Predicting Rapid Changes in Evaporative Stress Index (ESI) and Soil Moisture Anomalies over the Continental United States

DAVID J. LORENZ,[a] JASON A. OTKIN,[b] BENJAMIN ZAITCHIK,[c] CHRISTOPHER HAIN,[d] AND MARTHA C. ANDERSON[e]

[a] *Center for Climatic Research, University of Wisconsin–Madison, Madison, Wisconsin*
[b] *Space Science and Engineering Center, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin*
[c] *Department of Earth and Planetary Sciences, The Johns Hopkins University, Baltimore, Maryland*
[d] *Earth Science Office, NASA Marshall Space Flight Center, Huntsville, Alabama*
[e] *Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland*

(Manuscript received 4 December 2020, in final form 30 August 2021)

ABSTRACT: Probabilistic forecasts of changes in soil moisture and an evaporative stress index (ESI) on subseasonal time scales over the contiguous United States are developed. The forecasts use the current land surface conditions and numerical weather prediction forecasts from the Subseasonal to Seasonal (S2S) Prediction project. Changes in soil moisture are quite predictable 8–14 days in advance with 50% or more of the variance explained over the majority of the contiguous United States; however, changes in ESI are significantly less predictable. A simple red noise model of predictability shows that the spatial variations in forecast skill are primarily a result of variations in the autocorrelation, or persistence, of the predicted variable, especially for the ESI. The difference in overall skill between soil moisture and ESI, on the other hand, is due to the greater soil moisture predictability by the numerical model forecasts. As the forecast lead time increases from 8–14 to 15–28 days, however, the autocorrelation dominates the soil moisture and ESI differences as well. An analysis of modeled transpiration, and bare soil and canopy water evaporation contributions to total evaporation, suggests improvements to the ESI forecasts can be achieved by estimating the relative contributions of these components to the initial ESI state. The importance of probabilistic forecasts for reproducing the correct probability of anomaly intensification is also shown.

KEYWORDS: Probability forecasts/models/distribution; Seasonal forecasting; Short-range prediction; Regression

1. Introduction

Recently, Otkin et al. (2015a) and Lorenz et al. (2017a,b) developed probabilistic statistical forecasts of the U.S. Drought Monitor (USDM; Svoboda et al. 2002) using recent anomalies in precipitation, evaporative stress index (ESI) (Anderson et al. 2007b, 2011), and modeled soil moisture. Lorenz et al. (2018) added output from the North American Multi-Model Ensemble (NMME; Kirtman et al. 2014) to the original statistical forecasts to create hybrid statistical/dynamical model forecasts of USDM intensification. These studies, which focused on 2–8-week time scales, were motivated by the prediction of rapidly intensifying droughts. Such "flash" droughts (Svoboda et al. 2002; Hunt et al. 2009, 2014; Ford et al. 2015; Ford and Labosier 2017; Otkin et al. 2013, 2014, 2015a,b, 2016, 2018; Christian et al. 2019b) are particularly challenging to forecast (Pendergrass et al. 2020) because their time scales are too long for the best use of atmosphere initial conditions and too short for best use of external boundary conditions such as sea surface temperatures.

In this study, we extend the above studies to forecasts of ESI and soil moisture over 8–28-day time scales. While not as comprehensive as the USDM, ESI and soil moisture are continuous rather than discrete, categorical variables, and this likely improves their predictability. For example, Lorenz et al. (2017a,b) found that USDM forecasts could be improved by first estimating a continuous USDM that is an "optimal" blend of precipitation, ESI and soil moisture anomalies that is most consistent with the actual categorical USDM. Essentially, by discretizing a continuous process, the USDM loses information about the closeness of adjacent drought categories. Of course, the USDM must be categorical given the subjective nature of the index. ESI and soil moisture, on the other hand, are objective indices, and therefore are likely more consistent across time and space and more amenable to statistical prediction.

Like Lorenz et al. (2018), we use a hybrid statistical/dynamical model forecast where the observed initial state and numerical model output are predictors for a statistical regression. The predictand in this case is the *change* in ESI or soil moisture from the initial state. Previous studies (Nicolai-Shaw et al. 2016; Zhu et al. 2020) have found that the persistence or autocorrelation of land surface variables, such as soil moisture, plays a large role in their predictability. For example, areas with large soil moisture autocorrelation have large predictability resulting from knowledge of the initial condition alone. Since the focus of this work is rapid changes in ESI and soil moisture, we predict the change in ESI and soil moisture. In this case, the above relationships are reversed: small autocorrelation is associated with large predictability due to the initial condition because it implies a greater tendency for anomalies to decay toward climatology.

The paper begins with a description of the land surface data and the numerical model forecasts. Next, a method to quantify the role of autocorrelation in forecast skill is developed. In section 4, the statistical regression methodology is described. In section 5, we 1) describe the forecast results, 2) explore the

Corresponding author: David J. Lorenz, dlorenz@wisc.edu

DOI: 10.1175/JHM-D-20-0289.1

role of autocorrelation in the spatial variability of skill, 3) analyze the contributions of transpiration, bare soil, and canopy evaporation to ESI skill, and 4) discuss the importance of probabilistic forecasts for unbiased estimates of drought intensification. We end with conclusions and ideas for future research.

2. Data

a. ESI

The ESI, an index of moisture stress, is the standardized ratio of evapotranspiration to Penman–Monteith potential evapotranspiration (PET; Allen et al. 1998). Evapotranspiration is estimated from remotely sensed thermal infrared imagery using the Atmospheric Land Exchange Inverse (ALEXI) model (Norman et al. 1995; Anderson et al. 1997, 2007b). Over the United States, this energy balance model uses 1) lower tropospheric temperature profiles, surface meteorological variables and clear-sky solar radiation estimates from the Climate Forecast System Reanalysis dataset (Saha et al. 2010), 2) surface temperature measurements from the Geostationary Operational Environmental Satellite, and 3) an atmospheric boundary layer growth model (McNaughton and Spriggs 1986) to close the total surface energy budget from ~1.5 h after local sunrise until 1.5 h before local noon, thereby estimating sensible, latent, and ground heat fluxes. The resulting evapotranspiration estimates are reasonable for a variety of climate and vegetation types (Anderson et al. 2007a).

To compute the ESI, the mean and standard deviation are not computed over the entire record, but rather vary with the seasonal cycle. The mean seasonal cycle is smoothed using parabolic shaped weights of the form: $w_j = (n + 1)^2 - j^2$, for $j = -n, -n + 1, \ldots, n$. We use $n = 15$, which corresponds to a smoothing "window" of 31 days. The resolution of the ESI is $0.04° \times 0.04°$ in space and daily in time. At a given grid point, many days may be missing due to cloud cover, so composites are used to provide more complete coverage. Most of the results below use 7-day composites, which is a good compromise between smoothing versus temporal resolution. Because the ultimate goal of this research is forecasting rapid drought *intensification*, we forecast the future *change* in the ESI rather than the future value in the ESI. Ultimately, forecasts of changes rather than values contain the same information, so this is not restrictive.

b. NLDAS

For diagnostic purposes, model-based estimates of ESI from the Noah (NOAH, Ek et al. 2003; Barlage et al. 2010; Wei et al. 2013), Mosaic (MOS, Koster and Suarez 1994, 1996), and Variable Infiltration Capacity (VIC, Liang et al. 1996; Bowling and Lettenmaier 2010) models of the North American Land Data Assimilation System (NLDAS2; Mitchell et al. 2004; Xia et al. 2012a,b) are used. The 1) total evapotranspiration, 2) base soil evaporation, 3) canopy water evaporation, and 4) transpiration were taken from each of the three models. The PET was only available for the Noah model, so it was used to normalize the evaporation for all three models. Because all models are driven by the same atmospheric forcing, the Noah PET should be relevant for all models. The Noah PET uses the Penman–Monteith methodology (Chen et al. 1996). The resulting ESI was standardized in the same way as the remotely sensed ESI. The NLDAS ESI fields are used for diagnostic purposes only because there are pronounced model differences in the partitioning of evaporation into these components (e.g., Kumar et al. 2018). We also use the gridded precipitation, dewpoint, and temperature fields from the NLDAS forcing dataset to diagnose numerical model skill. The resolution of the NLDAS is $0.125° \times 0.125°$.

c. SMERGE

SoilMERGE (SMERGE; Tobin et al. 2019) is a "root zone" (0–40-cm depth) soil moisture dataset that combines output from the Noah NLDAS2 model with satellite retrievals of soil moisture from various passive and active C- and L-band microwave sensors from the European Space Agency Climate Change Initiative satellite. SMERGE uses an exponential filter to convert 0–5-cm soil moisture retrievals from the satellite observations to root zone soil moisture content (Tobin et al. 2017). The resolution of the SMERGE is $0.125° \times 0.125°$ over the contiguous United States and daily in time. For the results below, SMERGE anomalies are found by removing the smoothed mean seasonal cycle of SMERGE at each grid point. The mean seasonal cycle is smoothed using the same 31-day weights as ESI. Like the ESI, we forecast the future *change* in SMERGE rather than the future value of SMERGE. Because ESI benefits from weekly composites due to missing values, we also forecast weekly soil moisture composites for consistency.

d. Numerical forecast models

Starting in 2015, the Subseasonal to Seasonal (S2S) Prediction project (Vitart et al. 2017) established an extensive database of subseasonal reforecasts up to 60 days long for 11 atmospheric/coupled climate models. For this project, we use forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) model because it has almost complete daily coverage from 2000 to 2019. Another advantage of the ECMWF model is that the reforecasts are continuous up to present unlike many other models that stop at 2010 (leaving a gap from 2011 until start of the forecast period in 2015). For these reasons, only the ECMWF model is used in this study.

The ECMWF S2S model has 11 ensemble members (including the control) and the forecast length is 46 days with an output frequency of 6–24 h depending upon the variable. We considered the following variables as predictors: precipitation, maximum and minimum daily temperature, mean daily temperature, dewpoint temperature, sensible and latent heat flux, and net shortwave radiation at surface, as well as linear combinations of certain temperature variables. Prior to analysis, all predictors are daily averaged. The soil moisture fields were unusable due to a pronounced temporal inhomogeneity in the data, perhaps due to a land surface model change. In addition, there is an interpolation problem in the ECMWF data on day 16 of the reforecasts coinciding with a change in the model resolution from triangular spectral truncation at wavenumber 639 (T639) before day 16 to T319 at longer lead times. This problem

manifests in the 6-h accumulation fields (precipitation and sensible and latent heat) used to calculate the daily fields. We deal with this issue by ignoring the bad 6-h accumulation and multiplying the remaining 18-h accumulation by 4/3 to roughly correct for the amplitude lost in ignoring one 6-h accumulation. Only the ensemble mean values are used for the results below. The ensemble spread was also considered, but the change in skill was not consistently better.

To remove systematic model biases, the mean forecast value is calculated for each date in the seasonal cycle and for each forecast lead time. The seasonal cycle (for each lead time) is then smoothed using the same 31-day weights used for SMERGE. Next, the mean forecast was subtracted from each forecast to form the forecast anomalies that are used for our statistical prediction.

e. Time period and spatial resolution

The resolution of the archived ECMWF model output is 1.5° × 1.5°. Because this resolution is much coarser than the ESI and SMERGE, we do not interpolate the model data to the fine 0.04° or 0.125° resolutions of the ESI and SMERGE, respectively. Instead, all of the data are interpolated to an intermediate resolution 0.4° × 0.4° grid using bilinear interpolation for the ECMWF model and by averaging the ESI and SMERGE data. The domain is the contiguous United States for SMERGE and the continental United States and the surrounding regions for ESI. The forecasts are for the warm season (1 May–30 September) because flash droughts are most frequent during this time (Christian et al. 2019a). The length of the ESI and ECMWF datasets limit the time period of our forecasts, which are 2000–18 for ESI and 1999–2018 for the SMERGE.

3. Simple model for understanding predictability

Many geophysical time series are approximately red noise;[1] therefore, a forecast based on persistence alone can have significant skill (Nicolai-Shaw et al. 2016; Zhu et al. 2020). As such, a persistence forecast is a good null hypothesis to test whether Numerical model forecasts add value to the statistical forecast. When forecasting red noise, higher autocorrelation means the null hypothesis has more skill. On the contrary, when predicting the change in a red noise time series, lower autocorrelation means higher null hypothesis skill. Quantifying the precise impact of autocorrelation on skill is very important for understanding the results of this study; therefore, in this section we develop a simple model that relates the predictability of a change in a variable, S, to the numerical model skill and autocorrelation.

