# Improving Subseasonal Soil Moisture and Evaporative Stress Index Forecasts through Machine Learning: The Role of Initial Land State versus Dynamical Model Output

DAVID J. LORENZ,[a] JASON A. OTKIN,[b] BENJAMIN F. ZAITCHIK,[c] CHRISTOPHER HAIN,[d] THOMAS R. H. HOLMES,[e] AND MARTHA C. ANDERSON[f]

[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] *NASA Marshall Space Flight Center, Earth Science Branch, Huntsville, Alabama*
[e] *Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland*
[f] *Hydrology and Remote Sensing Laboratory, USDA, Agricultural Research Service, Beltsville, Maryland*

(Manuscript received 11 May 2023, in final form 25 March 2024, accepted 28 May 2024)

ABSTRACT: The effect of machine learning and other enhancements on statistical–dynamical forecasts of soil moisture (0–10 and 0–100 cm) and a reference evapotranspiration fraction [evaporative stress index (ESI)] on subseasonal time scales (15–28 days) are explored. The predictors include the current and past land surface conditions and dynamical model hindcasts from the Subseasonal to Seasonal Prediction project (S2S). When the methods are enhanced with machine learning and other improvements, the increases in skill are almost exclusively coming from predictors drawn from observations of current and past land surface states. This suggests that operational S2S flash drought forecasts should focus on optimizing use of information on current conditions rather than on integrating dynamically based forecasts, given the current state of knowledge. Nonlinear machine learning methods lead to improved skill over linear methods for soil moisture but not for ESI. Improvements for both soil moisture and ESI are realized by increasing the sample size by including surrounding grid points in training and increasing the number of predictors. In addition, all the improvements in the soil moisture forecasts predominantly impact soil moistening rather than soil drying—i.e., prediction of conditions moving away from drought rather than into drought—especially when the initial soil state is drier than normal. The physical reasons for the nonlinear machine learning improvements are also explored.

SIGNIFICANCE STATEMENT: Rapidly intensifying droughts pose extra challenges for predictability. Here, dynamical forecast model output is combined with nonlinear machine learning methods to improve forecasts of rapid changes in soil moisture and the evaporative stress index (ESI).

KEYWORDS: Drought; Soil moisture; Forecasting; Machine learning

## 1. Introduction

Accurate drought predictions are important early warning tools to help mitigate impacts on agriculture and water resources. Droughts that develop rapidly are especially difficult to forecast because their development time scale is between the short time scales most affected by atmospheric initial conditions and the longer time scales most affected by tropical sea surface temperature (SST) anomalies. Predicting these "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, 2018a, 2022; Christian et al. 2019; Pendergrass et al. 2020) requires different approaches from conventional droughts because the initial state of the land and/or atmosphere is very important for their evolution. For example, the particularly damaging flash drought of 2012 did not appear to be related to SST anomalies (Hoerling et al. 2014); instead, the drought came through "random" weather variability due to the chaotic nature of the atmospheric circulation. DeAngelis et al. (2020) found that skill in predicting the 2012 flash drought was restricted to short time scales (2 weeks) except for a few isolated initialization times when skill was present at 3–4 weeks. They also found that soil moisture initialization and accurate representation of atmospheric Rossby wave trains are essential for accurate 3–4-week forecasts. Hoell et al. (2021) also highlighted the essential role for internal atmospheric variability in the development of Midwest droughts except for long forecast lead times for the Ohio Valley.

Soil moisture and plant evaporative stress (see below) have shown particular promise as predictors of developing flash drought (Otkin et al. 2014, 2015a). In addition, soil moisture and evaporative stress are two of the indicators essential for monitoring agricultural drought. Therefore, to help further improve flash drought early warning, this study investigates the predictability of soil moisture and evaporative stress over the contiguous United States and adjacent regions of Canada and Mexico. The particular focus is on improving the hybrid statistical–dynamical forecasts developed in Lorenz et al. (2021) by utilizing machine learning. Lorenz et al. (2021) predicted soil moisture and a reference evapotranspiration

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

DOI: 10.1175/JHM-D-23-0074.1

fraction [evaporative stress index (ESI)] (Anderson et al. 2007a,b, 2011, 2013) using linear regression with predictors drawn from the current land surface and meteorological state and dynamical forecast model output from the Subseasonal to Seasonal Prediction project (S2S) (Vitart et al. 2017). They found that spatial variations in skill across the United States are tightly coupled with the autocorrelation of the predicted variable. They also found that S2S model forecasts provided significant benefits particularly for soil moisture. In the current study, we aim to improve the forecasts in Lorenz et al. (2021) using machine learning methods. In addition, the regressions in Lorenz et al. (2021) were fit locally and independently of information at other grid points. In this study, we explore the benefits of including surrounding grid points to increase sample size while still maintaining statistical methods that are tailored to each specific grid point. We find that increasing the sample size allows a larger number of predictors and increased skill. In the current study, we focus on the 15–28-day time scale, which was motivated by the prediction of flash droughts.

There are multiple studies that use machine learning to "predict" drought (e.g., Deo and Şahin 2015; Khan et al. 2020; Dikshit et al. 2020; Dikshit and Pradhan 2021; Zhu and Wang 2021); however, in these studies, "prediction" is used in a statistical sense but not a true forecast sense since the predictors and predictand overlap temporally. Brust et al. (2021) use machine learning in a true forecast setting to predict the U.S. Drought Monitor (USDM) (Svoboda et al. 2002). However, the USDM drought depiction may lag conditions on the ground by up to several weeks (Ford et al. 2015), and therefore, USDM predictions may not represent a true forecast setting in the same way as an index that responds immediately to conditions on the ground. Indeed, Lorenz et al. (2018) found very little improvement by including dynamical forecast model output in the USDM forecasts compared to the USDM forecasts that only used current/past conditions (Lorenz et al. 2017). For soil moisture forecasts, on the other hand, Lorenz et al. (2021) found that including output from the dynamical models doubled the skill for 14-day forecasts.

Several studies have used machine learning for subseasonal forecasts of surface air temperature and/or precipitation (Hwang et al. 2019; Wang et al. 2021; He et al. 2021, 2022). He et al. (2021) compared 10 machine learning methods for 3–4-week temperature forecasts and found that gradient boosting machine (Friedman 2001) performed the best overall. However, linear regression with least absolute shrinkage and selection operator (LASSO) (Tibshirani 1996; Jalali et al. 2013) performed almost as well. He et al. (2022), Slater et al. (2023), and Cao et al. (2022) combined machine learning with dynamical models by including dynamical model output in the suite of predictors.

A major focus of this study is comparing the nonlinear machine learning methods for predicting soil moisture and ESI with the linear method described in Lorenz et al. (2021). We will also outline a strategy for increasing sample size so that more predictors can be used, and machine learning can be more effective. Throughout, the linear methodology in Lorenz et al. (2021) will serve as a baseline for assessing the value of nonlinear machine learning and other enhancements.

We begin by describing the dataset and methodology in sections 2 and 3, respectively. Next, we discuss the effect of our enhancements on the baseline skill in Lorenz et al. (2021) and the relative role of initial/past state versus future dynamical model forecasts on the improvements relative to Lorenz et al. (2021). Then, we discuss the asymmetric impact of the improvements on the positive/negative tails of the error PDF of soil moisture. Finally, we end with the conclusions of the analysis.

## 2. Data

### a. ESI

The ESI represents standardized anomalies in the ratio of actual evapotranspiration (ET) to reference evapotranspiration (RET) (Anderson et al. 2007a,b, 2011, 2013). Actual ET is diagnosed from time changes in land surface temperature, typically retrieved from thermal infrared satellite imagery, using the Atmosphere–Land Exchange Inverse model (ALEXI) surface energy balance model (Anderson et al. 1997, 2007a), while reference ET is quantified by the FAO Penman–Monteith formulation (Allen et al. 1998) using surface meteorological data from Climate Forecast System Reanalysis (Saha et al. 2014). The ESI reflects the impacts of soil moisture on changes in vegetation transpiration and soil evaporation rates and has been demonstrated to be predictive of agricultural drought impacts on crop condition and yield (Anderson et al. 2016a,b). Otkin et al. (2014, 2015a) demonstrated the rapid response of ESI during flash drought events and showed that rapid changes in ESI can anticipate the development of flash drought.

To calculate the ESI, anomalies of the ratio of ET to RET are standardized at the pixel level by the seasonal cycle of its mean and standard deviation. In this study, the daily varying seasonal cycle is smoothed with a parabolic-shaped weight function of the form $w_j = (n + 1)^2 - j^2$, for $j = -n, -n + 1, \ldots, n$. We use $n = 15$, which implies the smoothing window is 31 days wide.

Although the smoothed ESI is computed at daily time steps, domain coverage on any given day is incomplete due to cloud cover. Therefore, we use 7-day running-mean ESI in this study. For our application, 7 days is a good compromise between temporal resolution and smoothing. The spatial resolution is $0.04° \times 0.04°$.

