{a-1}(1-t)^{b-1} dt}, \quad (9)$$

with the shape parameters a and b are defined as

$$a = m^2(1-m)/s^2 - m, \quad (10)$$

$$b = a(1-m)/m, \quad (11)$$

where $m = \alpha$ is the unperturbed albedo value and the mean of the beta distribution, and $s = p_\alpha m(1-m)$ is the standard deviation of the beta distribution. The perturbed albedo value is then the value of the beta distribution corresponding to the percentile of the value given by the spatial pattern. For example, if the spatial pattern at a grid point is equal to 0.5, which corresponds to approximately the 70th percentile for a standard normal distribution, the 70th percentile of the estimated beta distribution at that grid point is picked as the perturbed albedo value.

The parameter p_α , which scales the standard deviation of the beta distribution, is a tunable parameter. [Qu and Hall \(2005\)](#) estimate the seasonal standard deviation of surface albedo (their Fig. 2). Variability is largest over regions with seasonal snow cover (about 12%) and smaller in the tropics (closer to 2%–4%). Estimates of the diurnal variability of surface albedo, based on point measurements ([Grant et al. 2000](#), their Fig. 2) suggest a global land mean standard deviation of around 5%. For perturbation sizes used in previous studies, [Lavaysse et al. \(2013\)](#) show the standard deviation of the initial perturbed albedo values in their Fig. 1a with values between 2% and 7%. The standard deviation of albedo values in a 5-day control run is a measure of the diurnal albedo variability in the GEFS. This is around 10% over most land areas, 20% over deserts, and 40% over snow covered ground. Based on these considerations a conservative perturbation size of $p_\alpha = 0.2$ was chosen, which resulted in a standard deviation of the beta distribution of 0.05 for mean albedo values equal to 0.5 and the 70th percentile value from the example above would be about 0.59. For a mean albedo of 0.1 the 70th percentile value of the beta distribution corresponds to a perturbed albedo value of 0.11.

5) VEGETATION FRACTION

Similar to albedo, vegetation fraction σ is bounded between 0 and 1. Vegetation fraction data at 1-km resolution based on measurements from the Moderate

Resolution Imaging Spectrometer (MODIS) ([Broxton et al. 2014](#)) vary the most within the larger GEFS grid cell in regions with intermediate vegetation fraction values and vary the least in regions with vegetation fraction close to 0 or 1. Accordingly, vegetation fraction uncertainty is largest for intermediate values and smallest for vegetation fraction values near 0 and 1. Global average standard deviation of the MODIS vegetation fraction data within a GEFS T574 grid cell is 2% and the maximum standard deviation within a GEFS T574 grid cell can be as high as 50% (not shown). For intermediate mean vegetation fraction values the standard deviation (i.e., uncertainty) is around 20%–30% on the GEFS T574 grid. Based on these considerations the vegetation fraction was perturbed in an identical manner to the albedo, but with perturbation size $p_\sigma = 1$ so that the standard deviation of the perturbations is 0.25 for vegetation fraction values of 0.5 and the perturbations go to zero as σ approaches its upper or lower bound.

c. Stochastically perturbed soil temperature and moisture tendencies

Another method for parameterizing model uncertainty is to perturb soil tendencies directly. Here this was accomplished by extending the SPPT method from the atmosphere to soil moisture and temperature. SPPT multiplies the tendency computed by the model physics step (for soil, this would be the full tendency) by a random pattern with spatial and temporal coherence. For simplicity and consistency, the random spatial pattern generation procedure used for the atmospheric SPPT was used for the soil SPPT as well. The amplitude used to generate the soil SPPT pattern is equal to the free atmospheric SPPT amplitude. This SPPT pattern was then applied to the soil temperature and moisture tendencies at each soil level. Since the soil tendency perturbations were applied after surface fluxes were computed, this could cause local violations of the surface energy budget. However, it is unclear how big an impact this may have on forecast accuracy, as any perturbation would be counterbalanced at the next time step by an adjustment of the fluxes. Future research may explore modifying surface fluxes, either in response to soil SPPT or perturbing fluxes in place of perturbing soil tendencies directly.

d. Verification methods

One characteristic of a reliable ensemble is that the spread of the ensemble should match the ensemble-mean forecast error. The ensemble-mean forecast error is the root-mean-square error (RMSE) between the ensemble mean and the verification. Ensemble spread was

computed by computing the variance at each grid point and forecast hour, averaging over all initial times and then taking the square root (Fortin et al. 2014). Let $f_p(t_n, h_k, x_i, y_j)$ ($p = 1, \dots, N_p, n = 1, \dots, N_i, h_k = 0, 6, \dots, 120$) be the forecast of ensemble member p for initial date t_n and forecast hour h_k , and let $o(t_n, h_k, x_i, y_j)$ ($n = 1, \dots, N_i, h_k = 0, 6, \dots, 120$) be the verification valid at the same time as the forecast for initial date t_n , forecast hour h_k and location x_i, y_j . Here N_i is the number of initial dates (11 in January and 11 in August) and $N_p = 20$ is the number of ensemble members. The time-averaged ensemble spread at each grid point is given by

$$s(h_k, x_i, y_j) = \sqrt{\frac{1}{N_i} \sum_{n=1}^{N_i} \frac{1}{N_p - 1} \sum_{p=1}^{N_p} [f_p(t_n, h_k, x_i, y_j) - \bar{f}(t_n, h_k, x_i, y_j)]^2}, \quad (12)$$

where $\bar{f}(t_n, h_k, x_i, y_j)$ is the ensemble mean. For area averages, the variance $[s(h_k, x_i, y_j)]^2$ was area averaged over the region of interest and then the square root is taken. Area averages were computed by weighting by cosine of latitude. The RMSE between the verification and ensemble mean is computed at each grid point and forecast hour:

$$r(h_k, x_i, y_j) = \sqrt{\frac{1}{N_i} \sum_{n=1}^{N_i} [\bar{f}(t_n, h_k, x_i, y_j) - o(t_n, h_k, x_i, y_j)]^2}. \quad (13)$$

To compute the regional RMSE, the mean square error [MSE, $r(h_k, x_i, y_j)^2$] was area averaged over the region of interest weighted by the cosine of latitude and the square root was taken last, in the same manner as the regional spread estimates. Bias-corrected RMSE was computed in an identical manner, but using the bias-corrected ensemble mean. The bias was estimated as the difference between the verification and the ensemble averaged over all initial dates and ensemble members:

$$b(h_k, x_i, y_j) = \frac{1}{N_i} \sum_{n=1}^{N_i} \frac{1}{N_p - 1} \sum_{p=1}^{N_p} [f_p(t_n, h_k, x_i, y_j) - o(t_n, h_k, x_i, y_j)], \quad (14)$$

the bias-corrected ensemble $f'_p(t_n, h_k, x_i, y_j)$ is then given by

$$f'_p(t_n, h_k, x_i, y_j) = f_p(t_n, h_k, x_i, y_j) - b(h_k, x_i, y_j). \quad (15)$$

Statistical significance of the differences in spread or RMSE is based on estimating the uncertainty by

randomly removing 5% of the grid points 500 times and computing the lower and upper 2.5th percentiles of the resulting spread or RMSE distribution. This gives an estimate of the 95% confidence intervals for the control and all experiments. Different values for number of bootstrap samples and the amount of data removed were tested and only small differences were found. The difference in spread or RMSE is deemed to be significant if there is no overlap between the confidence interval of the control and any given experiment.

Rank histograms (Hammill 2001) for raw and bias-corrected ensembles were generated to assess ensemble spread and test for outliers. For an ensemble with forecast members and the analysis sampled from the same distribution, the rank histogram should be flat [i.e., each rank is equally likely and the fraction of occurrence of each rank is $1/(N_p + 1)$]. In the case presented here, with 20 ensemble members, that corresponds to 0.0475 and the fraction of occurrence of outliers (both on the low and high end) would be equal to 0.095.

The pattern correlation (Wilks 2011) between spread and RMSE patterns was computed to assess how well the spatial patterns of spread and RMSE match. The pattern correlation is the linear correlation between the cosine-weighted RMSE and spread pattern at a given forecast hour.

These metrics were evaluated for near-surface variables T_{2m} and U_{10m} . ERA-Interim (ERA-I; Dee et al. 2011) reanalysis data at 6-h intervals was used for global verification. ERA-I was used instead of the GEFS analysis so that the verification is independent of the forecast model. While ERA-I is still heavily influenced by its underlying model, T_{2m} observations were assimilated in this reanalysis and where those observations were available ERA-I is constrained by them. We also evaluated the performance over the contiguous United States (CONUS) where a dense network of observed T_{2m} data was available (UCAR 1987). Results show that while the bias differs between the CONUS surface observations and ERA-I, the general error characteristics of the GEFS were the same. For global evaluation, statistics were computed over the Northern Hemisphere land area (20°–60°N), the tropical land area (20°S–20°N), and the Southern Hemisphere land area (60°–20°S).