Simple model

Let an overbar, $\overline{y(t)}$, denote the time average of the variable y, which is a function of time t. Also, it is assumed that the time average has already been subtracted from all variables prior to analysis. In this case, the correlation between variables x and y is $\overline{xy}/\sqrt{\overline{x^2}\times\overline{y^2}}$ and the regression coefficient b that best predicts $y \approx bx$, is given by $b = \overline{xy}/\overline{x^2}$.

Returning to our specific problem, let a be the autocorrelation of the soil moisture S, for example, at lag T. Then, from the definition of autocorrelation:

$$a = \frac{\overline{S(t)S(t+T)}}{\sqrt{\overline{S(t)^2}\times\overline{S(t+T)^2}}} = \frac{\overline{S(t)S(t+T)}}{\overline{S(t)^2}}. \tag{1}$$

For the last equality we used the fact that $\overline{S(t+T)^2} = \overline{S(t)^2}$ for a stationary time series. The last equality is also in the form of the regression coefficient relating $S(t+T)$ and $S(t)$, and therefore:

$$S(t+T) = aS + F(t), \tag{2}$$

where F is the residual, or forcing of S. For soil moisture, F includes any precipitation and evaporation that are not linearly related to S. Next, square both sides of (2) and average:

$$\overline{S(t+T)^2} = a^2\overline{S^2} + 2a\overline{SF} + \overline{F^2}. \tag{3}$$

Because a is the least squares regression coefficient, the forcing F must be uncorrelated with S, and therefore $\overline{SF} = 0$ and

$$\overline{S(t+T)^2} = a^2\overline{S^2} + \overline{F^2}. \tag{4}$$

The first term on the right is the variance of future S explained by the initial condition and the second term is the variance from the forcing, which can potentially be forecast using numerical model output. Let the portion of the F that can be predicted by the numerical model output be F_P, then the total $S(t+T)$ variance that is explained by the initial condition and the numerical model is

$$\overline{S_P(t+T)^2} = a^2\overline{S^2} + \overline{F_P^2}. \tag{5}$$

Dividing (5) by the total variance $\overline{S(t+T)^2}$, using the assumption of stationarity $[\overline{S(t+T)^2} = \overline{S^2}]$, and noting from regression theory that the correlation squared c^2 is the fraction of variance explained:

$$c^2 = a^2 + \frac{\overline{F_P^2}}{\overline{S^2}}. \tag{6}$$

Stationarity can also be applied to (4) to get an expression for the S variance in terms of F:

$$\overline{S^2} = \frac{\overline{F^2}}{1-a^2}. \tag{7}$$

Substituting (7) into (6) and letting $c_F^2 = \overline{F_P^2}/\overline{F^2}$:

$$c^2 = a^2 + (1-a^2)c_F^2. \tag{8}$$

Note, $\overline{F_P^2}/\overline{F^2}$ is the fraction of forcing variance that is explained, and thus it is denoted as a correlation squared, c_F^2. Also, Eq. (8) is exact because F is simply the residual. The first term in (8) represents skill from the initial condition and the

[1] Red noise is a first-order autoregressive process with positive autocorrelation at unit time lag.

second term represents skill from the numerical model forecast. According to (8), large autocorrelation a is associated with large skill regardless of skill in predicting future atmospheric conditions c_F^2. On weekly time scales, soil moisture is persistent enough to be in this large autocorrelation regime.

To recast the above autocorrelation results in terms of the change, $\Delta = S(t + T) - S(t)$, we first find the variance of Δ:

$$\overline{\Delta^2} = \overline{[S(t+T) - S(t)]^2} = 2(1-a)\overline{S^2}, \tag{9}$$

where we use stationarity and the definition of autocorrelation, (1), to simplify the binomial term. Next, (2) is written in terms of Δ instead of $S(t + T)$:

$$\Delta = (a-1)S + F, \tag{10}$$

In analogy with (5), a portion of the variance is explained by the numerical model forecasts:

$$\overline{\Delta_P^2} = (1-a)^2\overline{S^2} + \overline{F_P^2}. \tag{11}$$

Finally dividing by the total Δ variance given by (9) and using (7) to again simplify the term involving F_P, we get

$$c_\Delta^2 = \frac{1-a}{2} + \frac{1+a}{2}c_F^2, \tag{12}$$

where c_Δ^2 is the fraction of variance explained for the change in S. As before, the first term in (12) represents skill from the initial condition and the second term represents skill from the numerical model forecast. Comparing (8) and (12), we see that the initial condition skill is smaller for the change c_Δ^2 when $(1 - a)/2 < a^2$, which simplifies to $(2a - 1)(a + 1) > 0$. Because a is always ≥ -1, the second factor is always positive and therefore the important factor is the first. Therefore, initial condition skill is smaller for c_Δ^2 when $2a - 1 > 0$, or $a > 1/2$.[2] For soil moisture (SMERGE), 81% of the domain has autocorrelation > 0.5 (8–14-day forecast). Therefore, predicting the change rather than the value is more desirable because the "trivial" skill from the initial condition is less emphasized. This further motivates our focus on forecasts of the change rather than the future value. For ESI, the autocorrelation condition is ambiguous as only 54% of the domain has $a > 1/2$.

Equation (12) is very important for understanding the spatial variability in forecast skill and the differences in skill, particularly for the ESI. The first term in (12) is the variance explained by the initial conditions. This term describes a relaxation of the anomaly back to climatology. If S is very persistent (i.e., a is close to 1) then this term is small. If persistence is weak (a is close to 0), then this term alone explains half of the total $S(t + T) - S(t)$ variance. The second term is the skill coming from the numerical model forecasts. If this component is perfect, then c_F^2 and the total skill is perfect ($c_\Delta^2 = 1$). For very persistent S, the effect of numerical model skill on $S(t + T) - S(t)$ skill is larger. For weakly persistent S, the effect of numerical model skill on predictability is weaker.

[2] The identical condition holds for the second term as well.

To help gauge the effect of autocorrelation on the spatial structure of skill, (12) is used to diagnose c_F^2 from the skill c_Δ^2 and the SMERGE autocorrelation a:

$$c_F^2 = \frac{2c_\Delta^2 - 1 + a}{1 + a}. \tag{13}$$

Next, the spatial mean c_F^2 is used at all grid points to calculate the portion of the correlation c_a^2, whose spatial variations are only due to the spatial variations in the autocorrelation:

$$c_a^2 = \frac{1-a}{2} + \frac{1+a}{2}\overline{\overline{c_F^2}}, \tag{14}$$

where $\overline{\overline{c_F^2}}$ is the domain average of c_F^2. We also look at the forcing skill c_F^2 itself and compare it to the numerical model skill of precipitation and dewpoint depression.

4. Methodology

To predict the change in SMERGE anomalies, ΔS, one needs the current state and the "forcing" over the forecast interval. For an 8–14-day forecast of the change in S (i.e., predict the change at days 8–14 from days −6 to 0), our predictors could be the current S and the precipitation and dewpoint temperature anomalies for days 1–14 of the numerical model forecast. The total number of daily predictors suggested by such a statistical forecast is 29 (= 1 + 14 + 14), which would be subject to severe overfitting for any standard regression scheme.[3] Fortunately, in our system, we have a priori knowledge of the appropriate, physically based sign of the regression coefficients: the precipitation coefficients should be positive, the dewpoint coefficients should be negative, and the initial state coefficient should be negative (see section 3). Sign-constrained regression has regularization properties similar to more sophisticated regression methods that need some technique to determine the optimal regularization parameter (Meinshausen 2013; Slawski and Hein 2013). To see the significant constraints imposed by a simple sign constraint, note that for n predictors, sign-constrained regression restricts the space of allowable coefficients by a factor of 2^n. For $n = 29$, this factor is 5×10^8. Lorenz et al. (2017a,b; 2018) demonstrated the ability of sign constrained regression to maintain the skill of drought forecasts on independent data. Algorithmically, sign constrained regression is implemented by changing the sign of all predictors with coefficients that are expected to be negative, and then using non-negative least squares (NNLS). For this study, we implement NNLS using the cyclic coordinate descent algorithm of Franc et al. (2005) to forecast the change in SMERGE and ESI.

All results presented below are cross validated: 1) one year is left out, 2) the regression is trained on all other years, 3) skill scores are tested on the left out year, and 4) the process is

[3] We also explored the use of 3-, 5-, and 7-day running-mean anomalies to reduce the number of predictors. This did not improve the cross-validated skill scores of the sign constrained regression, suggesting that overfitting is not significant in our case.

repeated until all years have had the chance to be left out. Our null hypothesis is the regression where only one predictor, the initial state, is used (e.g., persistence):

$$\Delta S = a_0 - a_1 S,$$

where a_0 is a constant and a_1 is positive constant that describes the rate the soil moisture ''relaxes'' to climatology. We then add the numerical model predictors for all daily forecast times:

$$\Delta S = a_0 - a_1 S + \sum b_j x_j,$$

where the b_j have the appropriate a priori sign constraint, j is an index over both variable and forecast lead time, and the x values are the numerical model predictors. In this study, the x values are normalized so that a simple comparison between regression coefficients provides a useful measure of their relative importance. This new numerical model regression is used if and only if the cross validated skill is better than the (cross validated) null hypothesis. The selection of the numerical model predictors (both number and kind) was tested using the domain average cross-validated skill. For both SMERGE and ESI the same two variables work the best: 1) precipitation and 2) the dewpoint depression (maximum daily temperature minus the dewpoint temperature).[4] Other combinations related to dewpoint depression and the diurnal temperature range work almost as well; however, using three variables did not improve the forecast skill so only two variables are used for simplicity. We analyzed the regression residuals, and it appears that the linearity assumption is valid (not shown).

A statistical forecast typically will not explain all variability of the predictand. Therefore, a statistical forecast is not complete without a characterization of the probability distribution function (PDF) of the residuals: $y - y_p$, where y is the observed predictand and y_p is the prediction. Least squares regression assumes the PDF of the residuals is a Gaussian distribution with a standard deviation that is independent of the predictors x_j. For ESI, we find that the standard linear regression assumption works well, so this distribution is used for the probabilistic forecasts below. For SMERGE, on the other hand, the residuals have nonzero skewness and excess kurtosis. Because the tails of the SMERGE residuals are of the form $\exp(-x^2)$, we choose to use a mixture of two Gaussian distributions to describe the PDFs of the residuals. While the PDF parameters of the mixed Gaussian are typically not robust, the quantiles generated from the fitted distribution are robust (Leytham 1984). Also, we will fit the mixed Gaussian PDFs using a combination of moments and likelihood instead of the pure likelihood approach because the combination approach has better log-likelihood when applied to independent data. The combination approach constrains four of the five parameters of the mixed Gaussian using the mean, standard deviation, skewness, and kurtosis (section 2.1 of Cohen 1967). The remaining degree of freedom is found by maximizing the likelihood over one dimension using the standard Brent (1971) method.

5. Results

a. Soil moisture (SMERGE)

We begin with forecasts of the change in 7-day composite SMERGE 8–14 days in the future. Forecasts are made for each day in the warm season (May–September). The cross-validated temporal correlation between the predicted and actual SMERGE tendency is shown in Fig. 1a. The highest skill is in the northern Great Basin, the northern Great Lakes region and northern New England. The least skill is located in the Mojave Desert and in an east–west band extending across the Corn Belt from Nebraska to Ohio. A portion of the total skill is captured by the null hypothesis, which only depends on the current anomaly and quantifies the tendency to relax toward climatology. The relative skill of the null hypothesis is quantified by taking the ratio of the variance explained from the null hypothesis to the total variance explained (Fig. 1b), which in our case is also the percent variance explained from the initial condition. We also show the complement of this ratio, which is the percent variance explained by the numerical model forecast (Fig. 1c). The skill in the northern Great Lakes and eastern seaboard is mostly a result of the initial condition. While this is real skill, a deterministic forecast based on this regression will seldom predict drought intensification since the relaxation to climatology is the dominant mechanism in the forecast. In section 5d, we analyze this intensification issue in more detail and show the advantages of probabilistic forecasts (as opposed to deterministic forecasts) in capturing intensification. In the Great Basin, on the other hand, 60%–70% of the skill is coming from the numerical model forecasts of precipitation and dewpoint depression. Furthermore, although the central United States has some of the least skill, it also has the largest fraction of skill coming from the model forecast (Fig. 1c).