### b. Land-cover classification

For increasing sample size, the degree of similarity between nearby grid points needs to be quantified. To this end, we use the 30 m × 30 m resolution land-cover classification dataset of Chen et al. (2015). These data are upscaled to the ESI grid by calculating the fractional coverage of each of eight land-cover types within each grid box. The eight land-cover types are bare, crop, forest, grass, shrub, urban, water, and wetland.

### c. Noah soil moisture

For soil moisture, we use estimates from the Noah model (Ek et al. 2003; Barlage et al. 2010; Wei et al. 2013) of the North American Land Data Assimilation System, version 2

(NLDAS2; Mitchell et al. 2004; Xia et al. 2012a,b). We predict both the 0–10- and 0–100-cm-depth soil moisture in this study. The soil moisture at 100–200 cm is also used as a predictor. The hourly Noah fields are averaged to daily values prior to all analyses in this study. The spatial resolution is 0.125° × 0.125°. Anomalies from the mean seasonal cycle are calculated as in ESI, except we do not normalize by the standard deviation for soil moisture. Due to issues with the precipitation forcing of NLDAS over Mexico and Canada, analyses involving spatial averages are restricted to the United States.

### d. ECMWF dynamical forecast model

Our statistical forecasts also use hindcasts from the S2S project (Vitart et al. 2017) as predictors. For this study, we use the European Centre for Medium-Range Weather Forecasts (ECMWF) model because this model has archived hindcasts for most days since 1999. In addition, unlike most other S2S models, the ECMWF model does not have a gap between the ends of the retrospective period (2010) until the start of the real-time period (2015).

The forecast length of the ECMWF model is 46 days with an output frequency of either daily or 6-hourly depending on the variable. All 6-hourly variables are converted to daily values prior to use as predictors. For this study, we use the mean of the 11 ensemble members. We explored using the individual ensemble members rather than simply the mean to potentially increase the sample size; however, this did not improve skill. Lorenz et al. (2021) explored the skill of the following output fields: precipitation, maximum and minimum daily temperature, mean daily temperature, dewpoint temperature, sensible and latent heat flux, net shortwave radiation at surface, soil moisture, and linear combinations of certain temperature variables. They found that precipitation and dewpoint depression (maximum 24-h temperature minus dewpoint temperature) are the best overall predictors, and only these predictors are considered further in this study. The ECMWF model resolution is T639 before day 16 and T319 at longer lead times.

To help remove systematic model biases, the mean seasonal cycle of the ECMWF hindcasts is removed for each initialization time and forecast lead time. The mean seasonal cycle is smoothed in the same way as the soil moisture before it is subtracted from the raw ECMWF hindcast fields.

### e. Spatial and temporal information

The ECMWF data are archived at 1.5° × 1.5° resolution. Because the predictable spatial scales are likely coarser than the relatively fine resolution of the ESI and soil moisture data, all fields are interpolated to an intermediate resolution 0.4° × 0.4° grid prior to analysis. We use bilinear interpolation for the ECMWF data, and we average the fine-resolution ESI and Noah soil moisture data over each 0.4° × 0.4° grid box. The results are not sensitive to the resolution of the intermediate grid (not shown). The spatial domain is the contiguous United States and the surrounding regions of Canada and Mexico. We forecast for the warm season (1 May–30 September) only since flash droughts are most common during this time (Christian et al. 2019). The length of the ESI and ECMWF datasets limits the time period of our hindcasts, which are 2000–18 for ESI and 1999–2018 for soil moisture. For the remainder of the paper, we use the term forecasts rather than hindcasts to describe the ECMWF data and our machine learning predictions even though all results and skill scores are evaluated on historic events.

## 3. Methodology

### a. Basic linear method

The linear methodology in this study is the same as in Lorenz et al. (2021). At each grid point in the domain, linear regression is used to predict either soil moisture or ESI anomalies from initial state anomalies and ECMWF forecast anomalies. Everything about the linear regression is standard except for a sign constraint on the regression coefficients. A sign constraint works well because we know a priori that increased precipitation and decreased dewpoint depression are associated with increased soil moisture and ESI. The sign constraint greatly improves the robustness of the regressions, much like alternative regularization methods such as ridge or LASSO regression (Meinshausen 2013; Slawski and Hein 2013). Signed constrained linear regression is implemented using cyclic coordinate descent (Franc et al. 2005). Cyclic coordinate descent works well because it is more efficient than active-set methods (Lawson and Hanson 1995) and it is easy to relax the sign constraint for certain predictors but not all. For the "basic" linear forecast of variable $x$, $n$ days in advance, we use the variable $x$ at the initial time and an ECMWF forecast variable out to day $n$. As discussed in Lorenz et al. (2021), in the absence of other forcing, one expects $x$ to relax back toward climatology; therefore, the sign of this regression coefficient is also known a priori and is negative when predicting the change in $x$ (future $x$ anomaly minus current $x$ anomaly).

For the basic forecast, we use ECMWF precipitation for the soil moisture variables and ECMWF dewpoint depression for ESI because these are the best-performing predictors in each case. Lorenz et al. (2021) found marginally better skill for some grid points using both precipitation and dewpoint depression for both soil moisture and ESI, but this improvement is not consistent across the spatial domain; therefore, a single ECMWF predictor is used here. Below, we will develop a strategy to increase the effective sample size and we will find that additional ECMWF predictors are beneficial for both soil moisture and ESI when sample size is increased. The land water budget states that the change in soil moisture is precipitation minus evaporation and runoff. Because precipitation directly forces soil moisture and moreover a significant portion of the variability in evaporation and runoff is associated with current and past precipitation, it is not surprising that precipitation anomalies have the most impact on soil moisture. For ESI, we expected the dewpoint depression to have a larger role based on the analysis of Otkin et al. (2018b).

As in Lorenz et al. (2021), the variable at each individual day of the ECMWF forecast is a separate predictor. Therefore, the basic 15-day forecast for soil moisture has 16 predictors: the initial soil moisture and 15 predictors for precipitation at each

day of the 15-day ECMWF forecast. The sign-constrained regression enables robust cross-validated skill despite the large number of predictors. For the basic forecast, an independent statistical fit is performed separately for each grid point in the domain as in Lorenz et al. (2021). In this study, we show results for 15- and 28-day forecasts of soil moisture. For ESI, which is a weekly composite, we predict the change from the −6 to 0-day composite to the 12–18- or 25–31-day composite. We use fivefold cross validation to evaluate skill on independent data. When the number of years in the data is not a multiple of 5, we still force the dividing point between cross-validation folds to be between years by allowing the length of the folds to vary. This ensures maximum independence between samples. Mean-square error is used to evaluate skill. All results shown below are cross validated in the sense that the statistical model has not seen the data used for evaluation.

Finally, for persistent predictands, the skill scores of a forecast of the future *change* in the predictand (i.e., anomaly at day $n$ minus the anomaly at the initial time) are more meaningful than skill scores of a prediction of the future anomaly itself (Lorenz et al. 2021). This is because forecasts of the change de-emphasize the trivial component of skill associated with high persistence. Of course, ultimately both forecasts give the same information. The 0–100-cm soil moisture is very persistent and so we forecast the future change in this variable. Lorenz et al. (2021) showed that to optimally de-emphasize the trivial component of skill, one should forecast the anomaly change when the predictand autocorrelation is greater than 0.5 and forecast the anomaly itself when the predictand autocorrelation is less than 0.5. ESI persistence is weaker than 0–100-cm soil moisture, and some regions are optimal with the future change (autocorrelation $> 0.5$), and some are optimal with the future anomaly (autocorrelation $< 0.5$). We choose to forecast the change in ESI in this study because the change in ESI is more meaningful in the central United States where flash droughts are more frequent (Christian et al. 2019). For 0–10-cm soil moisture, on the other hand, persistence is relatively weak, and therefore, we predict the future anomaly in this study. All enhanced forecasts use the same basic forecast choice established here regarding the future anomaly versus the change in the anomaly.

### b. Enhancements

Like the basic model, all enhancements use anomalies from the mean seasonal cycle for both predictors and predictands. We explore the effect of three potential enhancements to the baseline linear approach used by Lorenz et al. (2021): 1) machine learning, which is discussed in detail in the next subsection; 2) additional predictors; and 3) increasing sample size by including surrounding grid points (ISGPs) in the training of the statistical model. The testing of the model via cross validation is still performed on the central grid point alone so all skill scores reflect the localized skill even when ISGP is invoked. In some sense, ISGP is the most important enhancement because machine learning and additional predictors do not add skill unless sample size is increased via ISGP. We explored various distance radii for defining which surrounding grid points to include and found that an 8° longitude and latitude radius is optimal averaged over the domain. We also explored various schemes for weighting surrounding grid points in the statistical models so that grid points that are more like the central grid point carry more weight. We tested weighting based on land-cover classification and on predictand autocorrelation at the forecast lead time. The autocorrelation used for weighting is also cross validated because otherwise skill is artificially enhanced (not shown). For autocorrelation, each grid point is weighted:

$$w = \max(1 - 2|a^2 - a_0^2|, 0),$$

where $a_0$ is the autocorrelation of the central grid point and $a$ is the autocorrelation of the grid point being weighted. The weight $w$ ranges from 1 when the autocorrelation of the surrounding grid point matches the central grid point, and the maximum function prevents negative weights. We use the square of the autocorrelation because it is proportional to the percent variance explained. For land cover, let $L_j$ be the fractional coverage of eight land-cover types ($j = 1, 2, \ldots, 8$) at a grid point: bare, crop, forest, grass, shrub, urban, water, and wetland (Chen et al. 2015). The land-cover weight is

$$w = \max[1 - \sum_j (L_j - L_{0j})^2, 0],$$

where $L_{0j}$ is the land-cover fraction for the central grid point and the maximum function prevents negative weights. We find that autocorrelation weighting works best for soil moisture and land-cover classification works best for ESI.