4. Experiment setup

Each experiment used a 20-member ensemble initialized at 0000 UTC on 11 dates in January and 11 dates in August 2014 (every third day) and integrated out to 5 days at T574. While 5 days is very short compared to time scales of soil moisture variability, the atmospheric

response to methods of stimulating near-surface variability should be visible within a few days. The integration time and ensemble size were chosen based on the computational resources available. August and January were chosen to sample both boreal summer and winter cases. The sample size is 11 dates \times 20 members = 220 cases for each season. This is small, but large enough to get a qualitative estimate of the impact the surface perturbations have on ensemble spread and RMSE. All experiments, including the control, used the same atmospheric initial conditions and atmospheric SPPT, SKEB, and SHUM. The surface initial conditions in the control run were identical for each ensemble member.

The E_{sfcIC} experiment introduces soil temperature and moisture perturbations to the initial state, but keeps all other surface and atmospheric initial conditions and parameters the same as the control. Three experiments use one surface perturbation method individually without perturbing the initial soil state: E_{vfrac} perturbs only the vegetation fraction, E_{alb} perturbs only surface albedo. E_{ssppt} perturbs the soil moisture and temperature tendencies, via the soil SPPT extension, but no soil parameters nor the initial soil state. Finally, E_{all} combines all surface parameter, soil-state and soil-tendency perturbations. Table 3 provides a synthesis of the experiments and which parameters, states and/or tendencies were perturbed. During initial testing we also performed single parameter perturbation experiments for momentum roughness length, the ratio of heat and momentum roughness length, soil hydraulic conductivity and leaf area index, and we performed another experiment that combines albedo, vegetation fraction, momentum roughness length, the ratio of heat and momentum roughness length, soil hydraulic conductivity and leaf area index perturbations (E_{lsmparm}). We do not show results from these experiments as the impact is fairly small compared to the perturbations mentioned above, and they are included in the E_{all} experiment.

In addition to these T574 experiments to 5-day lead time, a smaller set of experiments (control, E_{sfcIC} , E_{all} , and E_{vfrac}) was run at lower resolution (T254) out to 10-day lead time with the same number of ensemble members and the same initial conditions. These forecasts permit evaluation of the spread and error growth of the surface perturbations to longer lead times. Because of limited computational resources, only select experiments were included in the simulations out to 10 days.

5. Results

Figure 3 shows that over the CONUS, the GEFS had slightly smaller RMSE when compared to ERAI than

TABLE 3. Details of the experiments discussed in the text. We also performed experiments for each parameter/ variable separately. For brevity, the results of experiments for perturbing z_0 , z_t/z_0 , SHC, and LAI individually are not shown.

Perturbation	Experiments						Control
	E_{all}	E_{lsmparm}	E_{sfcIC}	E_{alb}	E_{ssppt}	E_{vfrac}	
SMC and STMP IC	×		×				
α	×	×		×			
σ	×	×				×	
z_0	×	×					
z_t/z_0	×	×					
SHC	×	×					
LAI	×	×					
Landsppt	×				×		
SPPT, SKEB, SHUM	×	×	×	×	×	×	×

when compared to observations (Figs. 3a–h). The bias correction made the ensemble appear nearly properly dispersive when compared to ERAI, but less so when compared to observations (Figs. 3a–c, e–g). When introducing an observational and representativeness error of 1 K into the spread calculation following Berner et al. (2015), the ensemble appeared underdispersive only for the first 24 h of the forecast. The increase in spread due to the perturbations was also still within the 1-K observational error estimate. The lower-resolution T254 experiments showed a large increase in spread for both the E_{vfrac} and the E_{sfcIC} experiments for lead times greater than 5 days in winter. For E_{sfcIC} this was related to a large spread increase (≥ 0.5 K) in the central U.S. region, whereas for E_{vfrac} this was more related to a spread increase in the southern/southeastern United States (not shown). Recall that the E_{all} experiment includes all surface perturbations and not only E_{sfcIC} and E_{vfrac} . The E_{sfcIC} perturbations had a larger impact at the initial time for the lower-resolution T254 experiment than at T574, possibly a result of the larger spatial coherence of the soil perturbations at lower resolution. The spread increase due to the E_{all} experiment in the summer was largest around forecast hour 60 and decreased with lead time after that (Figs. 3j, l).

Comparing global land control ensemble spread and RMSE in Fig. 4 for T_{2m} shows that the ensemble was underdispersive near the surface, with spread being 1.0–2.5 K lower than RMSE (Figs. 4a–c, g–i). However, when considering the bias-corrected RMSE, it was clear that a large portion of the difference between spread and RMSE was due to systematic model error; in the tropics (Figs. 4b, h) and the summer hemispheres (Figs. 4c, g) spread and bias-corrected RMSE were comparable.

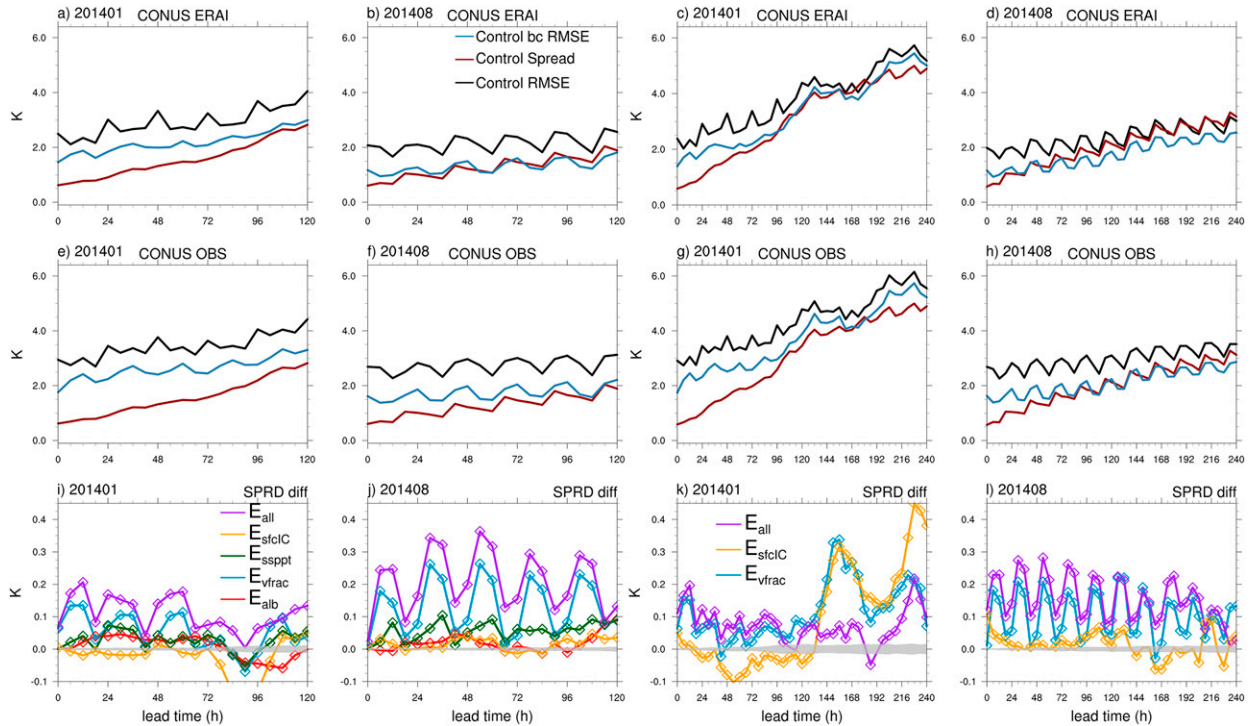


FIG. 3. Ensemble spread and RMSE for T_{2m} averaged over CONUS as a function of lead time for the T574 and T254 experiments. T574 RMSE, bias-corrected RMSE based on ERAI, and spread for (a) January and (b) August 2014, T254 RMSE, bias-corrected RMSE based on ERAI, and spread for (c) January and (d) August 2014. T574 RMSE, bias-corrected RMSE based on T_{2m} observations, and spread for (e) January and (f) August 2014, T254 RMSE, bias-corrected RMSE based on T_{2m} observations, and spread for (g) January and (h) August 2014. Spread difference (experiment – control) for T574 for (i) January and (j) August 2014. Spread difference (experiment – control) for T254 for (k) January and (l) August 2014. Shading indicates the 95% confidence interval around the control spread and diamonds mark where the spread is statistically significantly different from the control.

The impact of the surface perturbations was generally small but largest in the summer hemispheres, when the atmosphere is most responsive to the soil, and in the tropics (Figs. 4e,f,j,k). Overall, albedo, vegetation fraction, and stochastically perturbed soil moisture and temperature tendencies had larger impact on T_{2m} spread over land; the combined perturbations led to a spread increase of up to 0.4 K in the tropics (Figs. 4e,k) and Southern Hemisphere summer (Fig. 4f) and 0.2 K in the Northern Hemisphere summer land areas (Fig. 4j). However, even when compared to the bias-corrected RMSE, the spread of the ensemble was still too small in the winter hemispheres for lead times up to 96 h (Figs. 4d,l). The underdispersive character of the ensemble was less severe at longer lead times.