The role of precipitation and dewpoint depression on the forecast skill as a function of numerical model time lag is quantified by the regression coefficients from the NNLS (Fig. 1d). Because the general structure as a function of time does not depend that much on space, a domain average regression coefficient is shown. Due to the nonnegativity constraint and the normalization of the predictors, this is a reasonable measure of the relative importance of the two variables. The weights for precipitation are nearly constant for lead times of 1–7 days. After the start of the forecast composite at day 8, the weights start to gradually lose amplitude until they reach low values at day 12. For dewpoint depression, the weights maximize at day 1, with a secondary maximum at day 8.[5] The spatial structure of the weights given by the regression coefficients averaged over forecast lead time is shown in Figs. 1e and 1f. The average precipitation coefficient is largest in the north,

[4] Each variable corresponds to n individual predictors, where n is the maximum forecast lead time; $n = 14$ in this case.

[5] This secondary maximum structure seen in the domain average is robust because it is seen in most subdomains. The physical reasons for this structure are currently unknown.

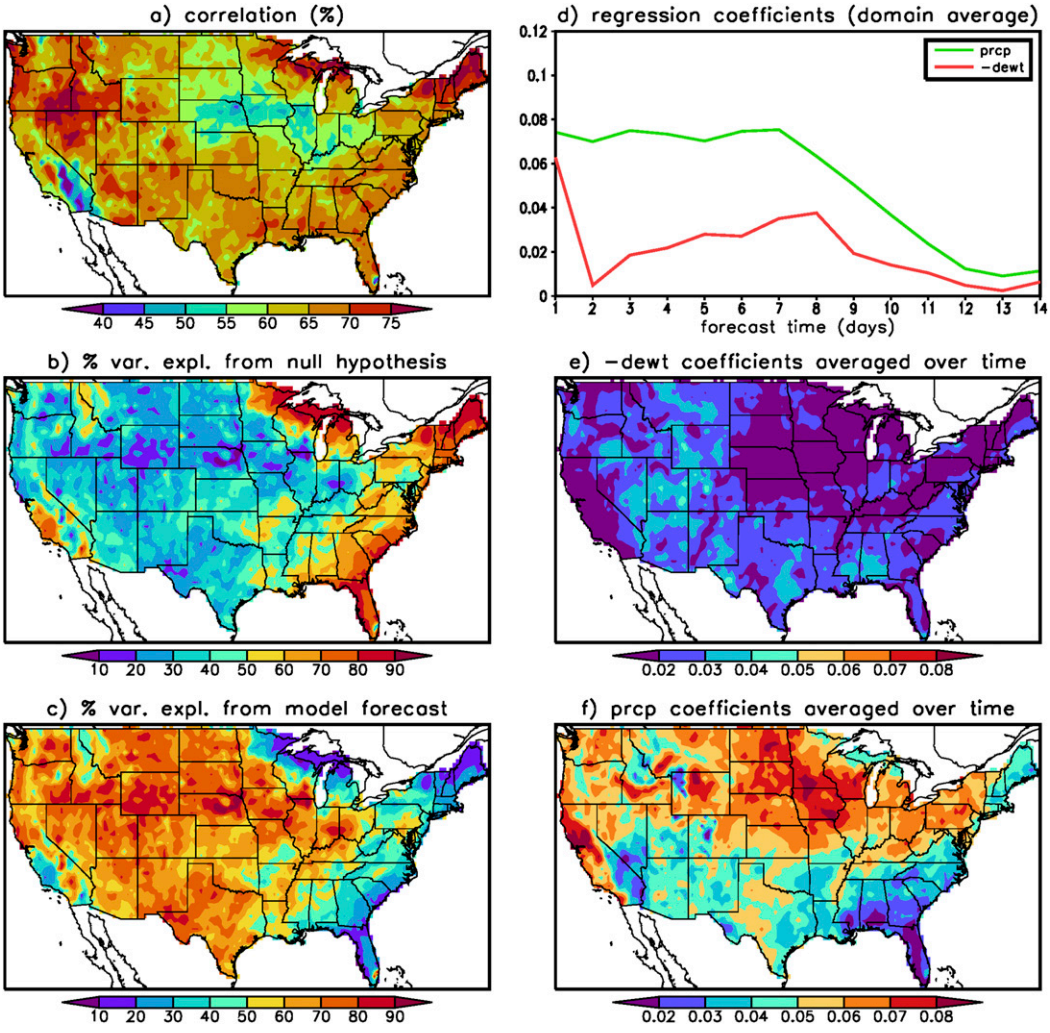

FIG. 1. (a) Cross-validated temporal correlation (%) between the observed and predicted SMERGE change (8–14-day composite SMERGE minus SMERGE from −6 to 0 day). (b) Percent variance in (a) explained by the null hypothesis (i.e., initial soil moisture anomaly alone). (c) Percent variance in (a) explained by the numerical model forecasts of precipitation and dewpoint depression. (d) Domain average regression coefficients of precipitation (green) and (the negative of) dewpoint depression (red) as a function of forecast lead time. (e) Temporal average regression weights for dewpoint depression (sign reversed). (f) Temporal average regression weights for precipitation.

while the dewpoint depression coefficient is largest across the southern United States and the Rocky Mountains. Regardless, almost everywhere precipitation is more important than dewpoint depression. A few exceptions are parts of the southwestern United States, especially the Mojave Desert, and the far southeastern United States, especially southern Alabama and Mississippi.

The effect of predictand autocorrelation on the forecast correlation c_a [see Eq. (14)], captures the high skill exhibited in the northern Great Lakes and the eastern seaboard (Fig. 2a, compare with Fig. 1a), therefore the high skill in these regions is caused by low autocorrelation (Fig. 2d). Such skill is due to the initial condition and captures the tendency of anomalies to rapidly decay back to climatology. On the contrary, the high skill in parts of the Intermountain West is not captured by c_a. Instead, the skill here is due to high predictability of the SMERGE forcing c_F^2 by the numerical model forecasts (Fig. 2c). Also shown in Fig. 2b is the percent variance explained by the null hypothesis (i.e., initial condition) as implied by spatial variations in the autocorrelation:

$$\frac{1-a}{(1-a)+(1+a)\overline{\overline{c_F^2}}}, \tag{15}$$

which is simply the first term on the right of (12) divided by the total right-hand side. This figure demonstrates that the autocorrelation explains a significant fraction of the spatial variability in the actual percent variance explained by the

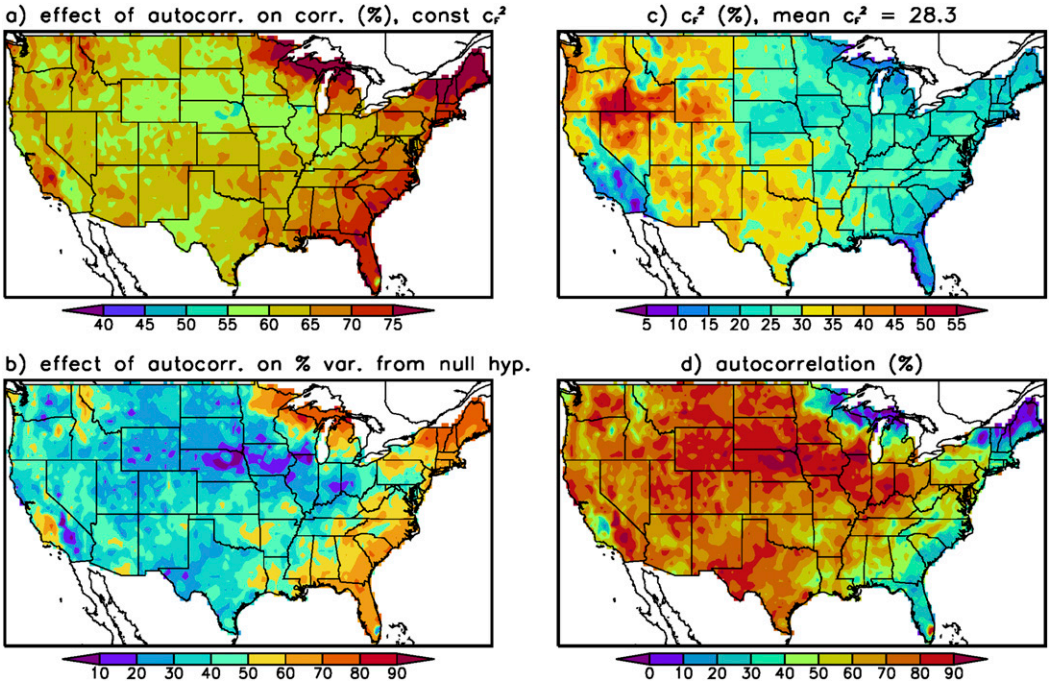

FIG. 2. (a) Hypothetical temporal correlation (%) between the observed and predicted SMERGE change when the value for c_F^2 at each grid point is replaced by its domain averaged value (see text). (b) As in (a), but for the hypothetical percent variance explained from the null hypothesis. (c) The value of c_F^2 (%), which is the fraction of the SMERGE "forcing" that is explained by the numerical model (see text). The domain averaged c_F^2, which is used for (a), is given in the caption. (d) The autocorrelation of the 7-day composite SMERGE at the time lag corresponding to the above forecasts (i.e., 7 days between composites).

null hypothesis (Fig. 1b). In summary, autocorrelation is quite important in explaining the spatial patterns of skill in SMERGE, although the skill of the numerical model forecasts c_F^2 is also important, especially in the western United States (Fig. 2c).

To help understand the spatial structure of c_F^2 in Fig. 2c, we also forecast the 1–14-day-average precipitation and dewpoint depression from corresponding numerical model fields. Because precipitation tends to be the most important for SMERGE (Figs. 1e,f), we expect precipitation skill to correspond most closely with c_F^2. Precipitation skill is largest in the northwest United States, particularly in the northern Great Basin and along the Oregon coast (Fig. 3a). These same regions show the highest values of c_F^2 (Fig. 2c). In the southwest United States and especially Texas, on the other hand, the c_F^2 values tend to be larger than that expected from the precipitation skill. This is explained by two factors: 1) the precipitation coefficients are smaller in the south (Fig. 1e) and therefore dewpoint depression skill becomes important and 2) dewpoint depression skill is largest in these regions (Fig. 3b). East of Texas, dewpoint depression skill decreases dramatically, which is consistent with the smaller c_F^2 in the eastern United States. A more detailed analysis that accounts for 1) covariability between precipitation and dewpoint depression at different lags and 2) the relationship between these variables and SMERGE would ultimately find complete consistency between the forcing skill, c_F^2 (Fig. 2c), and the precipitation and dewpoint depression skill (Fig. 3). This result follows from the fact that the simple model (section 3) is simply partitioning total skill between 1) the initial condition and 2) precipitation and dewpoint depression.

b. *Evaporative stress index*

Like SMERGE, we start with forecasts of the change in ESI 8–14 days in the future. The cross-validated temporal correlation between the forecast and actual ESI tendency is largest in the desert southwest, with isolated pockets of higher skill elsewhere. (Fig. 4a). In contrast to SMERGE, the percent variance explained by the null hypothesis is >50% across almost the entire domain (Fig. 4b).

The role of precipitation and dewpoint depression on ESI skill as a function of forecast lead time is shown in Fig. 4d. Unlike soil moisture, the dewpoint depression is more important than precipitation for most forecast times. The temporal structure of the weights, however, is similar. For example, the precipitation weights are largest at small forecast times and the dewpoint depression weights exhibit a relative maximum at 8 days. The dewpoint depression weights are very similar between ESI and SMERGE in both their spatial structure and amplitude (Figs. 4e,f). Like SMERGE, the precipitation weights tend to be more important in the northern United States; however, the weights are much smaller for ESI (Fig. 4f).