When sample size is increased via ISGP, additional predictors lead to improvement on cross-validated skill. We now use both precipitation and dewpoint depression from the ECMWF dynamical model because together they lead to consistent improvement for both soil moisture and ESI. In addition, we explored adding various combinations of the current and past anomalies of the soil moisture at all depths, the ESI, precipitation, and dewpoint depression. Unlike the predictors discussed previously, these additional predictors do not have a sign constraint on the regression coefficient because the sign is not known a priori. For both 0–10- and 0–100-cm soil moisture, we settle on the following predictors: the precipitation and the soil moisture at 0–10, 0–100, and 100–200 cm on the current day (day 0) and the same variables averaged over the previous week (from days −1 to −7). Averaged over the domain, these predictors lead to the best overall cross-validated skill. Using the ESI or current dewpoint depression as predictors for future soil moisture did not improve skill. A summary of the soil moisture predictors for the more "advanced" models with additional predictors is provided in Table 1. For ESI, we settle on the latest weekly composite of ESI (from days 0 to −6) as well as two prior weekly composites of ESI (from days −7 to −13 and days −14 to −20). In addition, we use the latest day's soil moisture at 0–10, 0–100, and 100–200 cm. The ESI forecasts include the ESI for the previous 3 weeks, and the signs of the fitted regression coefficients are consistent across the domain: The coefficient for the latest week is negative (like the basic forecast), the

TABLE 1. Full set of predictors used for the more advanced soil moisture forecasts (both 0–10 and 0–100 cm). All predictors are anomalies from the mean seasonal cycle. The last three columns state the time period for each predictor. Note that the mean anomaly from days −7 to −1 and the 0-day anomaly are distinct predictors. Current and past predictors are from NLDAS, and future predictors are from the ECMWF S2S model reforecasts. The "*n*" in the last column denotes the number of days of the forecast. Note that the basic forecast uses only the current anomaly of the predictand and the ECMWF precipitation forecast.

| Predictors | Past (NLDAS) | Current (NLDAS) | Future (ECMWF) |
|---|---|---|---|
| 0–10-cm soil moisture | Mean from day −7 to −1 | Day 0 | |
| 0–100-cm soil moisture | Mean from day −7 to −1 | Day 0 | |
| 100–200-cm soil moisture | Mean from day −7 to −1 | Day 0 | |
| Precipitation | Mean from day −7 to −1 | Day 0 | |
| Precipitation | | | Days 1 to *n* |
| Dewpoint depression | | | Days 1 to *n* |

−7 to −13-day composite is positive, and the −14 to −20-day composite is negative. Temporally alternating sign coefficients are preferentially triggered by high-frequency signals, and perhaps, these coefficients are empirically "tuned" to help identify and correct for high-frequency temporal noise in the ESI time series. A summary of the ESI predictors is provided in Table 2.

### c. *Machine learning*

In this study, we use the term "machine learning" for methods that can capture nonlinear relationships. We explore the following machine learning methods in this research: random forest (RF; Breiman 2001), gradient boosting machine (GBM; Friedman 2001), artificial neural network (ANN), multivariate adaptive regression spline (MARS; Friedman 1991), generalized additive model (GAM; Hastie and Tibshirani 1986), and support vector machine (SVM; Cortes and Vapnik 1995). In future work, we will also explore long short-term memory (LSTM) networks (Hochreiter and Schmidhuber 1997), which are well suited for forecasting time series. RF, ANN, and SVM were implemented using the scikit-learn package (Pedregosa et al. 2011). The remaining methods were implemented using py-earth (MARS), pyGAM (Servén and Brummitt 2018; GAM), and LightGBM (Ke et al. 2017; GBM). We also implemented GBM with scikit-learn and extreme gradient boosting (XGBoost) (Chen and Guestrin 2016) but found that LightGBM was faster and gave slightly better results. We found that GBM performed best on average, but ANN was close. This is consistent with the extensive survey of regression methods in Fernández-Delgado et al. (2019). Januschowski et al. (2022) hypothesize that tree-based methods like GBM and RF are successful because they are more robust than neural networks, which require skillful tuning of model structure and parameters. GBM is an ensemble of regression trees like RF. Whereas RF builds each tree independently of the others, GBM incrementally adds trees which correct the errors of the previous trees. Because GBM more directly minimizes the error, it can potentially find a better solution than RF, but it is also more prone to overfitting. For the remainder of the study, we only discuss results for the GBM machine learning method and how it compares with the linear method described previously.

For GBM, the maximum depth of the regression trees and the number of trees are the two hyperparameters that are tuned to optimize machine learning skill. The default values of these parameters are 5 and 100, respectively. He et al. (2021) find that the relatively short length of the observed climate record is a serious limitation on the performance of nonlinear machine learning methods for statistical forecasting; therefore, we try hyperparameter values less than or equal to the default. For both hyperparameters, lower values increase the robustness and decrease the flexibility of GBM. For the maximum depth, we try the values 3, 4, and 5. For the number of trees, we try 50 and 100. As mentioned above, the linear method uses fivefold cross validation where 80% of the data are used for training and 20% for testing and the process is repeated five times so that all data have a chance to be in the testing set. For GBM, we use nested cross validation for hyperparameter tuning. Specifically, the 80% training set is further divided into two halves. For each hyperparameter combination, we train on the first half and validate on the second half. Next, we train on the second half and validate on

TABLE 2. Full set of predictors used for the more advanced ESI forecasts. All predictors are anomalies from the mean seasonal cycle. The last three columns state the time period for each predictor. Note each current and past ESI weekly composite is a distinct predictor which gives three ESI predictors. Current predictors are from NLDAS except for ESI, and future predictors are from the ECMWF S2S model reforecasts. The *n* in the last column denotes the number of days of the forecast. Note that the basic forecast uses only the current anomaly of the predictand and the ECMWF dewpoint depression forecast.

| Predictors | Past | Current | Future (ECMWF) |
|---|---|---|---|
| ESI | Weekly composite at lags −1 and −2 weeks | Latest weekly composite | |
| 0–10-cm soil moisture | | Day 0 | |
| 0–100-cm soil moisture | | Day 0 | |
| 100–200-cm soil moisture | | Day 0 | |
| Precipitation | | | Days 1 to *n* |
| Dewpoint depression | | | Days 1 to *n* |

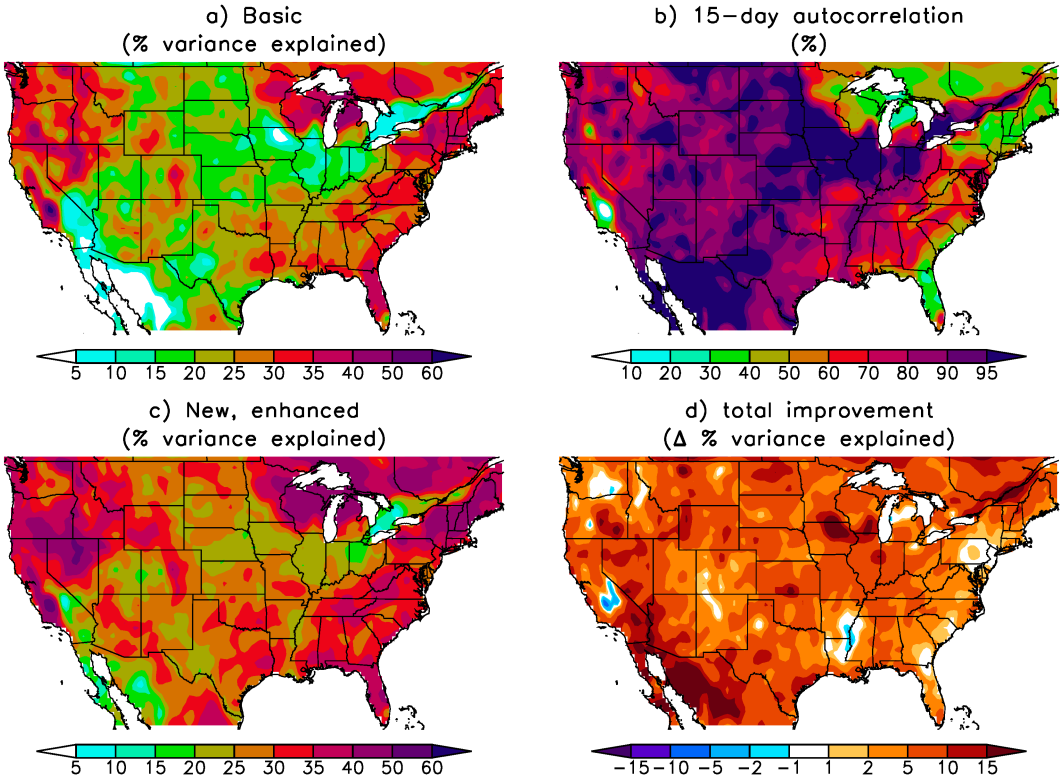


FIG. 1. (a) Cross-validated percent variance explained by the basic forecast (see text) of the change in 0–100-cm soil moisture at day 15. (b) Autocorrelation of 0–100-cm soil moisture at day 15. (c) As in (a), but for the enhanced forecasts with ISGP, additional predictors, and GBM. (d) Change in percent variance explained: (c) minus (a).

the first half. As before, break points between halves are forced to lie between years for maximum independence. The hyperparameter combination with the lowest mean square error over the validation data is then trained on the full 80% training set and then used to predict the 20% testing set. Finally, this procedure is repeated five times so that all data have a chance to be in the testing set. In addition, the predictors with a positive or negative sign constraint for the linear method are given a monotonically increasing or decreasing constraint for GBM, respectively. For all other hyperparameters, we use the default values.