Figure 5 shows the same quantities for U_{10m} . The difference between spread and RMSE was much smaller than for T_{2m} , and the unadjusted ensemble spread and RMSE were similar in magnitude in the summer hemispheres at lead times ≥ 72 h (Figs. 5c,g). When assuming an observational error of about 1 m s^{-1} (Gao et al. 2015), spread and RMSE were similar in magnitude for almost

all lead times. The spread increases from the surface perturbations were very small, on the order of 0.05 m s^{-1} and only E_{sspt} and E_{all} had a significant impact. This is consistent with the results of Lavaysse et al. (2013) who found that U_{10m} and precipitation are relatively insensitive to most surface perturbations. Koster et al. (2006) also show that temperature is much more controlled by the land surface than precipitation. T_{2m} also showed larger sensitivity than U_{10m} to soil moisture perturbations in the results presented by Tennant and Beare (2014).

The impact of the perturbations on precipitation, surface pressure, precipitable water and surface sensible heat flux was also examined, but no significant change in spread was found for the first three variables, and results are not shown. Area-averaged surface sensible heat flux spread increased about 5 W m^{-2} in the summer hemisphere and the tropics and $1\text{--}2 \text{ W m}^{-2}$ in the winter hemisphere. This corresponds to a relative increase in spread of 25%–30% in the summer hemisphere, 15%–20% in the tropics, and 15%–25% in the winter hemisphere.

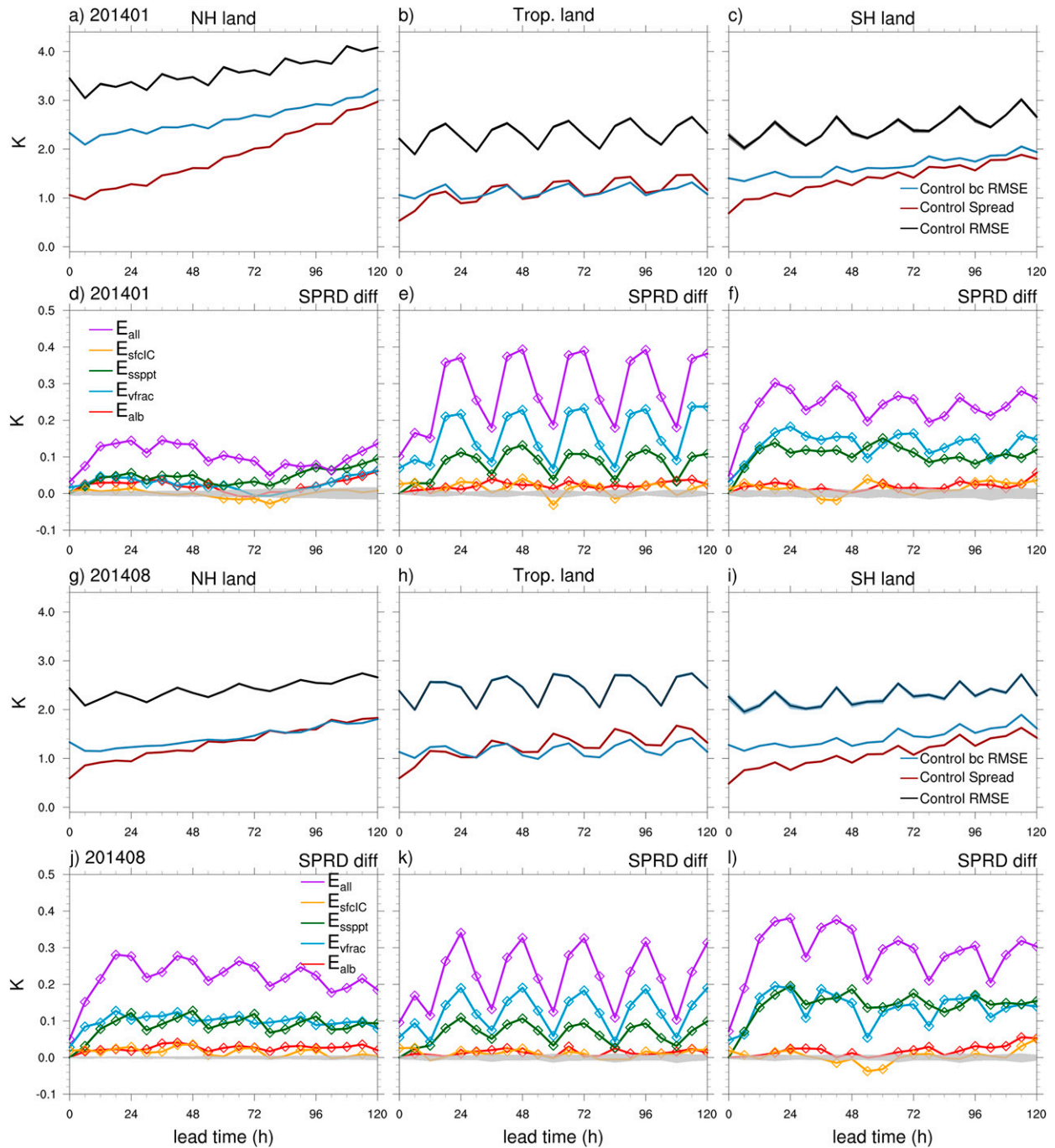


FIG. 4. Ensemble spread and RMSE for T_{2m} averaged over global land areas (excluding Antarctica and Greenland) as a function of lead time. RMSE, bias-corrected RMSE, and spread for (a) Northern Hemisphere January 2014, (b) tropics January 2014, and (c) Southern Hemisphere January 2014; spread difference (experiment – control) for (d) Northern Hemisphere January 2014, (e) tropics January 2014, and (f) Southern Hemisphere January 2014; RMSE, bias-corrected RMSE, and spread for (g) Northern Hemisphere August 2014, (h) tropics August 2014, and (i) Southern Hemisphere August 2014; and spread difference (experiment – control) for (j) Northern Hemisphere August 2014, (k) tropics August 2014, and (l) Southern Hemisphere August 2014. Shading indicates the 95% confidence interval around the control spread and diamonds mark where the spread is statistically significantly different from the control.

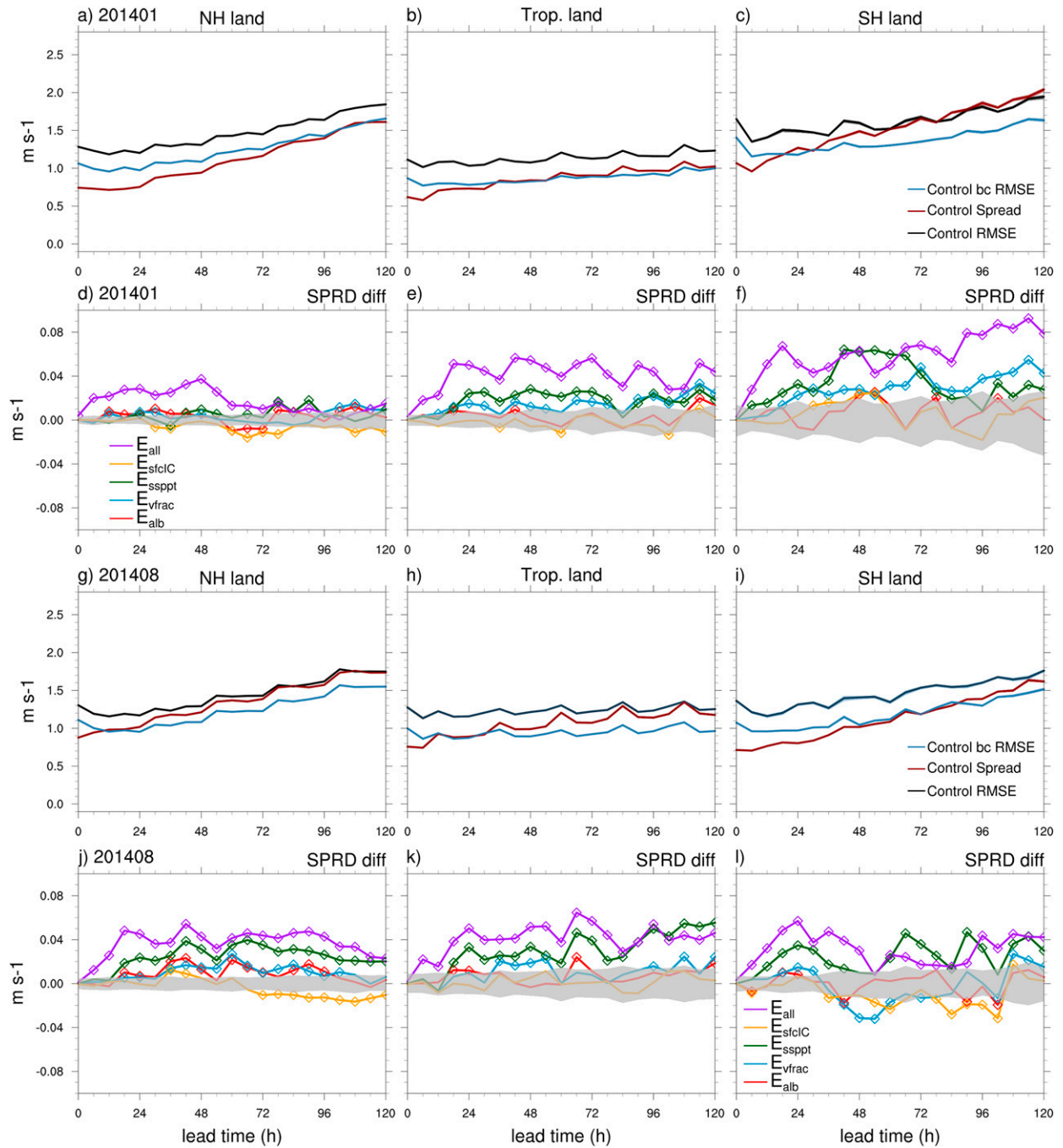


FIG. 5. As in Fig. 4, but for U_{10m} .