In Fig. 5, we repeat the diagnostics in the previous subsection in order to gauge the role of autocorrelation on the spatial

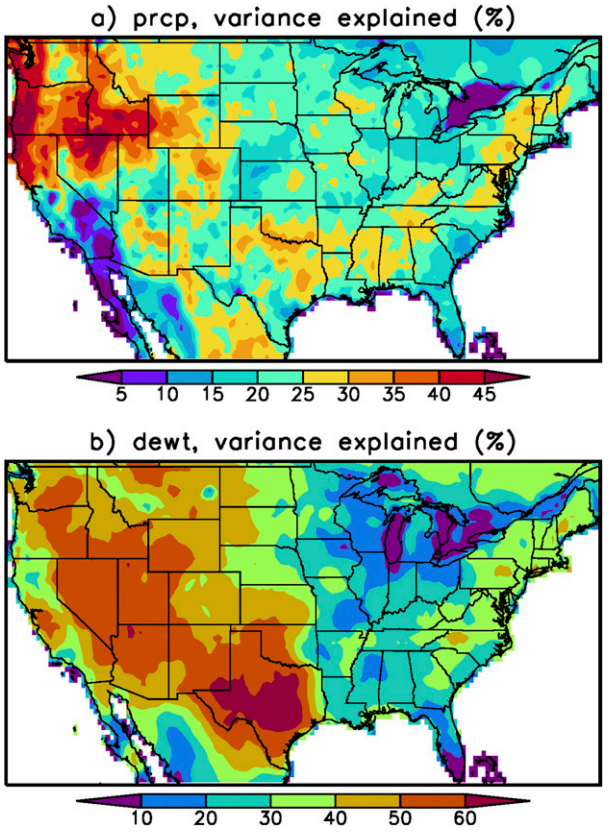

FIG. 3. (a) Percent variance explained for forecasts of 1–14-day-mean precipitation. (b) As in (a) but for dewpoint depression.

structure of skill. The role of autocorrelation c_a (Fig. 5a) is very similar to the actual correlation (Fig. 4a), demonstrating that much of the spatial variability in skill is due to spatial variability in autocorrelation of ESI rather than the numerical model forecast skill. In particular, the high skill in the Southwest is due to the low autocorrelation in that region. The autocorrelation also captures much of the fine scale structure in the central and eastern United States. On the other hand, the percent variance explained by the null hypothesis is not as well captured by the autocorrelation, although small scale features are well represented (Fig. 5b).

The largest difference between ESI and SMERGE is that the skill coming from the numerical model forecasts (c_F^2) is significantly less for ESI (Figs. 2c, 5c). For example, the area mean c_F^2 is 10.3% for ESI compared to 28.3% for SMERGE. Meanwhile, the autocorrelation for ESI is significantly smaller compared to SMERGE (Figs. 2d, 5d). Acting in isolation, this small autocorrelation would lead to higher total skill for ESI, which can be seen by a simple rearrangement of (12):

$$c^2 = \frac{1}{2}(1 + c_F^2) - \frac{a}{2}(1 - c_F^2). \tag{16}$$

(Note that the quantity $1 - c_F^2$ is never negative.) The fact that forecast skill is typically higher for SMERGE implies that c_F^2 is more important than the autocorrelation for between variable (i.e., SMERGE versus ESI) differences in skill.

An alternate way to highlight the differences between SMERGE and ESI is to plot skill versus autocorrelation for each grid point in the domain (Fig. 6a). Also superimposed on this plot are three lines depicting the relationship between c^2 and a described by (14) for three values of c_F^2: 0, 0.125, and 0.25. The black line, which corresponds to $c_F^2 = 0$, would typically represent an absolute lower bound on the scatter except that the skill c^2 is the cross-validated value instead of the within-sample value. The ESI tends to exhibit a linear relationship a small distance above the $c_F^2 = 0$ line, which implies that autocorrelation dominates the spatial variability in skill and that the null hypothesis dominates the total skill. For SMERGE, the data points are significantly above the $c_F^2 = 0$ line and moreover the relationship with autocorrelation is not linear for many grid points. For example, points with high autocorrelation that typically have lower skill instead have some of the highest skill in the entire domain. Only for a subset of points with low autocorrelation is a linear relationship between skill and autocorrelation observed.

The scatter between skill and autocorrelation is also a good way to see the effect of longer forecast lead times on skill. For example, Fig. 6b shows the same plot but for the 15–28-day forecasts (i.e., predict the change at days 15–28 from days −13 to 0). For ESI (blue), the points contract even closer to the $c_F^2 = 0$ line as the skill becomes more dominated by the initial conditions. In addition, there is a tendency for the autocorrelation to become smaller, which is typical for longer lead times. The accompanying leftward shift in the points implies an increase in skill due to autocorrelation. Indeed, the domain average ESI skill actually increases slightly for the 15–28-day forecasts. In some sense, however, the forecast quality has degraded as a and c_F^2 decrease because the initial condition skill only predicts the decay of anomalies. Such forecasts will not predict much drought intensification (see section 5d). A more dramatic change in the shape of the scatterplot is seen for the 15–28-day SMERGE forecasts: the c_F^2 values decrease so much that most of the nonlinearity in the scatter collapses to an essentially linear relationship. The SMERGE predictability becomes much like the ESI with skill mostly determined by the autocorrelation. Also, unlike the 8–14-day forecasts, the differences in ESI and SMERGE predictability are dominated by the autocorrelation: the ESI (SMERGE) tendency has more (less) skill because it is less (more) persistent. In summary, Fig. 6 shows that the 8–14-day forecasts are the most interesting because the skill from the numerical model forecasts is very inhomogeneous and quite large for SMERGE. As one moves toward longer time scales (the 15–28-day forecasts in Fig. 6b), the forecasts are less interesting because skill is dominated by the initial condition for both SMERGE and ESI. Hence, we do not discuss the 15–28-day results any further in this paper.

c. *Modeled ESI (NLDAS) and evaporation components*

To help understand the differences between ESI and SMERGE, we also forecasted modeled ESI derived from the NOAH, MOS, and VIC models. While there are some differences between the observed and modeled ESI (not shown), the differences are relatively minor, which suggests that the predictability of ESI is inherently different than SMERGE and that the differences are not due to any satellite retrieval issues associated with the observed ESI.

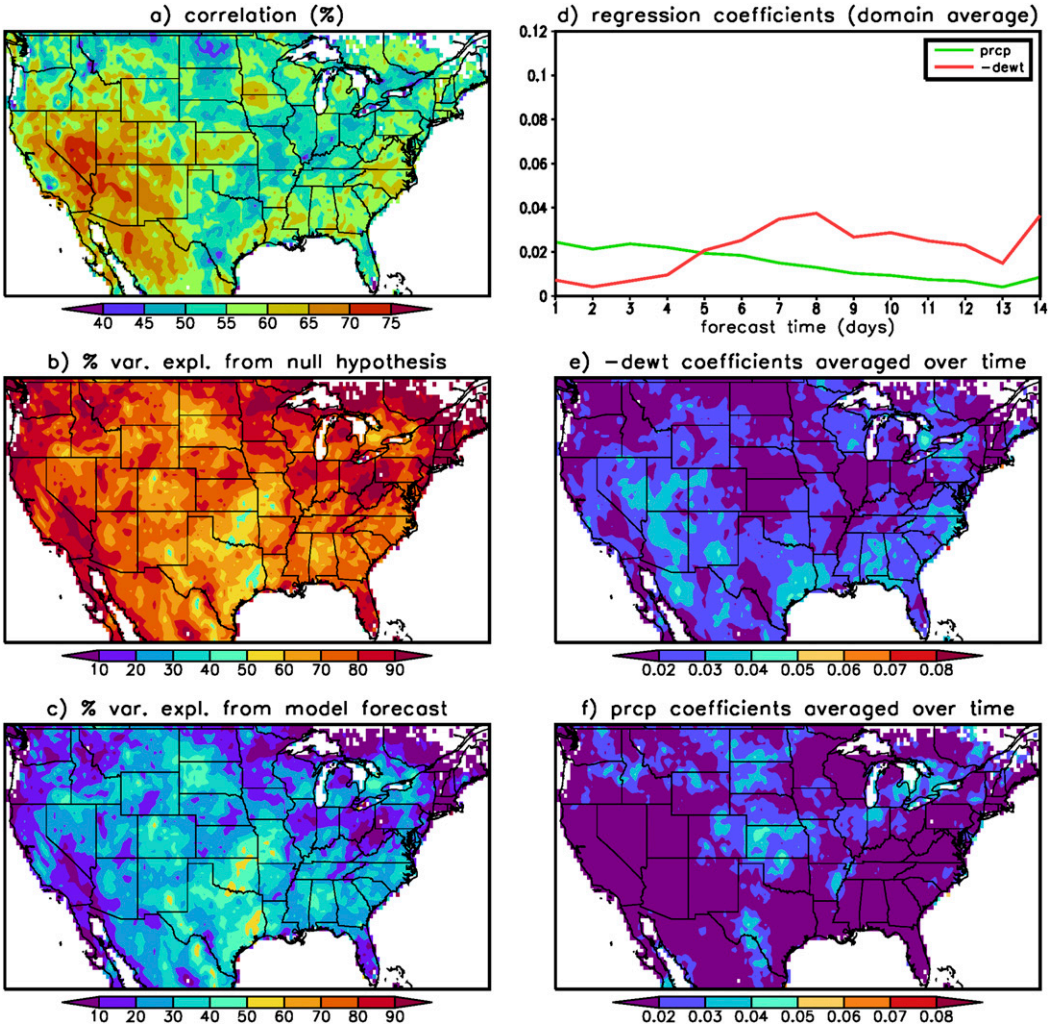

FIG. 4. (a) As in Fig. 1, but for the ESI. White areas over land have no cross-validated skill from the numerical model forecasts.

The modeled ESI is also used to assess the relative roles of transpiration, bare soil, and canopy evaporation on the time scale of ESI variability. While there are pronounced model differences in the relative partitioning of evaporation into these components (e.g., Kumar et al. 2018), the results of this study are robust among models. Here the total evaporation is divided into two components: 1) the bare soil and canopy water (BSCW) evaporation and 2) transpiration (TRANS). The BSCW and TRANS evaporation are then divided by the total PET and normalized to create a BSCW and TRANS ESI. The results from the three models are averaged together to create the ensemble mean autocorrelation for each component (Fig. 7). For almost all locations, the TRANS autocorrelation is significantly larger than the BSCW autocorrelation. This suggests it is useful to think of the ESI as two components: 1) a quickly evolving BSCW component and 2) a slowly evolving TRANS component. This is also consistent with our physical intuition. Knowing the ratio of the components in the observed ESI would help the predictability of the portion of the forecast involving the initial condition because the rate of decay for the two components are different:

$$\Delta E = -a_B E_B - a_T E_T, \tag{17}$$

where E is the ESI, the B and T subscripts denote the BSCW and TRANS components, and the a values are positive regression constants. Because the BSCW evaporation decays more rapidly, $a_B > a_T$. The partitioning of the NLDAS evaporation could be used in (15), however, we find that using the soil moisture (SMERGE) as a proxy for the transpiration E_T works best. To find the proper sign for the constrained regression, note that $E_B = E - E_T$ and substitute this into (15):

$$\Delta E = -a_B E + (a_B - a_T) E_T, \tag{18}$$

where $a_B - a_T$ is positive because $a_B > a_T$. Since we assume soil moisture is proportional to E_T, our improved regression should a priori assume that the SMERGE regression coefficient is positive.

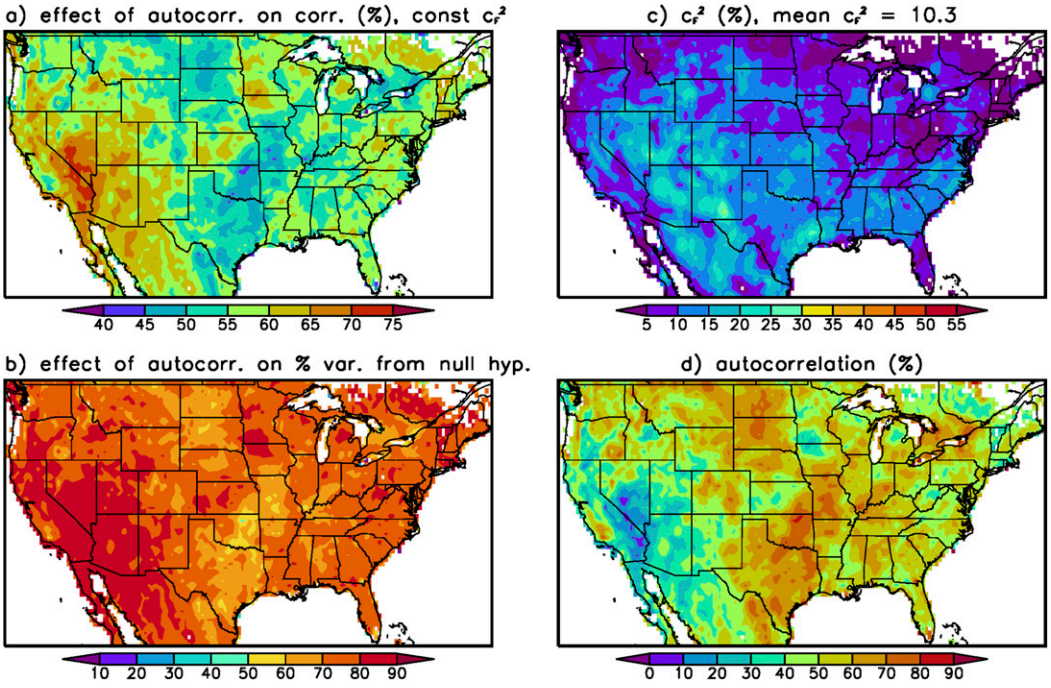

FIG. 5. (a) As in Fig. 2, but for the ESI. White areas over land have no cross-validated skill from the numerical model forecasts.