In addition, the machine learning methods we have tried do not perform as well as the sign-constrained linear method when given the full set of predictors, which is again likely related to the relatively small sample size. Therefore, we first fit the linear method and then use some of the regression coefficients as weights for dimension reduction. For example, if $a_j$ and $x_j$ are the regression coefficients and predictors, respectively, for the soil moisture $S$, then the linear method is

$$S = a_0 + \sum_{j=1}^{n} a_j x_j,$$

where $n$ is the number of predictors. For dimension reduction, we combine predictors together that are the same variable using weights from the linear method. For example, the 0–10-cm soil moisture at the initial time (day 0) and the previous week (mean from day −1 to −7) are two separate predictors for the linear method. For machine learning, they are summed together into a single predictor using the $a_j$ coefficients from the linear method. Similarly, a 15-day forecast with the linear method uses the precipitation for each of the 15 forecast days as a separate predictor. For machine learning, these 15 predictors are combined into a single predictor. After dimensionality reduction, six predictors are used for the soil moisture forecasts (see Table 1). For ESI, machine learning struggled to improve on the linear method, and therefore, we combined all current and past predictors together into a single predictor for machine learning. This gives either two or three total predictors after including the future dewpoint depression and/or precipitation. Note that the predictor reduction scheme uses the linear method itself to perform the reduction, so it is not possible to reduce predictors for the linear method in this way.

## 4. Results

### a. *Soil moisture*

The cross-validated skill of forecasts of the 15-day change in soil moisture (0–100 cm) is shown in Fig. 1. We predict the future change in soil moisture (future soil moisture anomaly minus current soil moisture anomaly) rather than the future anomaly itself because predictions of the change de-emphasize the trivial component of skill associated with the large persistence of

soil moisture. Forecasts of the change in soil moisture, on the other hand, place greater emphasis on the nontrivial components of skill (Lorenz et al. 2021). The basic forecast (Fig. 1a) performs a separate statistical fit for the time series at each grid point. The predictors are the initial 0–100-cm soil moisture anomaly and the precipitation from the ECMWF S2S dynamical forecast model. The methodology of the basic forecast is the same as Lorenz et al. (2021) except only the precipitation from ECMWF is used. Adding additional ECMWF predictors does not improve the domain-averaged skill in this case.

The cross-validated percent variance explained in the basic forecast case (Fig. 1a) is largest in the northern parts of the upper Midwest, the Eastern Seaboard, and the northwest United States. The skill is smallest in the Corn Belt and parts of the southwest United States and northwest Mexico. Lorenz et al. (2021) found that the spatial variations in skill are closely related to spatial variations in the soil moisture autocorrelation (Fig. 1b). For predictions of the future anomaly of soil moisture, skill increases with increased autocorrelation. Here, we predict the change in the soil moisture anomaly, and therefore, skill decreases with increased autocorrelation (Lorenz et al. 2021). For example, the soil moisture autocorrelation in the Corn Belt is larger than in any other region (Fig. 1b), and therefore, skill is weak in the Corn Belt (Fig. 1a). On the other hand, the soil moisture autocorrelation is small in the northern parts of the upper Midwest and in New England and skill is large in these regions. Small autocorrelation means the soil moisture has a strong tendency to relax toward climatology in the absence of opposing "forcing" from the meteorology. In this case, the depletion of soil moisture via evaporation, runoff, and/or percolation into deeper soil layers must be strongly correlated with the soil moisture anomaly itself, and therefore, the change in soil moisture is relatively easy to predict from the initial soil moisture anomaly alone. In the case of large autocorrelation, on the other hand, the change in soil moisture is dominated by future meteorology rather than the natural tendency of soil moisture anomalies to relax toward climatology (Lorenz et al. 2021). Recall that by subtracting the present anomaly from the future anomaly, one removes the persistent component of soil moisture and emphasizes the processes trying to change soil moisture such as precipitation and evaporation. Because future meteorology is more difficult to forecast, the skill in regions of large autocorrelation is relatively small.

The enhanced forecast case includes the improvements discussed in detail in section 3. The percent variance explained for the enhanced forecasts is shown in Fig. 1c, and the change in percent variance explained compared to the basic forecasts is shown in Fig. 1d. The skill improves almost everywhere, and in many cases, the improvements are largest where the skill is smallest in the basic case. For example, large improvements up to and exceeding 15% of the variance explained are seen in northeast Iowa and adjacent parts of Wisconsin and Illinois, in the St. Lawrence River Valley, and in parts of the southwest United States and northwest Mexico. These same regions have particularly small skill in the basic forecasts. Even outside these regions, most grid points show substantial improvements in skill. In the appendix, we compare our forecasts with the ECMWF soil moisture forecast.

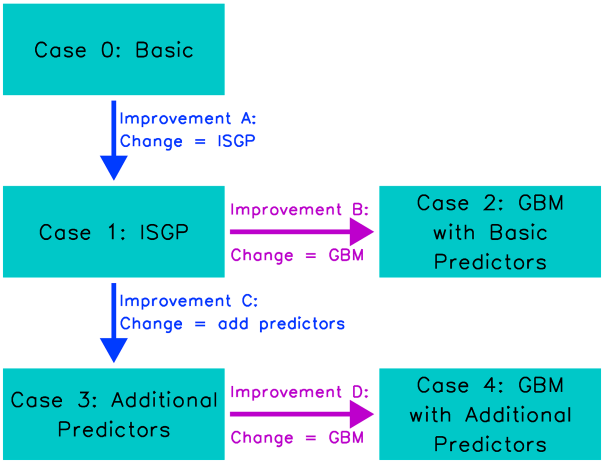


FIG. 2. Schematic of the relationship between the statistical models (boxes) considered in this study and the changes between statistical models (arrows). The purple arrows denote a change from linear to GBM.

Next, we look at the role of ISGP, additional predictors, and machine learning separately and incrementally on the forecast skill. A schematic of the individual statistical models shown here and the relationships between the models is shown in Fig. 2. The different statistical models are represented by squares, and the various improvements from one model to another model are represented by arrows. As mentioned previously, additional predictors and/or GBM do not improve the basic model unless ISGP is implemented first; therefore, we only include a single arrow emanating from the basic model (case 0), which represents the improvement from adding ISGP (improvement A). From the model with ISGP as the only improvement, we explore the role of either GBM (improvement B) or additional predictors (improvement C). Finally, we consider adding GBM to the model with ISGP and additional predictors (improvement D).