Figure 6 shows that at 6-h lead time, the albedo (E_{alb}) induced T_{2m} spread increase followed the local noon (Fig. 6b). The spread increase eventually spanned the whole globe within the first 24 h of the forecast (not shown). Somewhat counterintuitively, the vegetation fraction (E_{vfrac}) T_{2m} spread increase (Fig. 6c) was largest during the evening and night and in regions where the

vegetation fraction has an intermediate value. This may simply be a result of the vegetation fraction perturbations being largest for intermediate values of vegetation fraction. Similar spread increases during evening hours were seen for the soil moisture and temperature perturbations (E_{sfclC} , E_{ssppt}) (Figs. 6d,e), with the largest effects in arid and semiarid regions (see below).

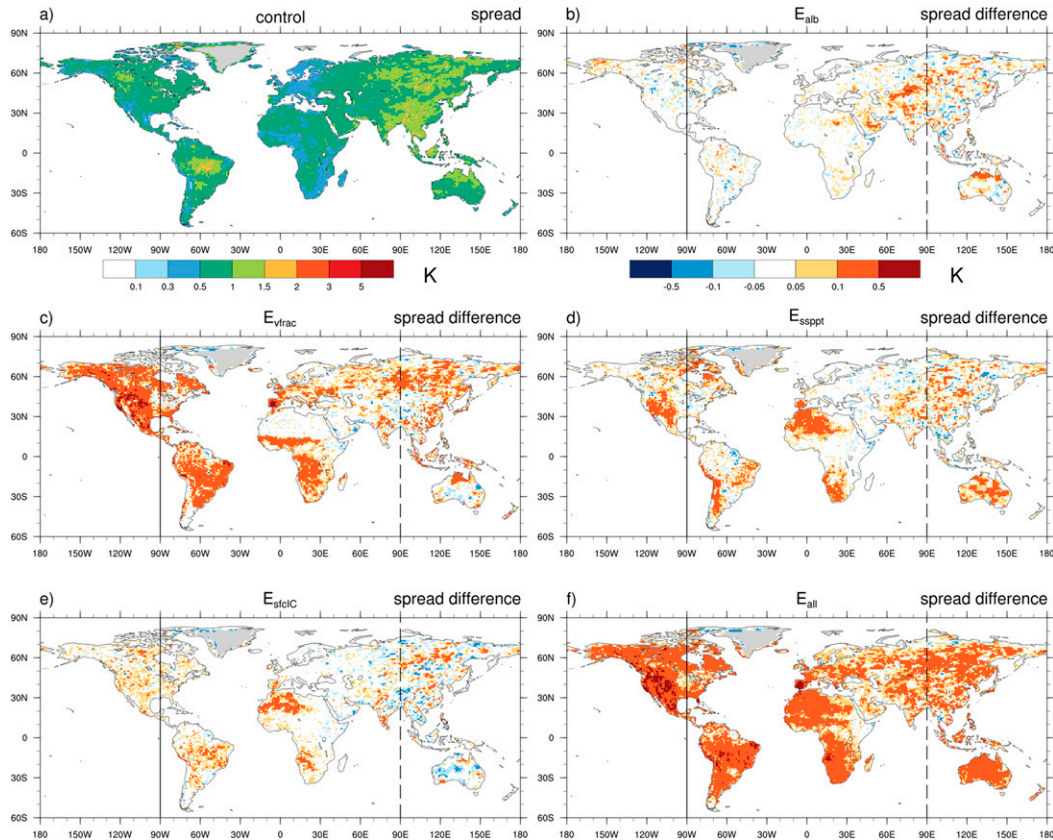


FIG. 6. Ensemble spread of T_{2m} (K) at lead time 6 h for August 2014 initial dates. (a) control ensemble spread, (b) difference between E_{alb} spread and control spread, (c) difference between E_{vfrac} spread and control spread, (d) difference between E_{ssppt} spread and control spread, (e) difference between E_{sfcIC} spread and control spread, and (f) difference between E_{all} spread and control spread. The black solid and dashed vertical lines indicate the locations of local midnight and noon respectively.

The nighttime peak in spread increase could be due to daytime perturbations being mixed through a greater depth with an unstable boundary layer. Considering all surface perturbations together, spread was increased across most land areas by ~ 0.3 K (Fig. 6f) at 6-h lead time. In some regions the spread increase was larger than ~ 0.5 K. Comparing the patterns indicates that those high values were mainly due to the vegetation fraction (E_{vfrac}) perturbations.

To support the claim that the soil moisture and temperature initial condition perturbations tended to increase T_{2m} spread more in arid regions, Figs. 7a, 7b, 7d, and 7e show the global land spread and spread difference as a function of initial soil moisture percentile. For the control run, with no soil state or parameter perturbations, T_{2m} spread initially was about the same across all percentiles of initial soil moisture. By day 3 of the forecast, the largest spread was associated with the 25th–95th percentiles. This is a consequence of the T_{2m} spread being largest at high to medium soil moisture percentiles

in the Northern Hemisphere and largest at low to medium percentiles in the tropics (not shown). The E_{sfcIC} experiment spread differences (Fig. 7d) increased the most for lower soil moisture percentiles with a distinct diurnal cycle. The diurnal cycle is related to a nighttime maximum of spread increase over the Sahara. As soil moisture and temperature perturbation patterns tend to be negatively correlated (Fig. 2), we speculate that during the night, the boundary layer becomes decoupled from the atmosphere, which together with negative soil moisture (positive temperature) perturbations cooling the surface less and positive soil moisture (negative temperature) perturbations cooling the surface more increases the spread in T_{2m} . Compared to the E_{vfrac} (Fig. 7e) and E_{ssppt} (Fig. 7b) experiments, the spread increase from E_{sfcIC} was small. Sensible heat flux differences due to E_{sfcIC} perturbations between the control and the E_{sfcIC} experiments were on the order of $10\text{--}30$ W m^{-2} over large regions and exceeded 50 W m^{-2} locally (not shown). These were slightly smaller than