In Fig. 8, we repeat the ESI forecast the same as before but with one additional predictor: the current value of SMERGE. The SMERGE predictor helps constrain the relative contribution of long versus short time scale contributions to the initial ESI state. The correlations for the new regression are similar but there is a consistent improvement in the central United States, particularly over Missouri and surrounding states. The difference in correlation (Fig. 8b) tends to show the most improvement in regions with relatively small skill and large autocorrelation (Fig. 5d), which is consistent with a larger role for root zone soil moisture in the ESI. Although the improvements are relatively small, for the remainder of this paper, we use these improved ESI forecasts. A similar analysis for SMERGE using the current ESI as a predictor did not yield improvements, perhaps because the root zone soil moisture does not have a rapidly evolving component like ESI.

d. *Probabilistic forecasts and intensification*

As mentioned previously, although deterministic forecasts dominated by null hypothesis skill can have high predictability, they are less able to forecast intensifying anomalies because

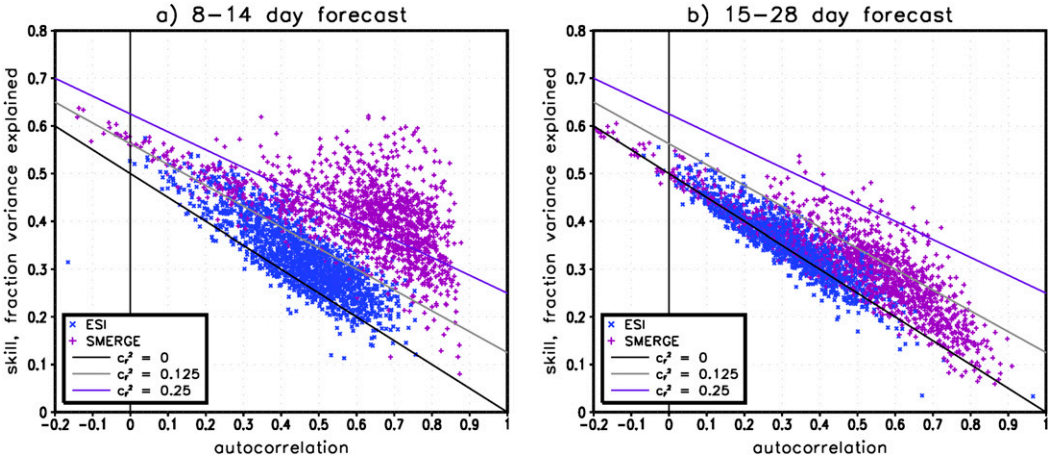

FIG. 6. (a) Scatterplot between autocorrelation (x axis) and skill (variance explained, y axis) for the ESI grid points (blue x) and the SMERGE grid points (purple +) for the 8–14-day forecasts. For clarity, every other grid point in both the x and y directions is skipped. The black, gray, and blue lines are the expected relationship between skill and autocorrelation when c_F^2 is 0, 0.125, and 0.25, respectively. (b) As in (a), but for the 15–28-day forecasts.

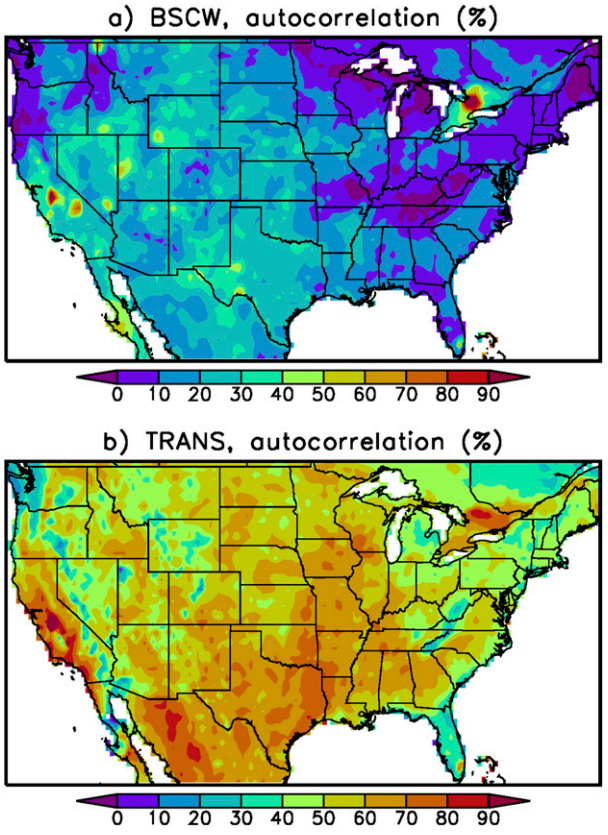

FIG. 7. (a) Autocorrelation (8–14 days) for the bare soil plus canopy water (BSCW) component of the evaporation averaged over three NLDAS models. (b) As in (a), but for the transpiration (TRANS) component of the evaporation.

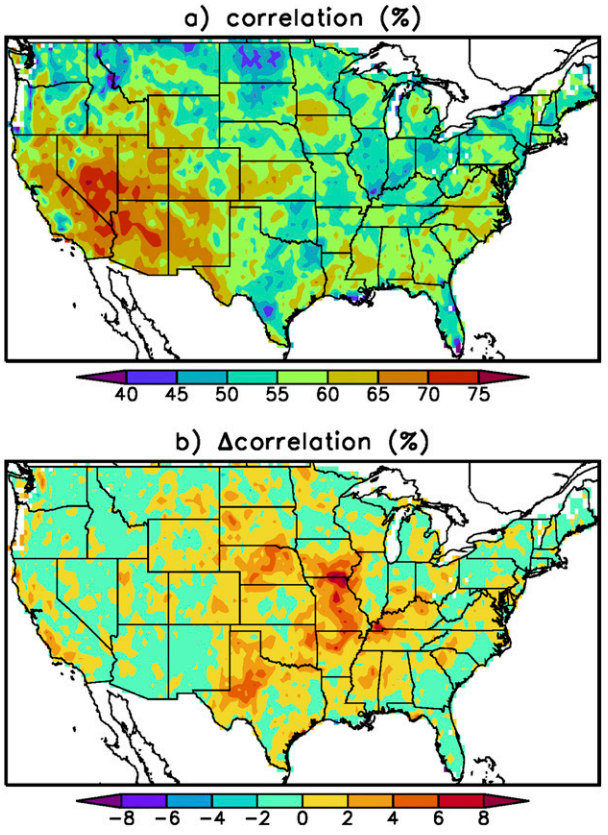

FIG. 8. (a) As in Fig. 1, but for the ESI with both initial ESI and initial SMERGE used as predictors. (b) The correlation in (a) minus the correlation in Fig. 4a (i.e., change in correlation). Note that the forecast no longer includes southern Canada and northern Mexico because these are not included in SMERGE.

the forecasts are strongly pulled toward the climatological mean. To diagnose such deficiencies, we first compute the percent of cases where a positive (wet) anomaly (7-day composite) is followed by a positive (wet) tendency 8–14 days later for the SMERGE observations (Fig. 9a). For uncorrelated anomalies and tendencies, one expects a value of 25% for this statistic. In reality, the values are less than 25% due to a natural tendency for anomalies to decay. Next, this diagnostic is repeated using forecasted tendencies rather than observed tendencies (Fig. 9b). While some patterns are captured, such as relatively high values in the lower Midwest, the forecasted amplitude is significantly smaller than observed. This is especially true in the northern Great Lakes, the Eastern Seaboard, and California. As expected, these are the same regions where the fraction of variance explained by the null hypothesis is very large (Fig. 1b). Since any forecast explains only a portion of the total variability, one needs to consider the distribution of the remaining unexplained variance. Using the residual PDFs (section 4) to expand the range of possible outcomes for a given y_p, we compute a revised probability of positive tendency and positive anomaly as estimated by our forecasts. One approach is to randomly sample from the residual PDF and then add that random value to each forecast. In our case, however, we integrate the PDFs directly (Fig. 9c). The agreement with the observed probability (Fig. 9a) is much improved, demonstrating that the residuals are well captured by the PDFs in this case. The PDF method corrects the above errors by assigning nonzero intensification probabilities even when the mean forecast predicts a reduction.

In Figs. 9d–f, the above analysis is repeated for the case of negative anomalies and negative tendencies, which is relevant for intensifying drought. The spatial patterns in this statistic are quite different than the positive case with large values in western United States instead of the Midwest. In fact, some regions exhibit values greater than 25%, signifying that negative anomalies have more tendency to amplify rather than decay. As before, the raw deterministic forecasts capture spatial features but with significantly reduced amplitude. The probabilistic forecasts definitely improve the chance of intensification but unlike the positive case there is a slight underestimation of the full amplitude. This suggest that perhaps alternate PDFs should be used for the SMERGE residuals.

In Fig. 10, the above intensification analysis is repeated for the ESI. Compared to SMERGE, the intensification probabilities tend be more or less uniform across the domain. In addition, the deterministic forecasts are significantly worse for ESI, which is consistent with the large fraction of skill due to the null hypothesis for ESI (Fig. 4b). On the other hand, the PDF of the ESI residuals is apparently a better fit to the true

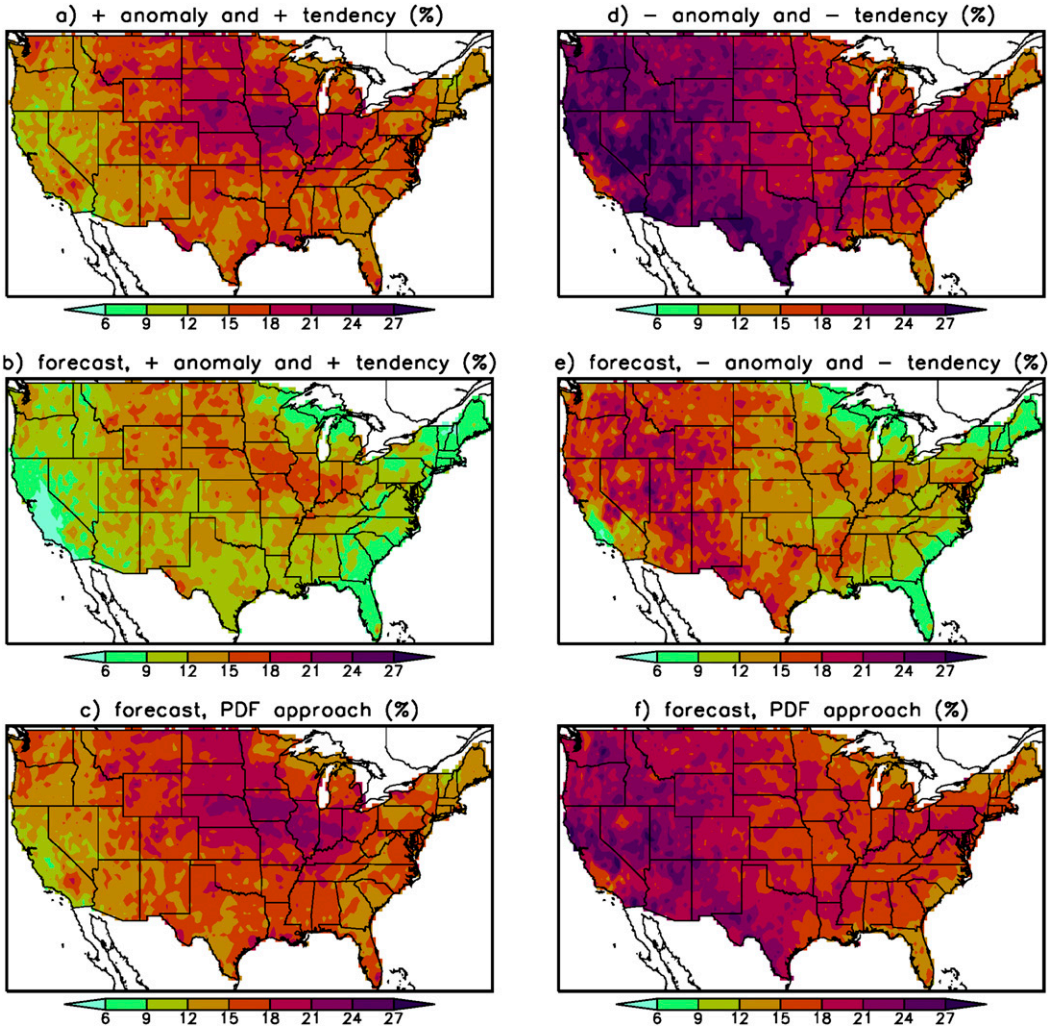

FIG. 9. (a) Percent of times where a positive SMERGE anomaly (7-day composite) is associated with a positive change in SMERGE 8–14 days later. (b) As in (a), except the change in SMERGE is replaced by the forecasted change in SMERGE. (c) As in (b), but the forecasted change is a PDF instead of a single value. The added spread in possible outcomes increases the chance of positive anomalies and positive changes in the forecasts. (d)–(f) As in (a)–(c), but for times where a negative SMERGE anomaly is associated with a negative change.