A map of the change in percent variance explained via ISGP (improvement A) is shown in Fig. 3a. Most grid points show significant improvement, but a few small regions, particularly in the west, show decreased skill. When using GBM (Fig. 3b), the improvements are distinctly more localized and regions of decreased skill are more pronounced. Also, the largest improvements tend to occur in the central United States between the Rockies and the Appalachians and in Mexico. Adding more predictors to the ISGP forecasts (Fig. 3c) produces the most consistent improvement across the domain. The largest improvements in this case are in parts of the southwest United States and northwest Mexico. When adding GBM to the additional predictors (Fig. 3d), the regions of increased skill are the most localized compared to the other cases. In the United States, the improvements are mostly in the upper Midwest, New England, and Texas. Even though GBM leads to overall improvement, localized regions of decreased skill exist. Apparently, GBM is overfitting in these cases, perhaps because

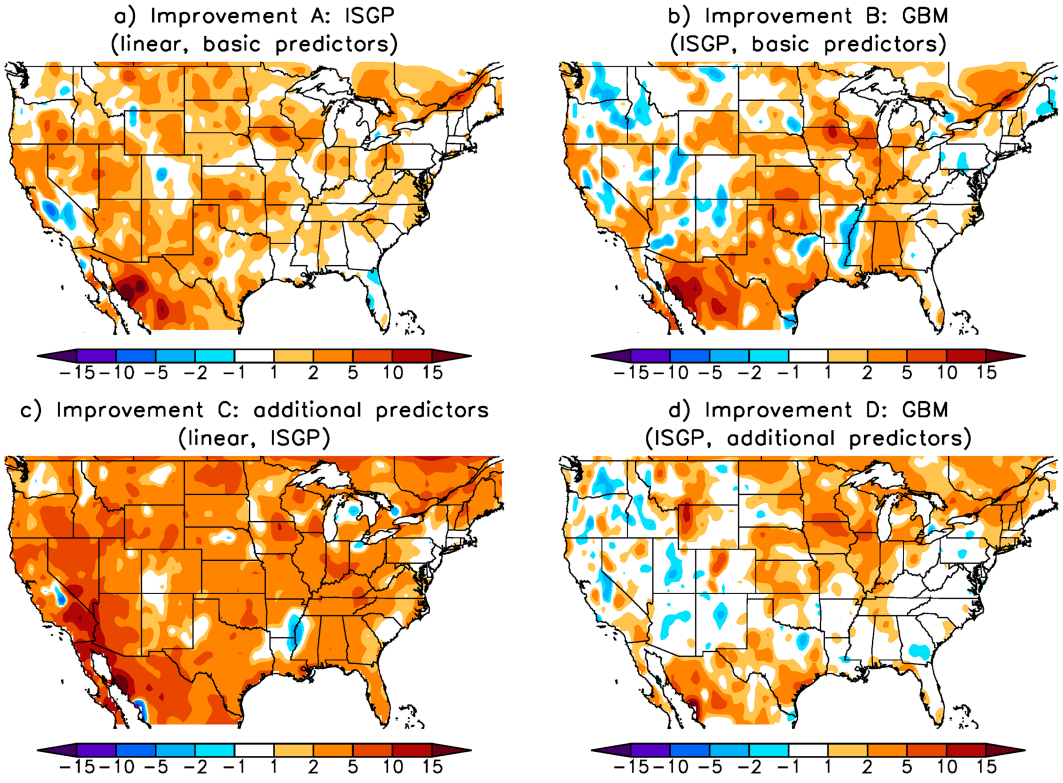


FIG. 3. (a) The change in cross-validated percent variance explained for the 15-day change in 0–100-cm soil moisture when going from the basic forecast to ISGP (improvement A; see Fig. 2). (b) As in (a), but for going from ISGP to ISGP and GBM. (c) As in (a), but for going from ISGP to ISGP and additional predictors. (d) As in (a), but for going from ISGP and additional predictors to ISGP, additional predictors, and GBM.

the signal-to-noise ratio is small in these regions. Note that Fig. 3d shows the change relative to Fig. 3c, so GBM with more predictors is more skillful than GBM with fewer predictors (Fig. 3b) because the improvements in Fig. 3c add to Fig. 3d to get the total improvement. In other words, the changes in Fig. 3 reflect the arrows given in the schematic in Fig. 2. In this way, each subplot is evaluating a single change in the methodology.

The total height of the bars in Fig. 4 shows the percent variance explained for each experiment averaged over the central United States only (between 104° and 85°W excluding Canada and Mexico). We focus on the central United States because the frequency of flash droughts tends to be largest in this region (Christian et al. 2019). Both 15- and 28-day forecasts are shown. The skill steadily increases with increasing case number. Note that the linear model with additional predictors has more skill than GBM with the basic predictor set. The red bars show the skill from a separate statistical forecast using only the initial/past predictors, and the blue bars show the contribution of future prediction from the ECMWF dynamical model, which is defined here as the difference between the total forecast skill and the skill from the initial/past state. For both 15- and 28-day forecasts, the contribution of the dynamical model is almost unchanged from case to case and instead almost all the improvements are coming from the initial/past predictors. This suggests that operational S2S flash drought forecasts should focus on optimizing use of information on current conditions rather than on integrating dynamically based forecasts, given the current state of knowledge. However, this does not mean future meteorology is not important. For example, if we use the actual observed future precipitation and dewpoint depression rather than that from the dynamical S2S model, then the skill increases dramatically to over 90% variance explained (not shown). So, significant improvements via better meteorological forecasts might be possible if a dynamical model is significantly less skillful than the predictability limit of weather (Zhang et al. 2019). It is also possible that alternative machine learning methods would be able to better utilize dynamical model output.

In Fig. 5, we repeat the summary analysis over the central United States for the 0–10-cm soil moisture. We predict the future anomaly of the 0–10-cm soil moisture rather than the change because the 0–10-cm soil moisture autocorrelation is weak, and therefore, the anomaly forecast de-emphasizes the trivial component of skill (see Lorenz et al. 2021). Unlike the deep soil moisture, GBM does not improve skill once additional predictors are added, especially for the 28-day forecasts (i.e., the skill for case 4 is less than or equal to case 3). In addition, the skill from the dynamical model varies more from case to case for the 0–10-cm soil moisture. However, the increased skill of the "best" model (case 4) is almost exclusively from the initial/past predictors, which is the same as found for the 0–100-cm soil moisture.

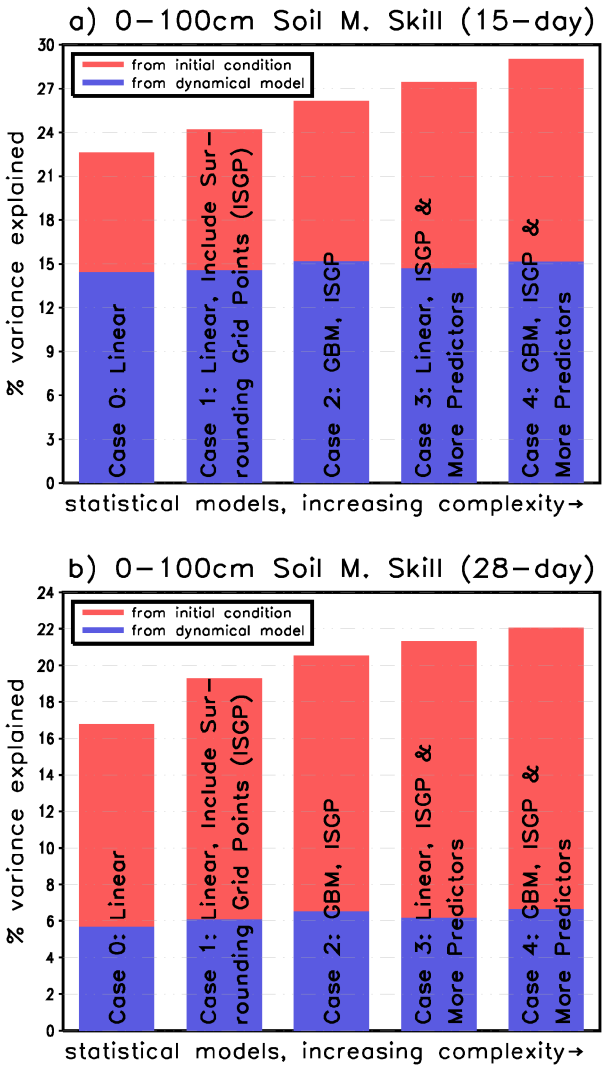


FIG. 4. (a) Central U.S. skill for the 15-day change in 0–100-cm soil moisture as a function of a hierarchy of statistical models: cases 0–5 (see Fig. 2 and text). The skill is partitioned into that from the initial state (red, no dynamical model predictors) and from the dynamical model (additional skill when dynamical model predictors are added). The central U.S. skill is defined as the percent variance explained averaged from 104° to 85°W excluding grid points in Canada and Mexico. (b) As in (a), but for the 28-day change in soil moisture.

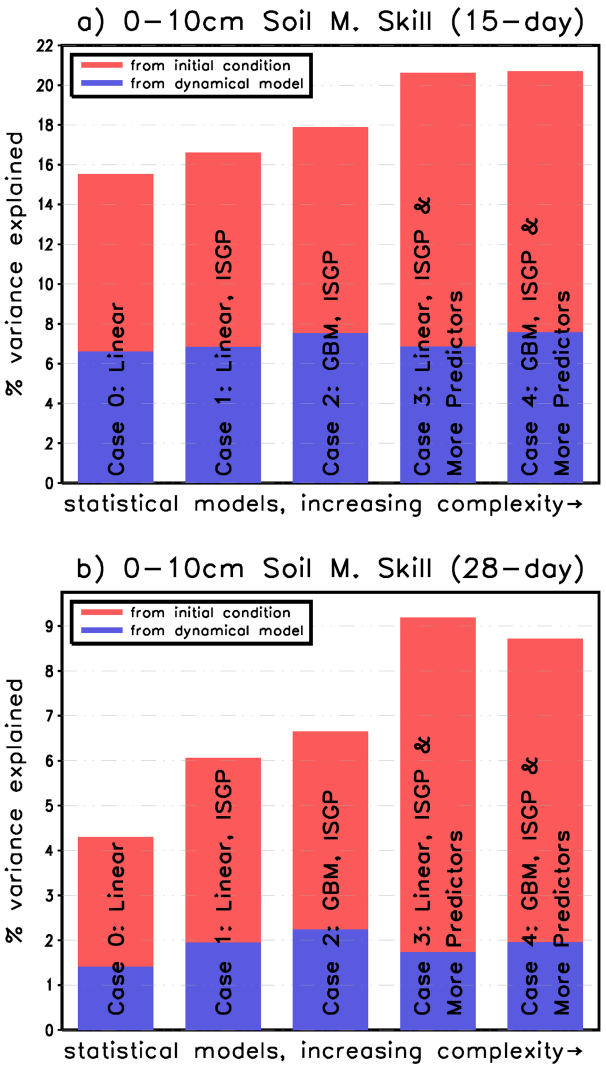


FIG. 5. As in Fig. 4, but for forecasts of the future anomaly of shallow (0–10 cm) soil moisture.