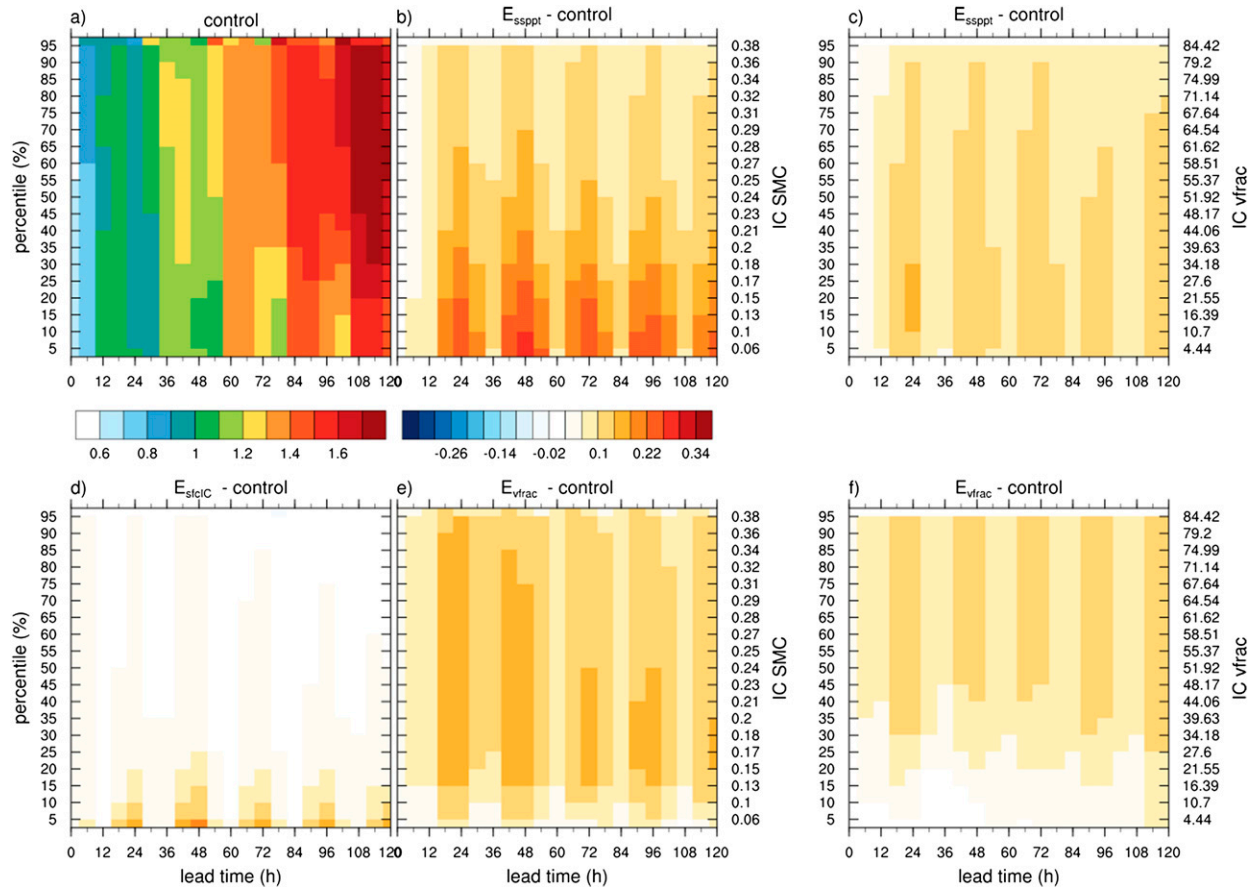


FIG. 7. Global land area T_{2m} spread increase (K) compared to control for August 2014 initial dates as a function of initial soil moisture content percentile for (a) control, (b) E_{ssppt} – control, (d) E_{sfclC} – control, and (e) E_{vfrac} – control experiments. The left y axes show the initial soil moisture percentiles and the right y axes show the corresponding volumetric soil moisture content. As a function of initial vegetation fraction for (c) E_{ssppt} – control and (f) E_{vfrac} – control experiments. The left y axes show the initial vegetation fraction percentiles and the right y axes show the corresponding vegetation fraction cover in %.

those presented by Sutton et al. (2006) and comparable to those presented by Chen et al. (1996). For E_{vfrac} the spread increase was largest between the 20th–80th percentiles of initial soil moisture. It was also largest in the 40th–90th percentile range of initial vegetation fraction (Figs. 7c,f). Recall that the vegetation fraction perturbations were largest for intermediate values of vegetation fraction. For E_{ssppt} (Fig. 7b), the spread increase pattern and diurnal cycle was very similar to E_{sfclC} , with the most spread increase for low soil moisture percentiles and with a diurnal cycle maximum at 0000 UTC and minimum at 1200 UTC, but much larger amplitude than E_{sfclC} .

Figure 8 shows the T_{2m} RMSE patterns and differences between control and the experiments at forecast hour 6. The differences locally were as large as 0.1–0.3 K, but the global average was on the order of +0.02 K. For most forecast hours the difference in global RMSE between the control and the experiments was statistically

insignificant, but for U_{10m} in January the differences for E_{ssppt} , E_{vfrac} , and E_{alb} were as large as 0.04 m s^{-1} and are significant. This means those experiments made the RMSE slightly worse compared to the control. Note that the increase in RMSE, although statistically significant, was small compared to the control RMSE of $2\text{--}3 \text{ m s}^{-1}$. The spatial patterns of the differences show only small-scale variability (Figs. 8b–d), except for the E_{sfclC} perturbations (Fig. 8e) where the difference shows a distinct large-scale spatial pattern consistent with the scale of soil-state perturbations.

A comparison of the RMSE, spread and bias-corrected RMSE patterns for T_{2m} at forecast hours 6 and 120 is given in Fig. 9 for the control and the E_{all} experiment. The raw and bias-corrected RMSE are shown only for the control (Figs. 9a,b,e,f), since the patterns are almost indistinguishable between control and E_{all} . Experiment E_{all} increased spread compared to the control, especially over central South America, southern Africa,

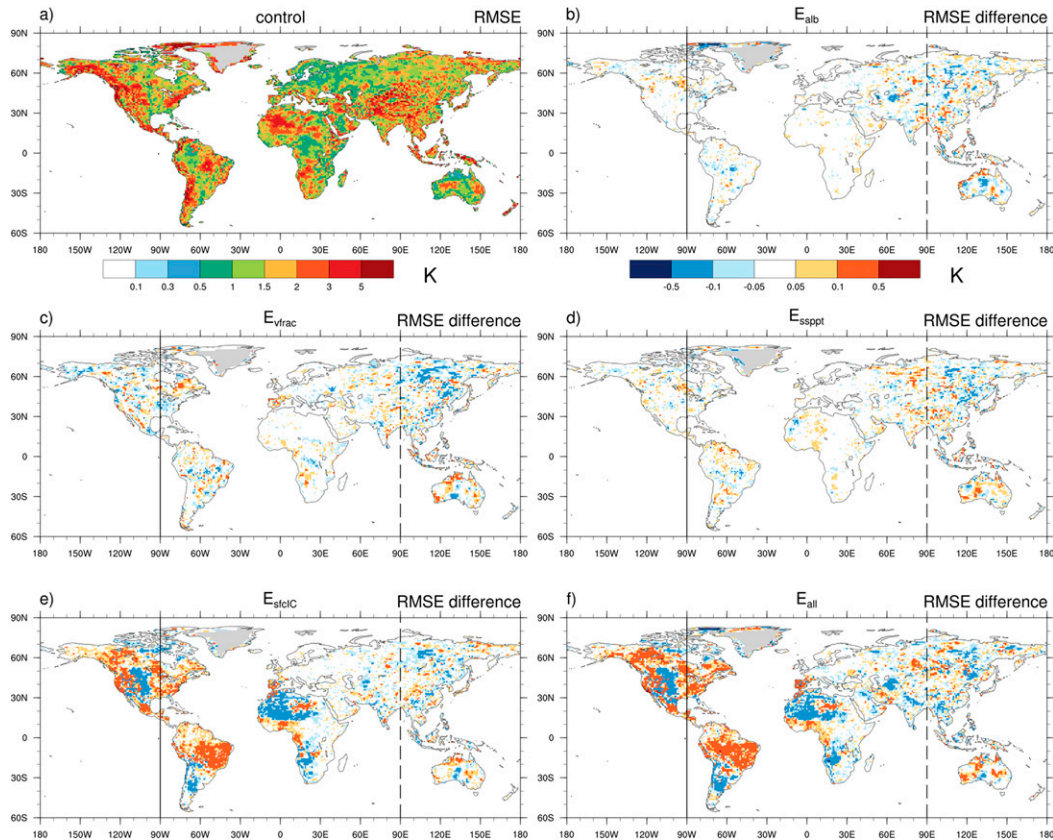


FIG. 8. RMSE of T_{2m} (K) at lead time 6 h for August 2014 initial dates. (a) control RMSE, (b) difference between E_{alb} RMSE and control RMSE, (c) difference between E_{vfrac} RMSE and control RMSE, (d) difference between E_{ssppt} RMSE and control RMSE, (e) difference between E_{sfcic} RMSE and control RMSE, and (f) difference between E_{all} RMSE and control RMSE.

and western North America at forecast hour 6 (Figs. 9c,d). By forecast hour 120 (Figs. 9g,h) the spread increase in E_{all} compared to the control was less localized and is visible across all land areas. In an ideally constructed ensemble, the spread pattern should match the RMSE pattern. While the spread amplitude was too small at most locations, it is hard to determine from the maps whether the patterns match closely or not; differences could in part be due to limited ensemble size and an insufficient number of cases.

To quantify how well the global patterns of spread and RMSE agree, the pattern correlation (spatial correlation) of the two quantities was computed. We note that this global metric may somewhat exaggerate spread-skill relationships for reasons explained in (Hamill and Juras 2006). Figure 10 shows that the pattern correlation over land areas (excluding Antarctica and Greenland) is about 0.2 larger when using the bias-corrected RMSE. Surface perturbations have a larger impact on the correlation for the bias-corrected RMSE than for the raw RMSE, especially

for short lead times < 48 h. The bias-corrected correlation curves cluster together, and the raw correlation curves cluster together after a lead time of 72 h. The correlation between spread and RMSE patterns increases with forecast hour, mainly due to the spread pattern changing and becoming more similar to the RMSE pattern rather than the RMSE pattern changing (not shown). Whitaker and Loughe (1998) also found that Northern Hemisphere spread-error correlations peak in the medium range and are associated primarily with the growth of synoptic-scale perturbations associated with baroclinic wave growth. The pattern correlation curves for the T254 experiments (Figs. 10c,d) flatten out around forecast hour 120 for the bias-corrected RMSE and saturate at around 0.9. Interestingly, the global pattern correlation for T254 in the August cases is much higher than for the T574 cases. The surface perturbations decreased pattern correlations during the first two forecast days in boreal winter. This effect was stronger for the bias-corrected RMSE and could be a consequence of using a perturbation methodology divorced