PDF and hence the probabilistic forecasts are able to nearly reproduce the observed intensification statistics (Figs. 10c,f)

We also explored additional probabilistic skill scores such as the ranked probability skill score (RPSS), which was applied to tercile forecasts SMERGE and ESI change. These results are not shown here because we found that the RPSS is well approximated in terms of the correlation via Eq. (21) in Tippett et al. (2010): $\text{RPSS} = 1 - \sqrt{1-\rho^2}$, where ρ is the correlation.

e. Examples

In this section, examples of observed and predicted SMERGE and ESI tendencies are given for the last eight years of the analysis. Here, only the deterministic forecast (i.e., the mean forecast) is shown. Because of missing ESI values, we show the average of all individual forecasts/observations over a month rather than an individual forecast/observation on a single day. For consistency, the same averaging is performed for SMERGE. We begin with SMERGE for the month of June (Fig. 11). For some years, the spatial structure and amplitude of the forecasts are quite good, which is encouraging given the intensification issues documented in the previous subsection. For example, the intensification of the 2012 flash drought is well captured by the forecasts (Figs. 11e,f). For other years, the spatial structures are reasonable, but their amplitudes are too weak (2013 and 2016), or the spatial structure is not good (2011). For August, the year 2013 shows good correspondence for both structure and amplitude (Fig. 12). However, there are two cases where the forecasts missed a significant soil moistening event in the central United States (2012, 2018).

Examples for ESI in June are shown in Fig. 13. Despite the intensification issues described in the previous subsection, the ESI forecasts are able to predict the amplitude of drought

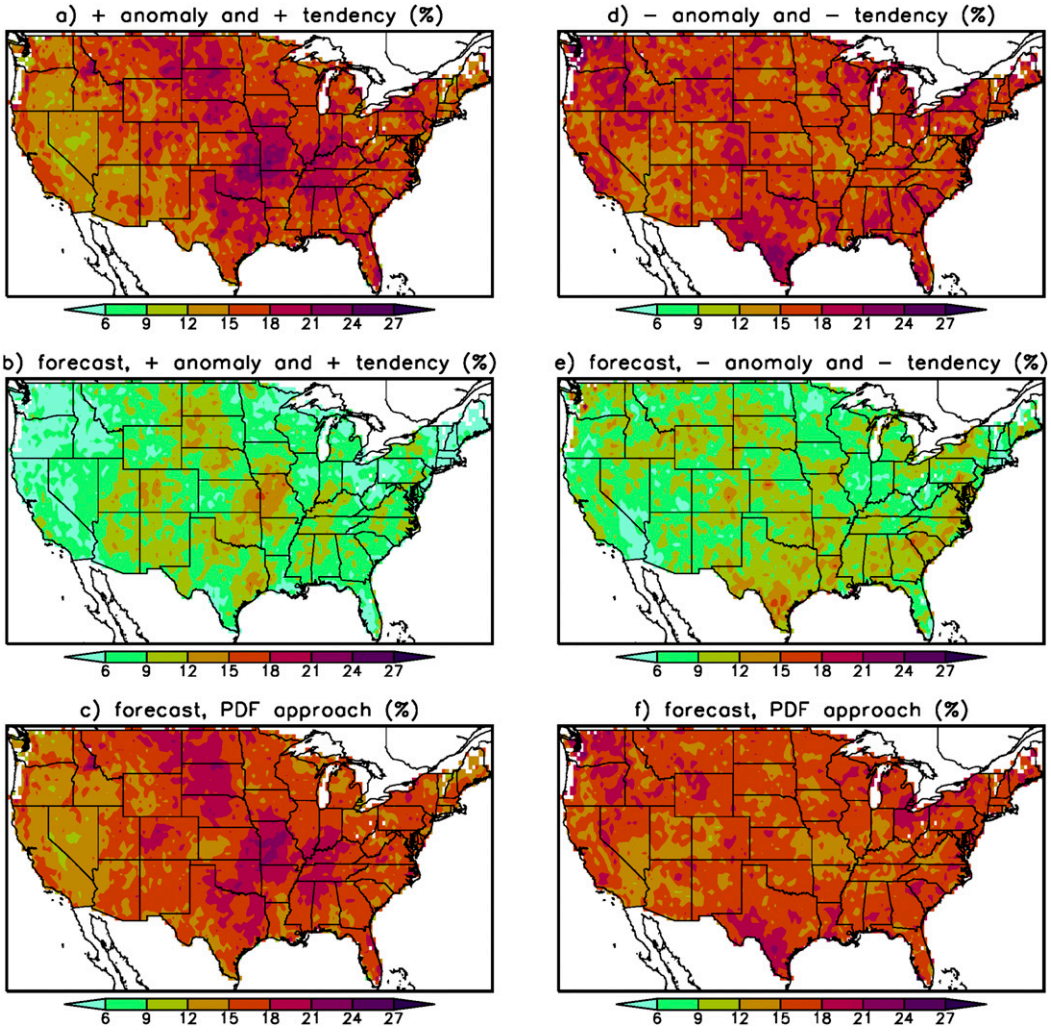

FIG. 10. As in Fig. 9, but for ESI.

intensification in the central United States in 2012. Similar agreement is seen for the drying in 2013 but in this case only the southern portion of the central United States was already anomalously dry (not shown). Like SMERGE, 2011 was a poor forecast for ESI as well. For August 2010 (Fig. 14), there is also strong correspondence between the observed and predicted ESI tendency and moreover significant portions of the drying regions are already dry and therefore "drought" is intensifying. However, like SMERGE, the ESI forecasts missed the moistening event in Illinois and Missouri in 2012 (but not 2018). Overall, the ESI forecasts do not look noticeably worse than the SMERGE forecasts in these examples. Another somewhat surprising feature of both the SMERGE and ESI examples is the close correspondence of the patterns at small spatial scales.

6. Conclusions

In this paper, we have developed probabilistic forecasts of changes in soil moisture (SMERGE) and evaporative stress index (ESI) on subseasonal time scales over the contiguous United States. The forecasts use current SMERGE and ESI conditions together with numerical weather forecasts from the ECMWF model of the S2S Prediction Project. We have also developed a deeper understanding of the observed predictability using a simple analytical model framework and output from NLDAS-2.

Changes in soil moisture are quite predictable on weekly time scales, with 50% or more of the variance explained over the majority of the contiguous United States. Changes in ESI are significantly less predictable than soil moisture except in the southwest United States. A simple autoregressive red noise model provides a greater understanding of the predictability. This model demonstrates that the spatial variations in skill are primarily a result of spatial variations in the autocorrelation, or persistence, of the predicted variable, especially for ESI. For example, the high ESI skill in the southwest United States is due to the small autocorrelation of the ESI in this region. The contribution of the numerical model forecasts to skill is more spatially homogeneous. In contrast to within variable predictability, the autocorrelation does not explain the differences between SMERGE and ESI. In fact, based on the autocorrelation alone, one

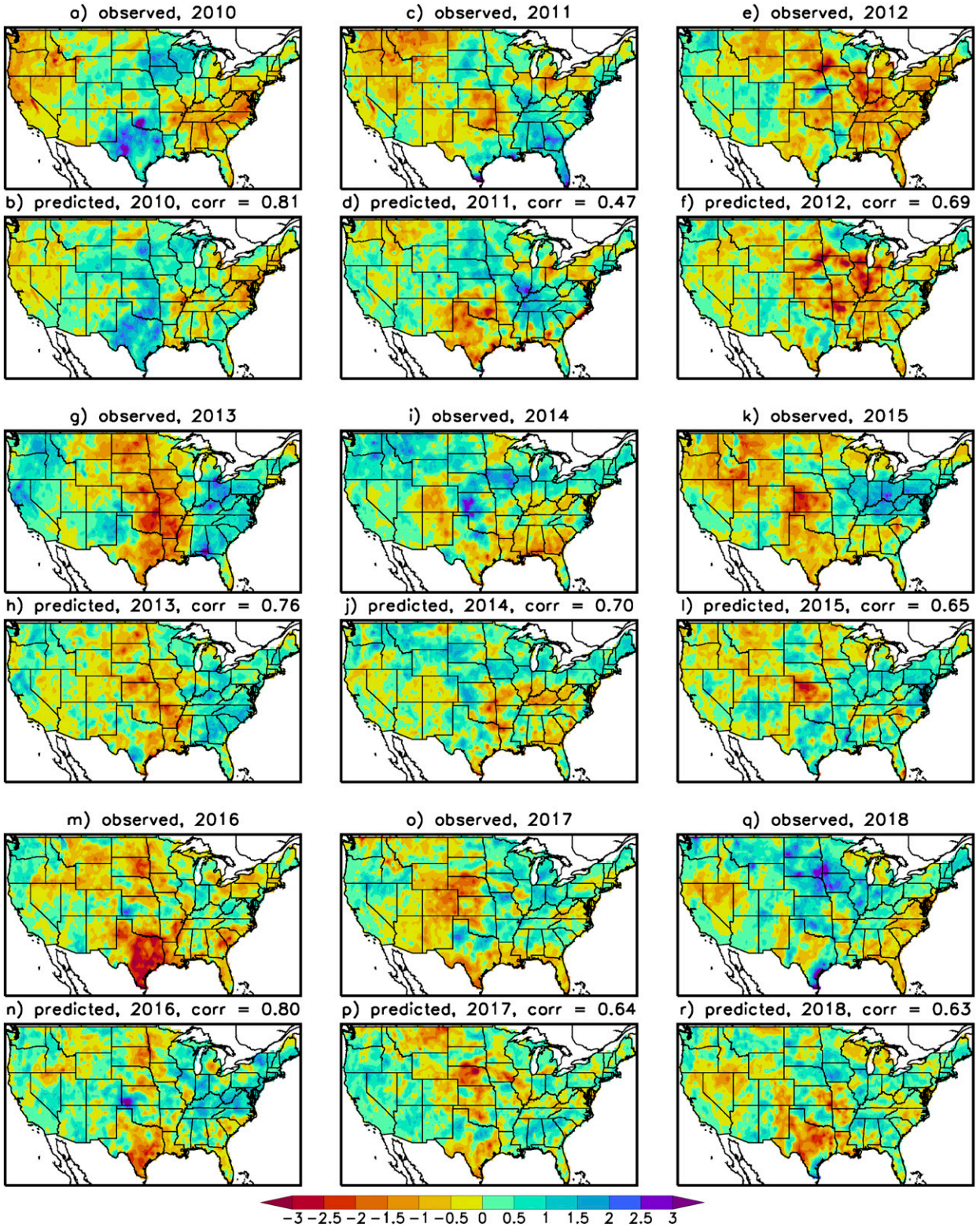

FIG. 11. Comparison of the change in SMERGE and the cross-validated predictions (i.e., the model has not "seen" the year shown) averaged over all June days for the latest 9 years. (a) Observed change in SMERGE for June 2010 and (b) prediction for June 2010. The spatial correlation between the observed and the prediction is shown, and so on. All examples are for the 8–14-day change of the 7-day composite. Units are percent water by volume.