### *b. ESI*

The cross-validated skill of forecasts of the change in ESI is shown in Fig. 6. Because we use 7-day composite ESI, we forecast the change from the −6 to 0-day average ESI to the 12–18-day average ESI, which is roughly analogous to a 15-day forecast. In the central United States, the ESI autocorrelation (Fig. 6b) tends to be greater than the 0.5 threshold identified in Lorenz et al. (2021), and therefore, we forecast the change in the ESI rather than the anomaly. The basic forecast (Fig. 6a) performs a separate statistical fit for the time series at each grid point. The predictors are the initial ESI anomaly and the dewpoint depression from the ECMWF S2S dynamical forecast model. This basic forecast is the same as Lorenz et al. (2021) except only the dewpoint depression from ECMWF is used. Adding additional ECMWF predictors does not improve skill in this case.

As in the soil moisture case, the forecast skill tends to be largest where the autocorrelation is smallest and vice versa. The percent variance explained for the enhanced forecasts is shown in Fig. 6c. For ESI, GBM does not improve skill, and therefore, the enhanced forecasts shown here use the linear method with ISGP and additional predictors (case 3, see below for more details). The change in percent variance explained compared to the basic forecasts is shown in Fig. 6d. The skill increases almost everywhere, and like the soil moisture case, the improvements tend to be largest where the skill is smallest in the basic case. For example, southern Iowa and eastern Missouri have relatively small skill in the basic case (Fig. 6a) but have among the largest increases in skill (Fig. 6d).

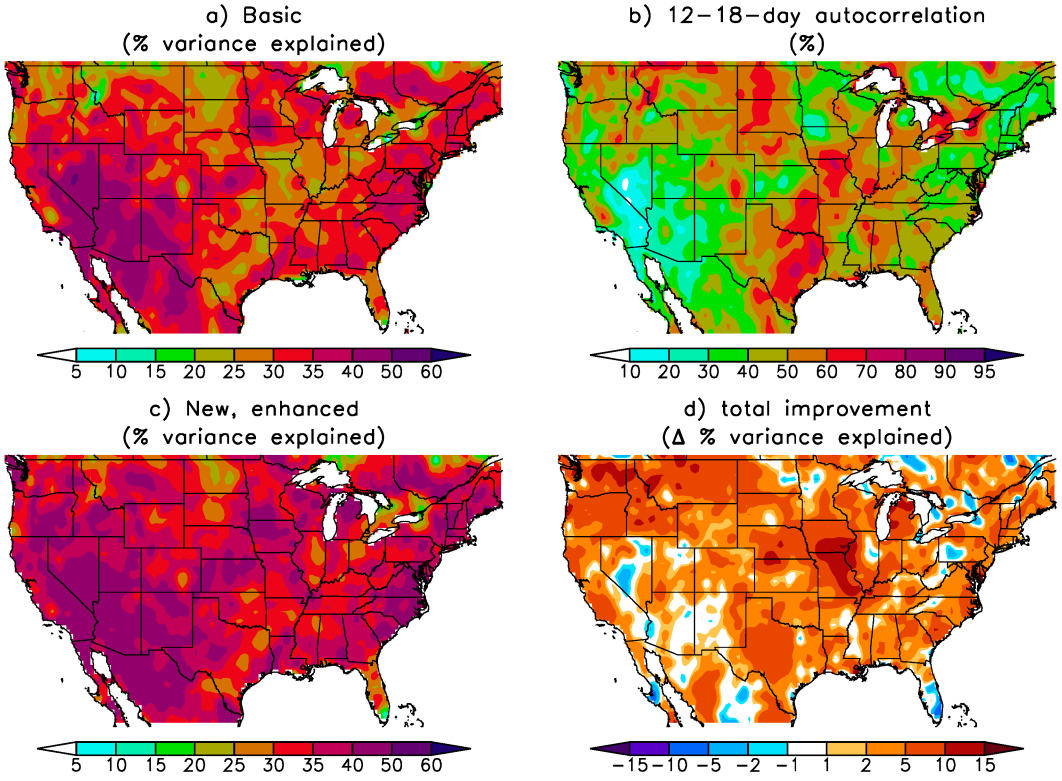


FIG. 6. As in Fig. 1, but for the change in the 7-day composite ESI for days 12–18 in the future.

In fact, in the enhanced forecast, this region no longer stands out as a region of lower skill (Fig. 6c). A similar situation holds in parts of Montana. On the other hand, significant parts of the southwest United States show relatively high skill in the basic forecast but no or very little improvement in the enhanced forecast.

In Fig. 7, we show the role of ISGP, additional predictors, and machine learning separately on ESI forecast skill. Of the four panels, ISGP leads to the most spatially uniform improvement in forecast skill (Fig. 7a). Additional predictors, on the other hand, lead to dramatic improvement in the central and northwest United States and very little improvement in the dry southwest (Fig. 7c). Lorenz et al. (2021) argue that improvements in ESI forecasts with additional predictors come from the fact that ESI is composed of two distinct components with dramatically different time scales: 1) a quickly evolving bare soil and canopy water component and 2) a slowly evolving transpiration component. By adding predictors, the statistical methodology can quantify the relative proportions of these two components at the initial state and can adjust the expected future changes accordingly. Soil moisture is particularly beneficial because it is a good proxy for the transpiration component of ESI (Lorenz et al. 2021). For the dry southwest, ESI is likely dominated by short bursts of bare soil and canopy water evaporation immediately after rainfall and the transpiration component is small. In this case, the ESI is dominated by the quickly evolving component of evaporation and the benefits of additional predictors are not significant. Note that the small ESI autocorrelation in the southwest (Fig. 6b) is consistent with this interpretation. As mentioned previously, GBM does not lead to consistent improvement in skill either with or without additional predictors (Figs. 7b,d).

A summary of the ESI skill averaged over the central United States is shown in Fig. 8. Like the 0–100-cm soil moisture, the improvements are almost exclusively coming from the initial/past predictors. For ESI, however, machine learning as currently implemented does not improve the skill of the forecasts. ESI also has a significantly larger fraction of the skill coming from the initial condition rather than the ECMWF dynamical model compared to soil moisture. In addition, when actual future observations of precipitation and dewpoint depression are used to "forecast" ESI, we find that skill does not improve to the same degree as soil moisture. For example, approximately 50% of the variance is explained compared to over 90% for soil moisture. The excellent skill for soil moisture in this "perfect" forecast setting is likely due to the direct role of precipitation on the soil moisture budget. For ESI, on the other hand, complex plant behavior is involved. In addition, we find that machine learning performs better than linear methods in a perfect forecast setting where future meteorology is known a priori. This suggests that there might be room for further improvement using machine learning in a realistic forecast setting.

The fact that the improvements are almost exclusively coming from the initial/past predictors has important implications