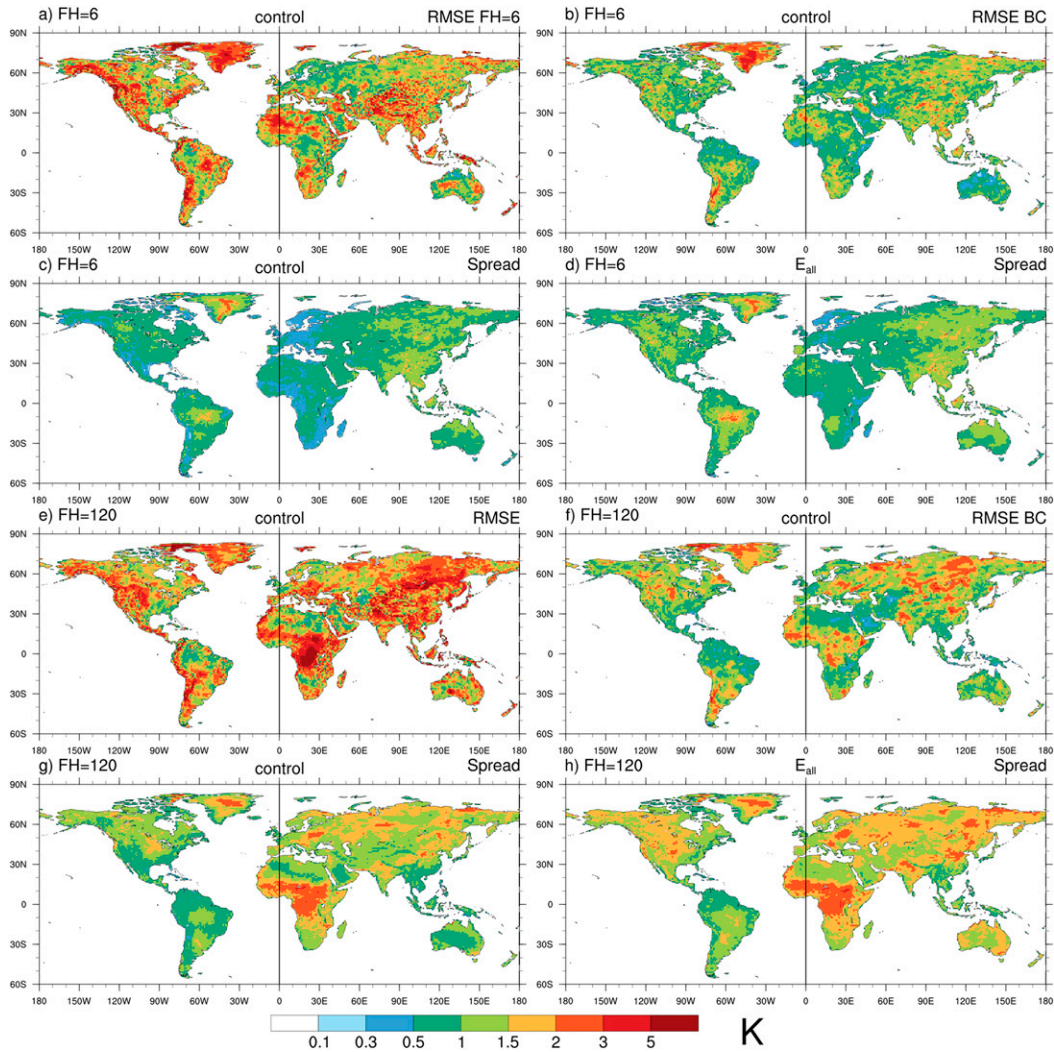


FIG. 9. Comparison of control and $E_{all} T_{2m}$ (K) for August 2014 initial dates. Lead time of 6 h for (a) control RMSE, (b) control bias-corrected RMSE, (c) control spread, and (d) E_{all} spread, and a lead time of 120 h for (e) control RMSE, (f) control bias-corrected RMSE, (g) control spread, and (h) E_{all} spread.

from data assimilation, which would ensure perturbations were consistent with observation uncertainty and with background error (Hamill et al. 2002).

Rank histograms are another way to assess the reliability of the ensemble. If the ensemble samples forecast uncertainty, a perfect observation would be equally likely to be at any rank in the ensemble, when averaged over enough realizations. Figures 11a and 11c show that at forecast hour 6 both the raw and the bias-corrected ensemble have U-shaped rank histograms for the control and for all experiments. There are only minor differences visible between the control and the perturbation experiments, with the E_{all} experiment exhibiting the smallest fraction of outliers. Even in that case the fraction of occurrence of outliers (the sum of the first and last rank) in the raw ensemble

was 0.4 instead of the expected 0.095. The fraction of occurrence of outliers was roughly halved when generating rank histograms with the bias-corrected ensemble. By forecast hour 120 (Figs. 11b,d), as the spread of the ensemble increased, the bias-corrected ensemble was closer to a flat rank distribution, especially for the E_{all} experiment, indicating that the bias-corrected ensemble has more realistic sampling properties. The raw ensemble however still had a U-shaped rank distribution at this lead time.

6. Conclusions

The goal of this work was to test the hypothesis that land surface state and parameter perturbations would improve consistency between spread and RMSE for key

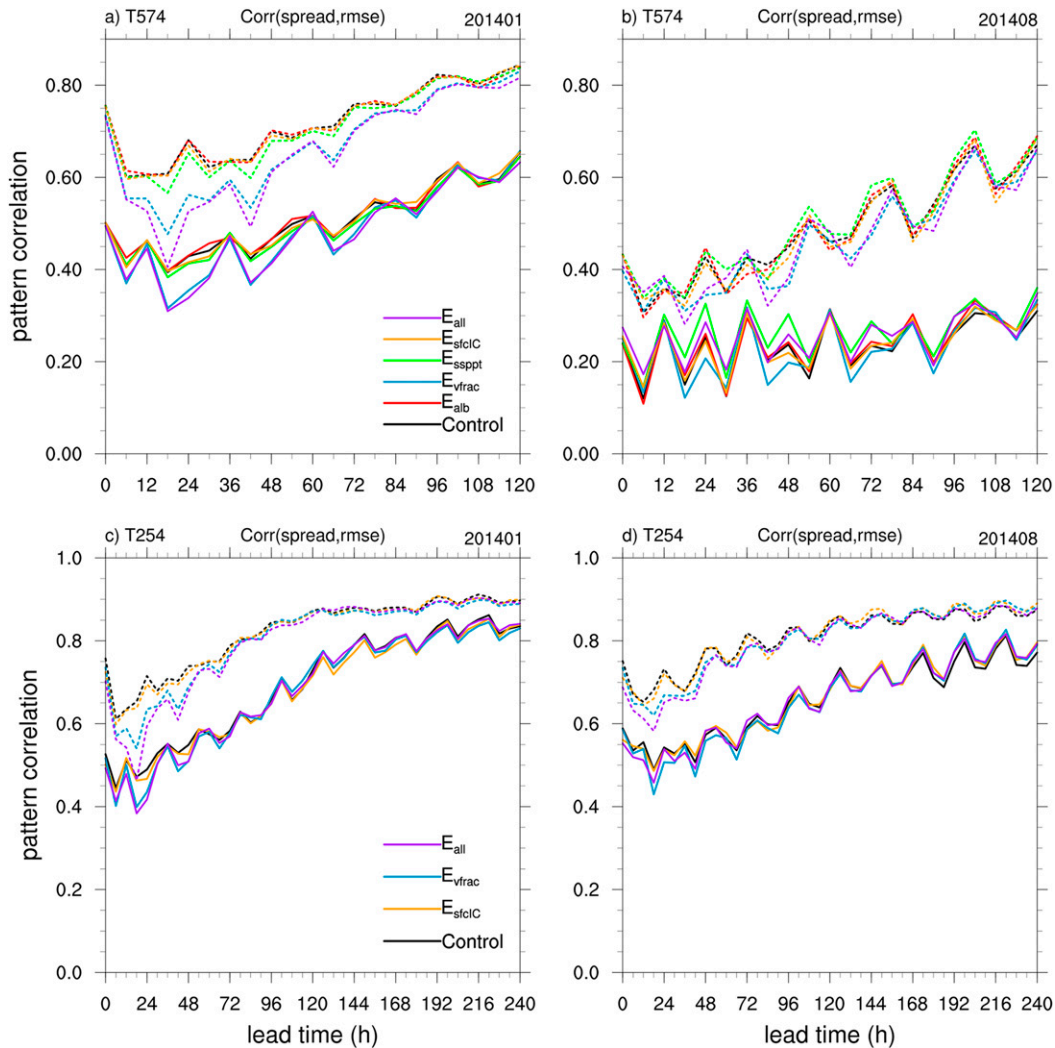


FIG. 10. Pattern correlation between T_{2m} global land area RMSE and spread patterns at all lead times. (a) T574 pattern correlation for all initial dates in January 2014 for RMSE and spread (solid) and bias-corrected RMSE and spread (dotted). (b) As in (a), but for initial dates in August 2014. (c) T254 pattern correlation for all initial dates in January 2014 for RMSE and spread (solid) and bias-corrected RMSE and spread (dotted) to 10 days lead for a subset of experiments. (d) As in (c), but for initial dates in August 2014. All land data points excluding Antarctica and Greenland are used to compute the pattern correlation.

surface variables such as T_{2m} in the GEFS. The surface perturbation magnitudes and spatial patterns were informed by previous research on state and parameter uncertainty and sensitivity where possible. Perturbations were applied to momentum roughness length, the ratio of heat and momentum roughness lengths, soil hydraulic conductivity, leaf area index, surface albedo, vegetation fraction, soil moisture and temperature tendencies, and the initial conditions of soil moisture and temperature. Collectively, these perturbation approaches had a small but positive impact on the spread of T_{2m} and U_{10m} . The GEFS showed the largest sensitivity to soil moisture and temperature tendency,

vegetation fraction, and albedo perturbations. The surface perturbations introduced here did not include snow cover and amount perturbations, and the lack of increased spread in winter could be due in part to this.