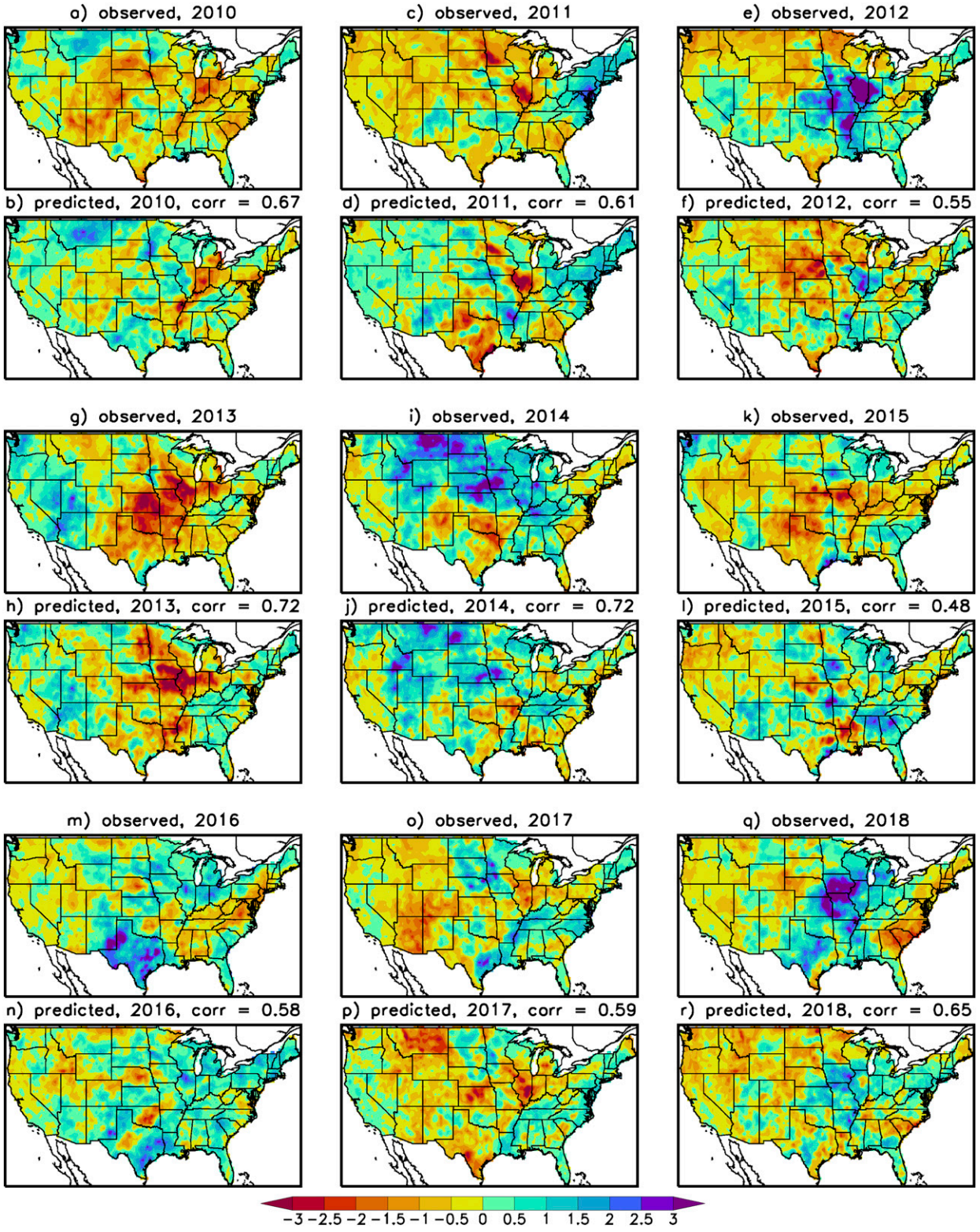

FIG. 12. As in Fig. 11, but for August.

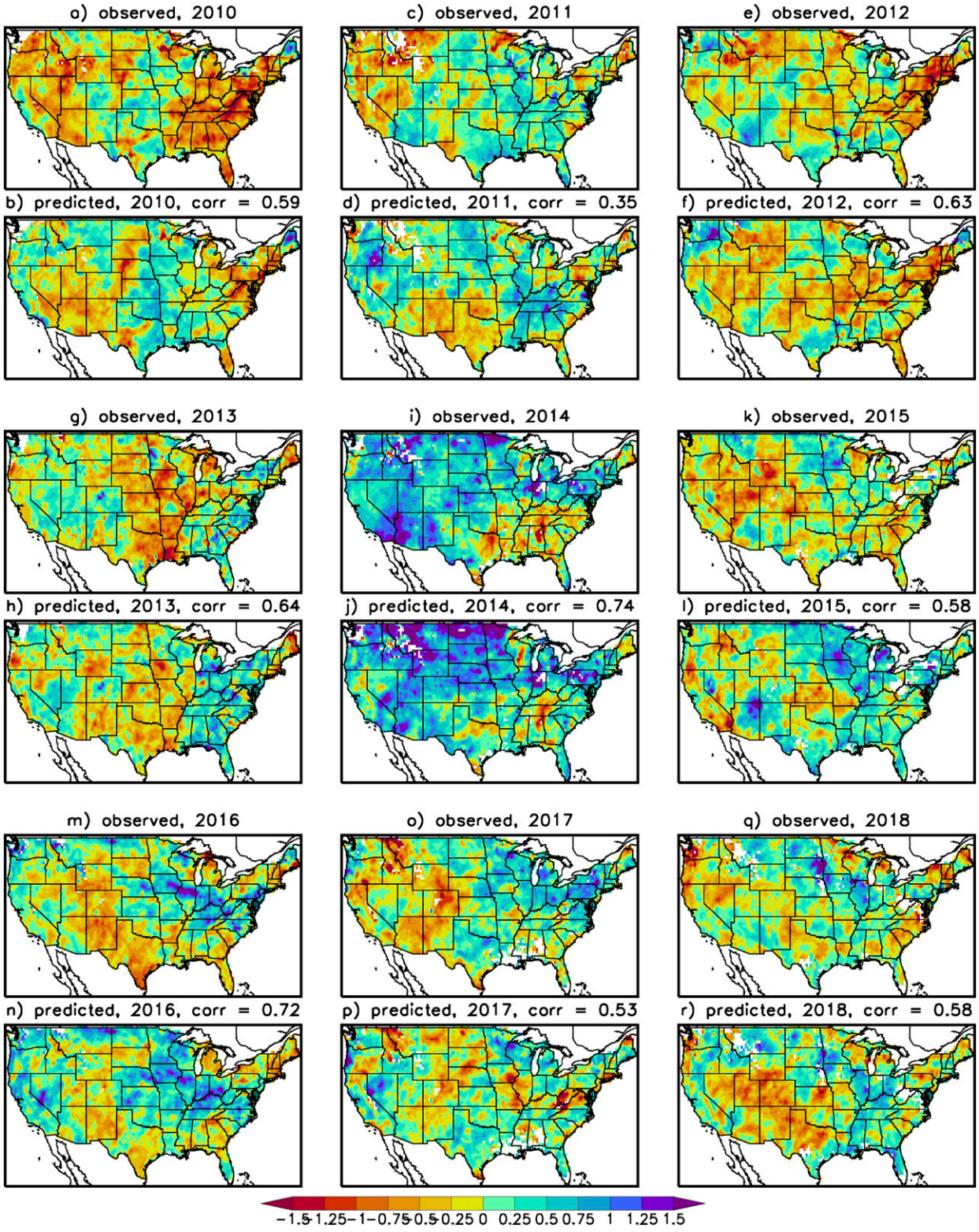

FIG. 13. As in Fig. 11, but for ESI and units are standardized anomalies.

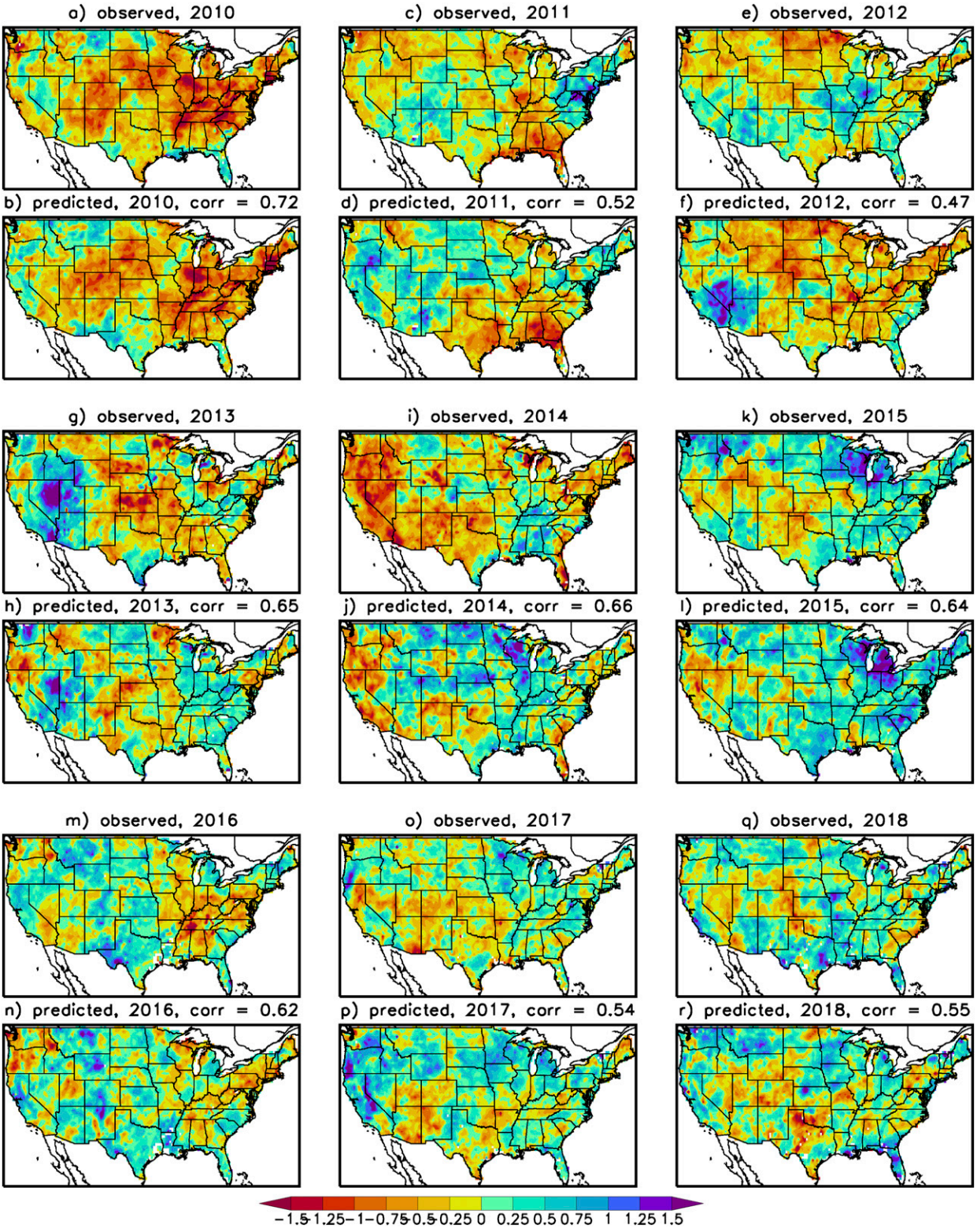

FIG. 14. As in Fig. 13, but for August.

would expect changes in soil moisture to be more predictable than ESI, but the opposite is the case. Instead, these differences are due to the greater SMERGE predictability by the ECMWF S2S model.

As the forecast lead time is increased from 8–14 to 15–28 days, the initial condition becomes more important for skill for ESI and especially SMERGE. For these forecasts, SMERGE skill is also dominated by the relaxation to climatology and the differences between SMERGE and ESI are due to differences in autocorrelation.

In addition, modeled ESI (NLDAS) was partitioned into the bare soil plus canopy water (BSCW) evaporation and transpiration (TRANS) components. These two components of ESI have dramatically different time scales, and therefore different rates of decay to climatology. This motivated an improved ESI prediction scheme that uses SMERGE to help partition the total ESI initial condition into its fast and slow components.

Forecasts dominated by a decay to climatology dramatically underestimate the number of cases of anomaly intensification. This problem is most evident for the ESI given the importance of the initial condition for forecast skill. This issue can be eliminated by developing probabilistic forecasts that properly characterize the spread of the forecast errors about the predicted value.

In earlier work predicting the USDM (Lorenz et al. 2018) from the North American Multi-Model Ensemble (NMME; Kirtman et al. 2014), we found only very small improvements due to the NMME. In this paper, we find significantly more skill from the ECMWF S2S model for both SMERGE and ESI. We believe a significant part of the difference stems from the fact that the USDM and the NMME are asynchronized: the USDM is weekly and the NMME was integrated every 5 days. This asynchronicity meant that the effective NMME forecast time was longer than intended. Because the contribution of the numerical model forecasts to predictability decreases rapidly as forecast lead time increases, this asynchronicity is consistent with less skill. In addition, the USDM is sometimes a lagged indicator of conditions on the ground, which can lead to timing issues for short forecast lead times and therefore lower skill. This is consistent with the fact that the NMME contributes more to the 4-week USDM forecasts than to the 2-week forecasts (Lorenz et al. 2018).