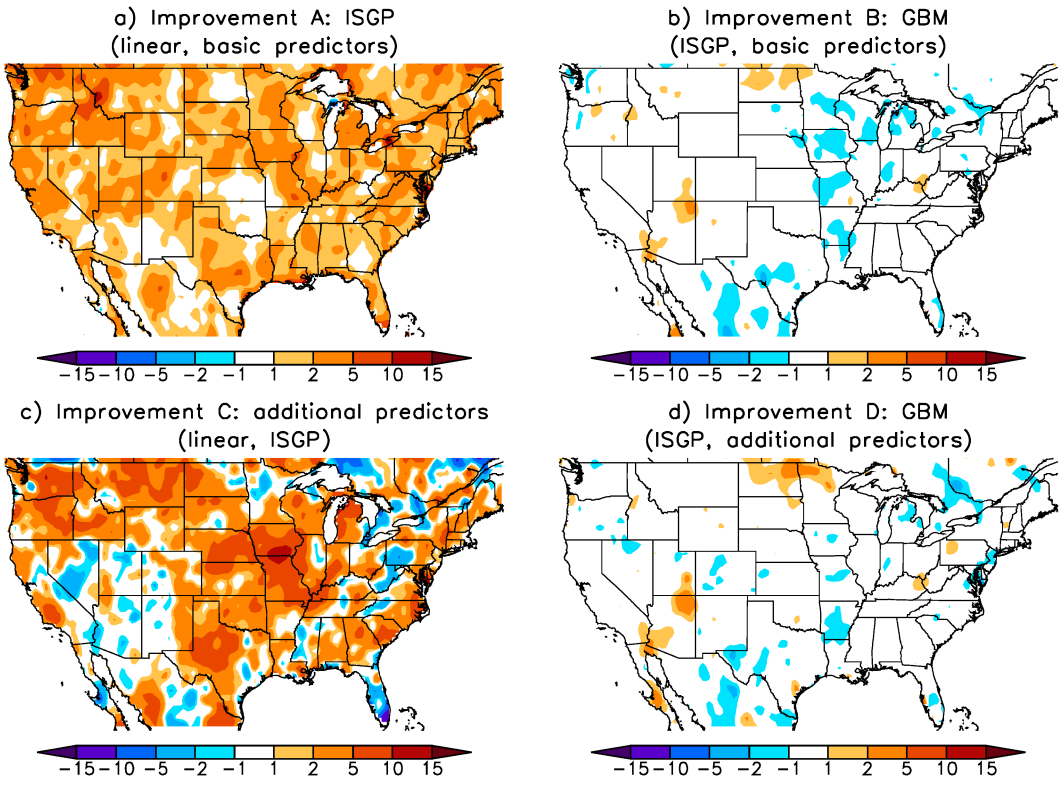


FIG. 7. As in Fig. 3, but for the change in the 7-day composite ESI for days 12–18 in the future.

in forecasting intensification of soil moisture and ESI anomalies. As discussed in Lorenz et al. (2021), the initial/past predictors tend to control the rate that anomalies relax back to climatology. Intensification of anomalies primarily comes from the dynamical forecast model. This also implies that flash droughts that are missed by dynamical models, such as the 2017 Northern Plains Drought, will also be missed by the current methodology.

### c. *Effect of improvements on error distribution*

As the model skill improves, one typically expects the error variance to decrease symmetrically. In other words, both negative and positive errors are reduced in absolute value at the same time. ESI exhibits this expected behavior (not shown). For 0–100-cm soil moisture, on the other hand, only the positive error is reduced, and the negative error hardly changes at all. For example, Fig. 9 shows the PDF of the predicted soil moisture tendency minus the actual soil moisture tendency (i.e., PDF of the error) for the 15-day forecast for both the basic case (case 0, red) and the GBM with additional predictors case (case 4, blue). For robustness, the individual PDFs at each grid point are averaged over the central United States. Only the positive tail changes between cases. This result is not dependent on machine learning because the linear model with additional predictors also has improvement restricted to the positive tail of the distribution only (not shown). Note also that the error distribution is strongly skewed to the left. This skewness is from the actual soil moisture tendency, which is skewed to the right due to the skewness in precipitation. Because the error involves the negative of the actual tendency, the distribution in Fig. 9 is skewed to the left.

Looking in more detail, we show the two-dimensional PDF averaged over the central United States in Fig. 10. The additional dimension in Fig. 10 is the initial soil moisture anomaly ($x$ axis). Evidently, the decreases in the positive tail of the PDF shown in Fig. 9 are due to the reductions in the top-left quadrant in Fig. 10. The positive tail improves when the soil moisture is already dry, so, in other words, the number of false drought recoveries decreases. There are also additional interesting structures in Fig. 10 that are not evident in the one-dimensional PDF. For example, when the current soil moisture anomaly is close to zero, the sense of the improvement changes: The new model preferentially improves errors in future drying. In some sense, the prediction at zero initial soil moisture anomaly is degraded in the new model because the likelihood of zero error decreases slightly (note the change at the origin in Fig. 10 is negative). The 0–10-cm soil moisture exhibits similar behavior (not shown).

### d. *Interpretation of machine learning improvements*

In this subsection, we seek to explain how GBM improves forecasts of 0–100-cm soil moisture change. In general, GBM is not an interpretable method; however, there are strategies for understanding the functional form of black-box machine learning methods. Here, we use a method that distills the

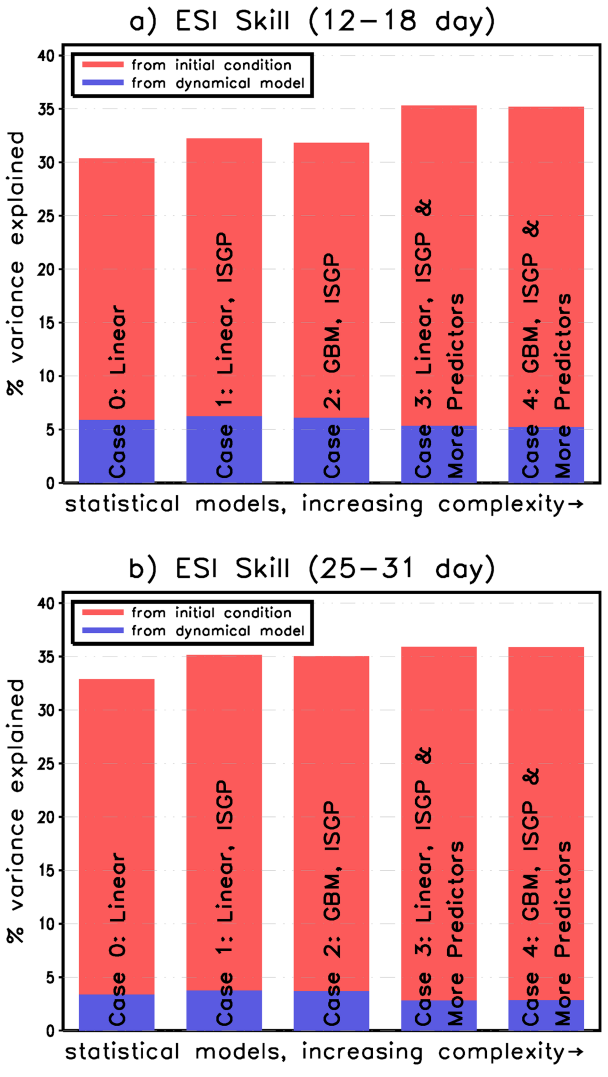


FIG. 8. As in Fig. 4, but for the change in the 7-day composite ESI for (a) days 12–18 and (b) 25–31 days in the future.

full GBM function of $n$ variables to a single variable by summarizing the main effects of an individual predictor. Accumulated local effect (ALE) plots (Apley and Zhu 2020) are the current state-of-the-art method for understanding the effect of individual predictors; however, it requires saving the machine learning model at each grid point and fold of the cross validation. Instead, we use the marginal plot (Apley and Zhu 2020), which simply requires the predictions and predictors at each grid point and time step.[1] The marginal plot is also beneficial because the actual soil moisture changes can be plotted and compared, unlike the ALE plot. Let $f(x_1, x_2, x_3, \ldots, x_n)$ be the function fitted by GBM. Suppose we want to understand the individual effect of the

[1] A marginal plot is also used to describe a scatterplot with histograms on the margins of the $x$ and $y$ axes. This is not the marginal plot used here.

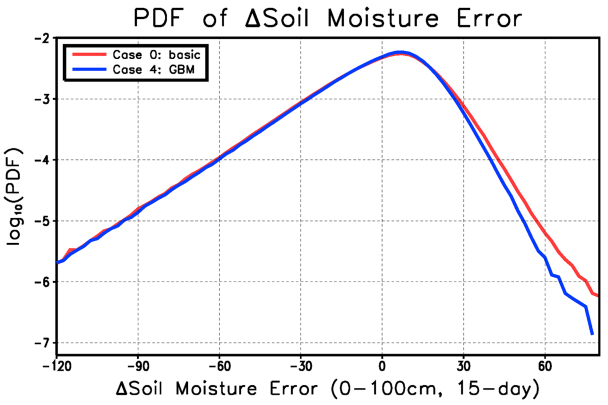


FIG. 9. PDF of the error in the 15-day 0–100-cm soil moisture (kg m$^{-2}$) forecast averaged over the central United States. Both the basic (red, case 0) and enhanced (blue, case 4) forecasts. To better see the tails, the logarithm of the PDF is plotted.

variable $x_1$. First, we choose an appropriate number of bins that span the range of $x_1$. For each $x_1$ bin, we calculate the average value of $f$. The marginal plot is the average $f$ in each bin as a function of the value of $x_1$ in each bin. The marginal plot for the change in the 0–100-cm soil moisture is shown in Fig. 11. To increase sample size, we average the results from 95° to 100°W and 40°–45°N, which is the region where GBM is most beneficial (see Fig. 3d).