Spread increase from soil initial condition perturbations (E_{sfcIC}) was small. Soil-state perturbations were produced from a weighted, random linear combination of EOFs of soil-state uncertainty estimated from normalized anomalies between two soil analyses, an admittedly more ad hoc approach than through, say, ensemble-based data assimilation. While sensible heat flux differences due to E_{sfcIC} were comparable to other studies, the impact on T_{2m} spread of the E_{sfcIC}

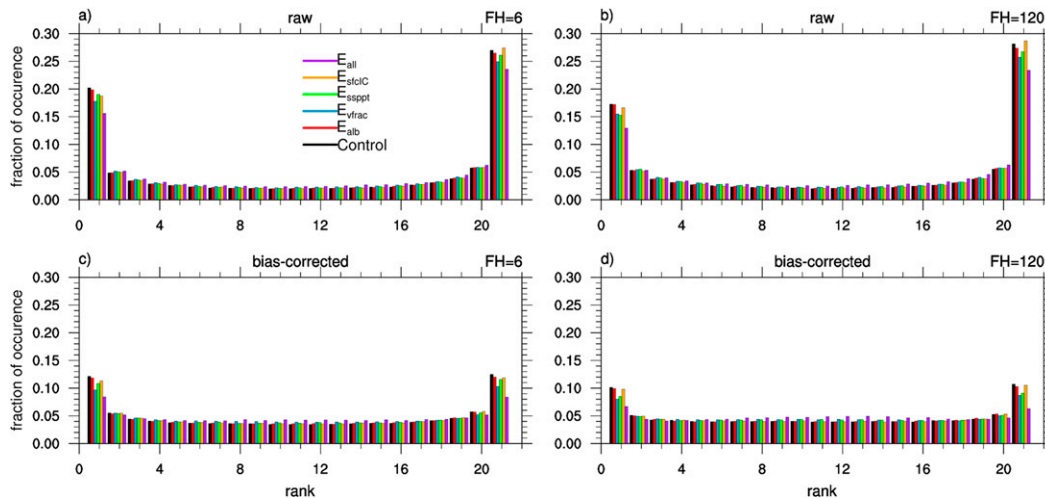


FIG. 11. Rank histogram of T_{2m} for August 2014 initial dates for the raw ensemble (a) at lead time of 6 h and (b) at lead time of 120 h, and the bias-corrected ensemble (c) at lead time of 6 h and (d) at lead time of 120 h. Rank histograms are based on all land data points excluding Antarctica and Greenland.

perturbations was very small, typically 0.1 – 0.3 K and localized compared to other surface perturbations. Possible explanations include that the perturbations were unrealistically small. Additionally, [Koster et al. \(2006\)](#) and [Zhang et al. \(2011\)](#) showed for previous versions of the Global Forecast System that soil perturbations did not impact the atmosphere as much as in other atmospheric models (i.e., the soil and atmosphere may be too weakly coupled in this version of the GEFS).

The experiments with the largest impact on ensemble spread (E_{vfrac} , E_{sspt} , E_{all}) had a lower bias-corrected pattern correlation at 24-h lead time than the control, but this difference disappeared at longer lead times. Pattern correlation increased with lead time for all experiments.

Similar to the pattern correlations, the rank histograms indicated substantial underdispersion but became slightly flatter as lead time increased. As with the pattern correlation, the bias correction of the ensemble had a much larger impact on rank histogram improvement than surface perturbations, although the surface perturbations also improved the distribution.

While the increase in spread was small, collectively the surface perturbations did increase T_{2m} spread to match bias-corrected RMSE in the tropics and the summer hemisphere during the first few days of the forecast. The difference between the bias-corrected RMSE and raw RMSE was larger than the difference between bias-corrected RMSE and spread. This indicates that the ensemble spread being smaller than RMSE near the surface was not entirely due to insufficient spread in the ensemble. Rather, model bias contributed equally or more to

the lack of calibration of the ensemble near the surface as did the lack of spread. This bias could be a result of biased initial soil states, imperfect LSMs, and/ or biased forcings (downward solar and longwave radiation) used by the LSM. There is evidence that the GEFS is not alone in having a large bias compared to the uncertainty introduced through perturbation strategies. [Bernier et al. \(2015\)](#) showed that bias correcting a WRF ensemble has a similar impact on T_{2m} skill as introducing a combination of stochastic perturbation schemes.

The perturbations presented here likely did not account for the entire parameter uncertainty in the model. The perturbation sizes were chosen to be conservative even within the uncertainty estimates that were found. The different perturbations are likely not additive. The ratio of the sum of the spread increase of all single perturbation experiments to the spread increase due to the E_{all} experiment ranges from 0.5 to 1.5 for hemispheric averages. There are nonlinear interactions between the perturbations that can lead E_{all} spread to be much larger or much smaller than the sum of spread of the individual experiments. In conclusion, it remains possible that perturbing even more parameters or perturbing the tested ones with larger amplitudes would lead to more spread increase.

Overall, it appears that land surface temperature bias in the GEFS is a much larger contributor to the spread deficiency than the lack of perturbations to land surface states, parameters and tendencies. These results suggest that while the surface perturbations introduced here are useful and address important aspects of model uncertainty, addressing model bias is equally important and should be a priority for model developers.

Acknowledgments. This research was supported by NGGPS project NA15OAR4320137 and NWS Sandy Supplemental Grant NA14NWS4830003. We would like to acknowledge the contribution to the soil SPPT scheme development by Yuejian Zhu, Xiaqiong Zhou, and Dingchen Hou. We thank Jesse Meng for many discussions of the Noah LSM. We would also like to thank Judith Berner and two anonymous reviewers for their thorough and detailed reviews of this manuscript.

REFERENCES

- Beljaars, A. C. M., and A. A. M. Holtslag, 1991: Flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteor.*, **30**, 327–341, [https://doi.org/10.1175/1520-0450\(1991\)030<0327:FPOLSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2).
- Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF Ensemble Prediction System. *J. Atmos. Sci.*, **66**, 603–626, <https://doi.org/10.1175/2008JAS2677.1>.
- , K. Fossell, S.-Y. Ha, J. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, <https://doi.org/10.1175/MWR-D-14-00091.1>.
- Betts, A. K., and A. C. M. Beljaars, 1993: Estimation of effective roughness length for heat and momentum from FIFE data. *Atmos. Res.*, **30**, 251–261, [https://doi.org/10.1016/0169-8095\(93\)90027-L](https://doi.org/10.1016/0169-8095(93)90027-L).
- Bonan, G. B., K. W. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R. E. Dickinson, and Z.-L. Yang, 2002: The land surface climatology of the community land model coupled to the NCAR Community Climate Model. *J. Climate*, **15**, 3123–3149, [https://doi.org/10.1175/1520-0442\(2002\)015<3123:TLSCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3123:TLSCOT>2.0.CO;2).
- Bonavita, M., L. Isaksen, and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **138**, 1540–1559, <https://doi.org/10.1002/qj.1899>.
- Bowler, N. E., 2008: Accounting for the effect of observation errors on verification of MOGREPS. *Meteor. Appl.*, **15**, 199–205, <https://doi.org/10.1002/met.64>.
- , A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, <https://doi.org/10.1002/qj.234>.
- Bréda, N. J. J., 2003: Ground-based measurements of leaf area index: A review of methods, instruments and current controversies. *J. Exp. Bot.*, **54**, 2403–2417, <https://doi.org/10.1093/jxb/erg263>.
- Broxton, P., X. Zeng, W. Scheffic, and P. Troch, 2014: A MODIS-based 1 km maximum green vegetation fraction dataset. *J. Appl. Meteor. Climatol.*, **53**, 1996–2004, <https://doi.org/10.1175/JAMC-D-13-0356.1>.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- , J. Barkmeijer, T. N. Palmer, and D. S. Richardson, 2000: Current status and future developments of the ECMWF Ensemble Prediction System. *Meteor. Appl.*, **7**, 163–175, <https://doi.org/10.1017/S1350482700001456>.
- Campana, K., and P. Caplan, 2005: Technical procedures bulletin for the T382 Global Forecast System. Tech. Rep., NOAA/NWS/NCEP Environmental Modeling Center, accessed 2 February 2018, http://www.emc.ncep.noaa.gov/gc_wmb/Documentation/TPBOct05/T382.TPB.FINAL.htm.
- Chen, F., and Coauthors, 1996: Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101**, 7251–7268, <https://doi.org/10.1029/95JD02165>.
- Ciach, G. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.*, **22**, 585–595, [https://doi.org/10.1016/S0309-1708\(98\)00043-8](https://doi.org/10.1016/S0309-1708(98)00043-8).
- Clapp, R. B., and G. M. Hornberger, 1978: Empirical equations for some soil hydraulic properties. *Water Resour. Res.*, **14**, 601–604, <https://doi.org/10.1029/WR014i004p00601>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Diak, G., S. Heikkinen, and J. Rates, 1986: The influence of variations in surface treatment on 24-hour forecasts with a limited area model, including a comparison of modeled and satellite-measured surface temperatures. *Mon. Wea. Rev.*, **114**, 215–232, [https://doi.org/10.1175/1520-0493\(1986\)114<0215:TIOVIS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<0215:TIOVIS>2.0.CO;2).
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. 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>.
- Evensen, G., 2009: *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, 307 pp., <https://doi.org/10.1007/978-3-642-03711-5>.
- Fortin, V., M. Abaza, F. Anctil, and R. Turcotte, 2014: Why should ensemble spread match the RMSE of the ensemble mean? *J. Hydrometeorol.*, **15**, 1708–1713, <https://doi.org/10.1175/JHM-D-14-0008.1>.
- Gao, F., X.-Y. Huang, N. A. Jacobs, and H. Wang, 2015: Assimilation of wind speed and direction observations: Results from real observation experiments. *Tellus*, **67A**, 27132, <https://doi.org/10.3402/tellusa.v67.27132>.
- Grant, I. F., A. J. Prata, and R. P. Cechet, 2000: The impact of the diurnal variation of albedo on the remote sensing of the daily mean albedo of grassland. *J. Appl. Meteor.*, **39**, 231–244, [https://doi.org/10.1175/1520-0450\(2000\)039<0231:TIOTDV>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<0231:TIOTDV>2.0.CO;2).
- Guo, Z., and P. A. Dirmeyer, 2006: Evaluation of the Second Global Soil Wetness Project soil moisture simulations: 1. Intermodel comparison. *J. Geophys. Res.*, **111**, D22S02, <https://doi.org/10.1029/2006JD007233>.
- Hacker, J. P., C. Snyder, S.-Y. Ha, and M. Pocerlich, 2011: Linear and non-linear response to parameter variations in a mesoscale model. *Tellus*, **63A**, 429–444, <https://doi.org/10.1111/j.1600-0870.2010.00505.x>.
- Hamill, T. M., 1997: Short-range ensemble forecasting using the ETA/RSM forecast models. Ph.D. thesis, Cornell University, 432 pp.
- , 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- , 2006: Ensemble-based data assimilation. *Predictability of Weather and Climate*, T. N. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.