In the future, improvements to the forecast methodology developed during this study might be possible through a better characterization of the initial state. For example, we found that quantifying the contribution of soil moisture to the initial ESI state improved skill in the central United States. Further improvements might be possible to SMERGE by quantifying the role of shallow versus deep soil moisture to the total root zone soil moisture. In addition, the timing of precipitation events relative to the nonmissing ESI days might better characterize the canopy water and bare soil contribution to ESI. Also, this study used the ECMWF S2S numerical model simulations because of the nearly complete daily coverage and the relatively large number of ensemble members (11). Unfortunately, S2S simulations are not available in real time. In the future, we will apply our methodology to the models of the Subseasonal Experiment (SubX) (Pegion et al. 2019), which are available in real time.[6] For the SubX, however, the ensemble sizes tend to be smaller (4), and forecasts are initialized less frequently, which might impact the robustness of our results. We also plan to apply more advanced machine learning methods to soil moisture and ESI predictability. Such nonlinear methods can potentially account for dependencies between soil moisture predictability and initial soil moisture state that has been observed in previous studies (Orth and Seneviratne 2013).

[6] The SubX includes NOAA's Global Ensemble Forecast System (GEFS).

Acknowledgments. This work was supported by funds provided by the NSF PREEVENTS ICER-1854902.

REFERENCES

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith, 1998: Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, 300 pp., www.fao.org/docrep/X0490E/X0490E00.htm.

Anderson, M. C., J. M. Norman, G. R. Diak, W. P. Kustas, and J. R. Mecikalski, 1997: A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sens. Environ.*, **60**, 195–216, https://doi.org/10.1016/S0034-4257(96)00215-5.

——, W. P. Kustas, and J. M. Norman, 2007a: Upscaling flux observations from local to continental scales using thermal remote sensing. *Agron. J.*, **99**, 240–254, https://doi.org/10.2134/agronj2005.0096S.

——, J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas, 2007b: A climatological study of evapotranspiration and moisture stress across the continental U.S. based on thermal remote sensing: 1. Model formulation. *J. Geophys. Res.*, **112**, D10117, https://doi.org/10.1029/2006JD007506.

——, C. R. Hain, B. Wardlow, J. R. Mecikalski, and W. P. Kustas, 2011: Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States. *J. Climate*, **24**, 2025–2204, https://doi.org/10.1175/2010JCLI3812.1.

Barlage, M., and Coauthors, 2010: Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *J. Geophys. Res.*, **115**, D22101, https://doi.org/10.1029/2009JD013470.

Bowling, L. C., and D. P. Lettenmaier, 2010: Modeling the effects of lakes and wetlands on the water balance of Arctic environments. *J. Hydrometeor.*, **11**, 276–295, https://doi.org/10.1175/2009JHM1084.1.

Brent, R. P., 1971: An algorithm with guaranteed convergence for finding a zero of a function. *Comput. J.*, **14**, 422–425, https://doi.org/10.1093/comjnl/14.4.422.

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.

Christian, J. I., J. B. Basara, J. A. Otkin, and E. D. Hunt, 2019a: Regional characteristics of flash droughts across the United States. *Environ. Res. Commun.*, **1**, 12 5004, https://doi.org/10.1088/2515-7620/ab50ca.

——, ——, ——, ——, R. A. Wakefield, P. X. Flanagan, and X. Xiao, 2019b: A methodology for flash drought identification: Application of flash drought frequency across the United States. *J. Hydrometeor.*, **20**, 833–846, https://doi.org/10.1175/JHM-D-18-0198.1.

Cohen, A. C., 1967: Estimation in mixtures of two normal distributions. *Technometrics*, **9**, 15–28, https://doi.org/10.1080/00401706.1967.10490438.

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.

Ford, T. W., and C. F. Labosier, 2017: Meteorological conditions associated with the onset of flash drought in the eastern United States. *Agric. For. Meteor.*, **247**, 414–423, https://doi.org/10.1016/j.agrformet.2017.08.031.

——, D. B. McRoberts, S. M. Quiring, and R. E. Hall, 2015: On the utility of in situ soil moisture observations for flash drought early warning in Oklahoma, USA. *Geophys. Res. Lett.*, **42**, 9790–9798, https://doi.org/10.1002/2015GL066600.

Franc, V., V. Hlaváč, and M. Navara, 2005: Sequential coordinate-wise algorithm for the non-negative least squares problem. *CAIP 2005: Computer Analysis of Images and Patterns*, A. Gagalowicz and W. Philips, Eds., Springer, 407–414, https://doi.org/10.1007/11556121_50.

Hunt, E., K. G. Hubbard, D. A. Wilhite, T. J. Arkebauer, and A. L. Dutcher, 2009: The development and evaluation of a soil moisture index. *Int. J. Climatol.*, **29**, 747–759, https://doi.org/10.1002/joc.1749.

——, M. Svoboda, B. Wardlow, K. Hubbard, M. J. Hayes, and T. Arkebauer, 2014: Monitoring the effects of rapid onset of drought on non-irrigated maize with agronomic data and climate-based drought indices. *J. Agric. For. Meteor.*, **191C**, 1–11, https://doi.org/10.1016/j.agrformet.2014.02.001.

Kirtman, B. P., and Coauthors, 2014: The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Koster, R. D., and M. J. Suarez, 1994: The components of a SVAT scheme and their effects on a GCM's hydrological cycle. *Adv. Water Resour.*, **17**, 61–78, https://doi.org/10.1016/0309-1708(94)90024-8.

——, and ——, 1996: Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo. 104606, Vol. 9, 76 pp., http://gmao.gsfc.nasa.gov/pubs/docs/Koster130.pdf.

Kumar, S., T. Holmes, D. M. Mocko, S. Wang, and C. Peters-Lidard, 2018: Attribution of flux partitioning variations between land surface models over the continental US. *Remote Sens.*, **10**, 751, https://doi.org/10.3390/rs10050751.

Leytham, K. M., 1984: Maximum likelihood estimates for the parameters of mixture distributions. *Water Resour. Res.*, **20**, 896–902, https://doi.org/10.1029/WR020i007p00896.

Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface and soil moisture parameterization of the VIC-2L model: Evaluation and modifications. *Global Planet. Change*, **13**, 195–206, https://doi.org/10.1016/0921-8181(95)00046-1.

Lorenz, D. J., J. A. Otkin, M. Svoboda, C. R. Hain, M. C. Anderson, and Y. Zhong, 2017a: Predicting U.S. Drought Monitor (USDM) States using precipitation, soil moisture, and evapotranspiration anomalies. Part I: Development of a nondiscrete USDM index. *J. Hydrometeor.*, **18**, 1943–1962, https://doi.org/10.1175/JHM-D-16-0066.1.

——, ——, ——, ——, ——, and ——, 2017b: Predicting U.S. Drought Monitor (USDM) States using precipitation, soil moisture, and evapotranspiration anomalies. Part II: Intraseasonal drought intensification forecasts. *J. Hydrometeor.*, **18**, 1963–1982, https://doi.org/10.1175/JHM-D-16-0067.1.

——, ——, ——, ——, and Y. Zhong, 2018: Forecasting rapid drought intensification using the Climate Forecast System (CFS). *J. Geophys. Res. Atmos.*, **123**, 8365–8373, https://doi.org/10.1029/2018JD028880.

McNaughton, K. G., and T. W. Spriggs, 1986: A mixed-layer model for regional evaporation. *Bound.-Layer Meteor.*, **34**, 243–262, https://doi.org/10.1007/BF00122381.

Meinshausen, N., 2013: Sign-constrained least squares estimation for high-dimensional regression. *Electron. J. Stat.*, **7**, 1607–1631, https://doi.org/10.1214/13-EJS818.

Mitchell, K. E., and Coauthors, 2004: The multi-institutional North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, https://doi.org/10.1029/2003JD003823.

Nicolai-Shaw, N., L. Gudmundsson, M. Hirschi, and S. I. Seneviratne, 2016: Long-term predictability of soil moisture dynamics at the global scale: Persistence versus large-scale drivers. *Geophys. Res. Lett.*, **43**, 8554–8562, https://doi.org/10.1002/2016GL069847.

Norman, J. M., W. P. Kustas, and K. S. Humes, 1995: A two-source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. *Agric. For. Meteor.*, **77**, 263–292, https://doi.org/10.1016/0168-1923(95)02265-Y.

Orth, R., and S. I. Seneviratne, 2013: Predictability of soil moisture and streamflow on subseasonal timescales: A case study. *J. Geophys. Res. Atmos.*, **118**, 10 963–10 979, https://doi.org/10.1002/jgrd.50846.

Otkin, J. A., M. C. Anderson, C. Hain, I. Mladenova, J. Basara, and M. Svoboda, 2013: Examining flash drought development using the thermal infrared based Evaporative Stress Index. *J. Hydrometeor.*, **14**, 1057–1074, https://doi.org/10.1175/JHM-D-12-0144.1.

——, ——, ——, and M. Svoboda, 2014: Examining the relationship between drought development and rapid changes in the Evaporative Stress Index. *J. Hydrometeor.*, **15**, 938–956, https://doi.org/10.1175/JHM-D-13-0110.1.

——, ——, ——, and ——, 2015a: Using temporal changes in drought indices to generate probabilistic drought intensification forecasts. *J. Hydrometeor.*, **16**, 88–105, https://doi.org/10.1175/JHM-D-14-0064.1.

——, and Coauthors, 2015b: Facilitating the use of drought early warning information through interactions with agricultural stakeholders. *Bull. Amer. Meteor. Soc.*, **96**, 1073–1078, https://doi.org/10.1175/BAMS-D-14-00219.1.

——, and Coauthors, 2016: Assessing the evolution of soil moisture and vegetation conditions during the 2012 United States flash drought. *Agric. For. Meteor.*, **218**, 230–242, https://doi.org/10.1016/j.agrformet.2015.12.065.

——, M. Svoboda, E. D. Hunt, T. W. Ford, M. C. Anderson, C. Hain, and J. B. Basara, 2018: Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bull. Amer. Meteor. Soc.*, **99**, 911–919, https://doi.org/10.1175/BAMS-D-17-0149.1.

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1.

Pendergrass, A. G., and Coauthors, 2020: Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Climate Change*, **10**, 191–199, https://doi.org/10.1038/s41558-020-0709-0.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, https://doi.org/10.1175/2010BAMS3001.1.

Slawski, M., and M. Hein, 2013: Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electron. J. Stat.*, **7**, 3004–3056, https://doi.org/10.1214/13-EJS868.

Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, https://doi.org/10.1175/1520-0477-83.8.1181.

Tippett, M. K., A. G. Barnston, and T. DelSole, 2010: Comments on "Finite samples and uncertainty estimates for skill measures for seasonal prediction." *Mon. Wea. Rev.*, **138**, 1487–1493, https://doi.org/10.1175/2009MWR3214.1.

Tobin, K. J., R. Torres, W. T. Crow, and M. E. Bennett, 2017: Multi-decadal analysis of root-zone soil moisture applying the exponential filter across CONUS. *Hydrol. Earth Syst. Sci.*, **21**, 4403–4417, https://doi.org/10.5194/hess-21-4403-2017.

——, W. T. Crow, J. Dong, and M. E. Bennett, 2019: Validation of a new root-zone soil moisture product: Soil MERGE. *IEEE J. Selected Topics Appl. Earth Obs. Remote Sens.*, **12**, 3351–3365, 10.1109/JSTARS.2019.2930946.

Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Wei, H., Y. Xia, K. E. Mitchell, and M. B. Ek, 2013: Improvement of the Noah land surface model for warm season processes: Evaluation of water and energy flux simulation. *Hydrol. Processes*, **27**, 297–303, https://doi.org/10.1002/hyp.9214.

Xia, Y., M. B. Ek, H. Wei, and J. Meng, 2012a: Comparative analysis of relationships between NLDAS-2 forcings and model outputs. *Hydrol. Processes*, **26**, 467–474, https://doi.org/10.1002/hyp.8240.

——, and Coauthors, 2012b: Continental-scale water and energy flux analysis and validation of the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, https://doi.org/10.1029/2011JD016048.

Zhu, S., H. Chen, X. Dong, and J. Wei, 2020: Influence of persistence and oceanic forcing on global soil moisture predictability. *Climate Dyn.*, **54**, 3375–3385, https://doi.org/10.1007/s00382-020-05184-8.