- , and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, [https://doi.org/10.1175/1520-0493\(1997\)125<1312:VOERSR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2).
- , and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, <https://doi.org/10.1256/qj.06.25>.
- , and J. S. Whitaker, 2007: Ensemble calibration of 500-hPa geopotential height and 850-hPa and 2-m temperatures using reforecasts. *Mon. Wea. Rev.*, **135**, 3273–3280, <https://doi.org/10.1175/MWR3468.1>.
- , C. Snyder, and R. E. Morss, 2002: Analysis-error statistics of a quasigeostrophic model using three-dimensional variational assimilation. *Mon. Wea. Rev.*, **130**, 2777–2790, [https://doi.org/10.1175/1520-0493\(2002\)130<2777:AESOAQ>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2777:AESOAQ>2.0.CO;2).
- , J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688, <https://doi.org/10.1175/2010MWR3456.1>.
- Houtekamer, P. L., 1993: Global and local skill forecasts. *Mon. Wea. Rev.*, **121**, 1834–1846, [https://doi.org/10.1175/1520-0493\(1993\)121<1834:GALSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<1834:GALSF>2.0.CO;2).
- , and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, <https://doi.org/10.1175/MWR-D-15-0440.1>.
- , X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1143–1162, <https://doi.org/10.1175/MWR-D-13-00138.1>.
- Koster, R. D., and M. J. Suarez, 1996: Energy and water balance calculations in the Mosaic LSM. Tech. Rep. 104606, NASA, 58 pp.
- , and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeorol.*, **7**, 590–610, <https://doi.org/10.1175/JHM510.1>.
- , Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, <https://doi.org/10.1175/2009JCLI2832.1>.
- Krzysztofowicz, R., 1997: Transformation and normalization of variates with specified distributions. *J. Hydrol.*, **197**, 286–292, [https://doi.org/10.1016/S0022-1694\(96\)03276-3](https://doi.org/10.1016/S0022-1694(96)03276-3).
- 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>.
- Masson, V., J.-L. Champeaux, F. Chauvin, C. Meriguet, and R. Lacaze, 2003: A global database of land surface parameters at 1-km resolution in meteorological and climate models. *J. Climate*, **16**, 1261–1282, <https://doi.org/10.1175/1520-0442-16.9.1261>.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433, [https://doi.org/10.1175/1520-0493\(2000\)128<0416:AAEFK>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0416:AAEFK>2.0.CO;2).
- Mullen, S. L., and R. Buizza, 2001: Quantitative precipitation forecasts over the United States by the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **129**, 638–663, [https://doi.org/10.1175/1520-0493\(2001\)129<0638:QPFOTU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0638:QPFOTU>2.0.CO;2).
- Palmer, T., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. Tech. Memo. 598, European Centre for Medium-Range Weather Forecasts, 44 pp., <https://doi.org/10.21957/ps8gbwbdv>.
- Pitman, A. J., 1994: Assessing the sensitivity of a land-surface scheme to the parameter values using a single column model. *J. Climate*, **7**, 1856–1869, [https://doi.org/10.1175/1520-0442\(1994\)007<1856:ATSOAL>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1856:ATSOAL>2.0.CO;2).
- Qu, X., and A. Hall, 2005: Surface contribution to planetary albedo variability in cryosphere regions. *J. Climate*, **18**, 5239–5252, <https://doi.org/10.1175/JCLI3555.1>.
- Reichle, R., J. Walker, R. Koster, and P. Houser, 2002: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.*, **3**, 728–740, [https://doi.org/10.1175/1525-7541\(2002\)003<0728:EVEKFF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2).
- Reynolds, C. A., J. A. Ridout, and J. G. Mclay, 2011: Examination of parameter variations in the U. S. Navy Global Ensemble. *Tellus*, **63A**, 841–857, <https://doi.org/10.1111/j.1600-0870.2011.00532.x>.
- Ries, H., K. H. Schlünzen, B. Brümmner, M. Claussen, and G. Müller, 2010: Impact of surface parameter uncertainties on the development of a trough in the Fram Strait region. *Tellus*, **62A**, 377–392, <https://doi.org/10.1111/j.1600-0870.2010.00451.x>.
- Rodell, M., and H. K. Beaudoin, 2007: GLDAS_CLM10SUBP_3H: GLDAS CLM land surface model L4 3 hourly 1.0 × 1.0 degree subsetted V001. Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, accessed 1 August 2016, <https://doi.org/10.5067/83NO2QDLG6M0>.
- , and —, 2015: GLDAS Noah land surface model L4 3 hourly 1.0 × 1.0 degree V2.0. Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, accessed 1 August 2016, <https://doi.org/10.5067/L0JGCNVBNRAX>.
- , and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Schaake, J. C., and Coauthors, 2004: An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS). *J. Geophys. Res.*, **109**, D01S90, <https://doi.org/10.1029/2002JD003309>.
- Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, <https://doi.org/10.1256/qj.04.106>.
- Sutton, C., T. M. Hamill, and T. T. Warner, 2006: Will perturbing soil moisture improve warm-season ensemble forecasts? A proof of concept. *Mon. Wea. Rev.*, **134**, 3174–3189, <https://doi.org/10.1175/MWR3248.1>.
- Tennant, W., and S. Beare, 2014: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quart. J. Roy. Meteor. Soc.*, **140**, 1150–1160, <https://doi.org/10.1002/qj.2202>.
- Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113**, D18101, <https://doi.org/10.1029/2007JD009284>.
- UCAR, 1987: TDL U.S. and Canada surface hourly observations. Research Data Archive, National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO, accessed 22 February 2017, <http://rda.ucar.edu/datasets/ds472.0/>.
- von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Wang, J., J. Chen, J. Du, Y. Zhang, Y. Xia, and G. Deng, 2018: Sensitivity of ensemble forecast verification to model bias.

- Mon. Wea. Rev.*, **146**, 781–796, <https://doi.org/10.1175/MWR-D-17-0223.1>.
- Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, <https://doi.org/10.1175/MWR-D-12-00141.1>.
- Whitaker, J. S., and A. F. Loughé, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292–3302, [https://doi.org/10.1175/1520-0493\(1998\)126<3292:TRBESA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2).
- Wilks, D., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.
- Zhang, D., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer-sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594–1609, [https://doi.org/10.1175/1520-0450\(1982\)021<1594:AHRMOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1982)021<1594:AHRMOT>2.0.CO;2).
- Zhang, L., P. A. Dirmeyer, J. Wei, Z. Guo, and C.-H. Lu, 2011: Land–atmosphere coupling strength in the Global Forecast System. *J. Hydrometeorol.*, **12**, 147–156, <https://doi.org/10.1175/2010JHM1319.1>.
- Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.