First, we explore the dependence of soil moisture change on the initial/past 0–100-cm soil moisture (=$x_1$) for our various forecasts (Fig. 11a). Also shown is the actual change in soil moisture averaged in each $x_1$ bin (black, dashed line). When the 0–100-cm soil moisture anomaly is positive, the change in soil moisture tends to be negative and vice versa. Note that the actual soil moisture change decreases rapidly with the soil moisture anomaly when the soil moisture anomaly is positive. For negative soil moisture anomalies, on the other hand, the soil moisture change saturates at a relatively small, finite value. A possible explanation of this behavior is

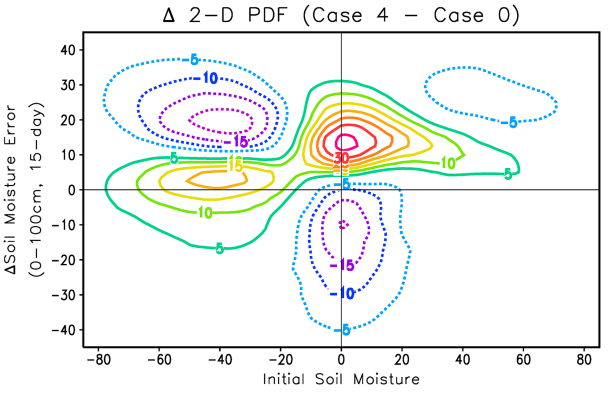


FIG. 10. As in Fig. 9, but for the two-dimensional PDF as a function of initial 0–100-cm soil moisture ($x$ axis; kg m$^{-2}$) and the error in the 15-day 0–100-cm soil moisture forecast ($y$ axis; kg m$^{-2}$). The PDF is multiplied by $1 \times 10^6$ prior to plotting.

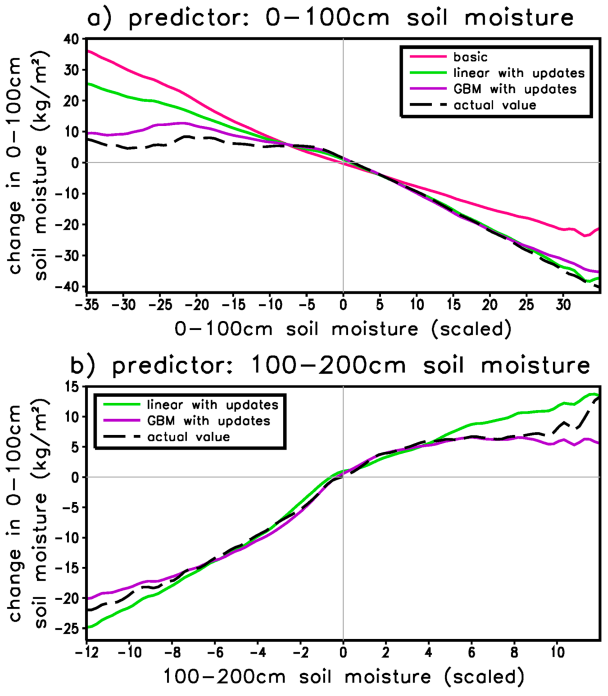


FIG. 11. (a) The dependence of the prediction (change in 0–100-cm soil moisture at 15 days) as a function of the initial/past 0–100-cm soil moisture as estimated from the marginal plot (see text). Recall the initial and post-soil moisture anomalies are combined using the regression coefficients from the linear method as part of dimensionality reduction (see section 3c); therefore, the *x*-axis units are scaled instead of raw anomalies. The *y* axis is the predicted change in soil moisture at 15 days (kg $m^{-2}$). The basic linear method (magenta), linear method with ISGP and additional predictors (green), GBM with ISGP and additional predictors (purple), and the actual soil moisture change (black, dashed) are shown. To increase sample size, the results are averaged from 95° to 100°W and 40° to 45°N. There are 101 bins over the *x* axis, and the results are smoothed with two applications of a 1-2-1 filter. (b) As in (a), but for initial/past 100–200-cm soil moisture as the individual predictor.

as follows: Positive soil moisture anomalies tend to lead to increased evaporation, runoff, and/or percolation to deeper soil layers, which then act as a negative feedback that decreases the soil moisture anomaly. For negative soil moisture anomalies, on the other hand, decreases in evaporation, runoff, and/or percolation to deeper soil layers are bounded by zero, and therefore, the "restoring force" for soil drying is weaker than soil moistening. The marginal plot for the basic linear method (magenta) cannot capture the change in slope. Instead, the "best fit" line is too steep for negative soil moisture anomalies and too shallow for positive soil moisture. When additional predictors are added to the linear method (green), the additional degrees of freedom somehow allow the linear method to capture some of the slope change. GBM further improves the representation of the true relationship (purple).

Another predictor that benefits from GBM is the 100–200-cm soil moisture (Fig. 11b). This variable appears to influence the change in 0–100-cm soil moisture via its effects on percolation: Moist deep soil limits the reduction in soil moisture above via decreased percolation and vice versa (black, dashed). In this case, the lower bound on percolation causes the function to saturate when 100–200-cm soil moisture is very moist. We do not show the basic method in Fig. 11b because 100–200-cm soil moisture is not a predictor for this method. The linear method with additional predictors (green) can capture a surprising amount of the actual relationship, but the added flexibility of GBM (purple) leads to additional improvement. The remaining predictors do not show significant improvements with GBM. This is consistent with our finding that the GBM improvements are leveraging information in the initial/past state rather than the dynamical forecasts. In future work, we will try to understand why GBM does not improve forecasts everywhere even though the above physics is active everywhere. How the additional predictors in the linear method can capture some of the improvements is another interesting issue to explore.

## 5. Conclusions

In this study, we improve the subseasonal forecasts of soil moisture and ESI developed by Lorenz et al. (2021). Like Lorenz et al. (2021), we combine the current/past land state and dynamical forecast model output to create gridded forecasts over the contiguous United States and the surrounding regions of Canada and Mexico. We explore three strategies for improving the forecasts: 1) nonlinear machine learning methods, 2) additional predictors, and 3) increasing sample size by ISGP in the training of the statistical models. For nonlinear machine learning, GBM performed the best of the six methods tried. In some sense, ISGP is the most important improvement because the other methods do not increase skill unless sample size is also increased. Nonlinear machine learning methods can improve soil moisture forecasts; however, ISGP makes a similar contribution to the enhanced skill and additional predictors make an even greater contribution. Therefore, averaged over the domain, the best linear method performs only slightly worse than GBM and there are certain regions where the linear method outperforms GBM. This is consistent with the machine learning–based subseasonal temperature forecasts of He et al. (2021). For ESI, nonlinear machine learning does not improve upon linear methods in a true forecast setting. Because the ESI is a remotely sensed observed quantity, unlike the Noah soil moisture, perhaps, a larger time period is needed for a robust signal to appear above the noise. This is consistent with the fact that reducing the noise by substituting actual precipitation and temperature observations for S2S forecasts enables nonlinear machine learning methods to beat the linear method.

As our statistical methods improve, the current/past land state is the source of almost all the increases in skill for both soil moisture and ESI. The variance explained by the dynamical forecast model output, on the other hand, is essentially unchanged from method to method. This suggests that understanding the processes involving the initial land/vegetation state is the most fruitful approach to increasing

hydrometeorological subseasonal skill, although improvements to dynamical forecasts are required to better capture the intensification of anomalies (Lorenz et al. 2021). In future work, we will explore whether alternate machine learning methods can better leverage the information in the dynamical forecasts.

The improvements to the soil moisture forecasts do not uniformly improve for all situations. Instead, our methods improve forecasts of rapid increases in soil moisture, while forecasts of rapid decreases are hardly impacted. The improvements preferentially occur when soil moisture is already dry, so the new methods primarily improve forecasts of drought recovery. For ESI, on the other hand, there is no pronounced asymmetry in forecast improvement. Analysis of the functional form of the GBM predictions for soil moisture shows that GBM is better able to capture the nonlinear relationship between 1) soil moisture change and the initial soil moisture and 2) soil moisture change and soil moisture at deeper levels of the soil profile. Both nonlinear relationships appear to be related to the fact that evaporation, runoff, and/or percolation are more strongly bounded in the negative direction than in the positive.

*Acknowledgments.* We thank four anonymous reviewers for their very helpful comments and suggestions. This work was supported by funds provided by the NSF PREEVENTS ICER-1854902.

*Data availability statement.* The NLDAS and ECMWF S2S data are publicly available. For the code and scripts, please contact the lead author.

## APPENDIX

### Comparison to Raw Soil Moisture Forecasts from ECMWF

In this appendix, the skill of the 0–100-cm soil moisture forecasts in this paper is compared with the raw 0–100-cm soil moisture from the ECMWF reforecasts. To help correct for simple scaling biases, the change in ECMWF soil moisture is a predictor in a simple univariate linear regression and the NLDAS change in soil moisture is the predictand. The skill of the ECMWF soil moisture forecasts is subtracted from the skill of the basic method (Fig. A1a) and new method with ISGP, additional predictors, and GBM (Fig. A1b). The basic method is more skillful than ECMWF over most of the domains, but there are a few areas where ECMWF is more skillful. These include northern Iowa and southern Minnesota, northwestern Mexico, and small isolated pockets in the western and central United States. The new method is a noticeable improvement, and now, skill is greater than the ECMWF everywhere except a small pocket in far southern Ontario. Unfortunately, the 0–10-cm soil moisture and a reference evapotranspiration (required to calculate ESI) are not available from the S2S archive so the skill for these variables cannot be evaluated in ECMWF.

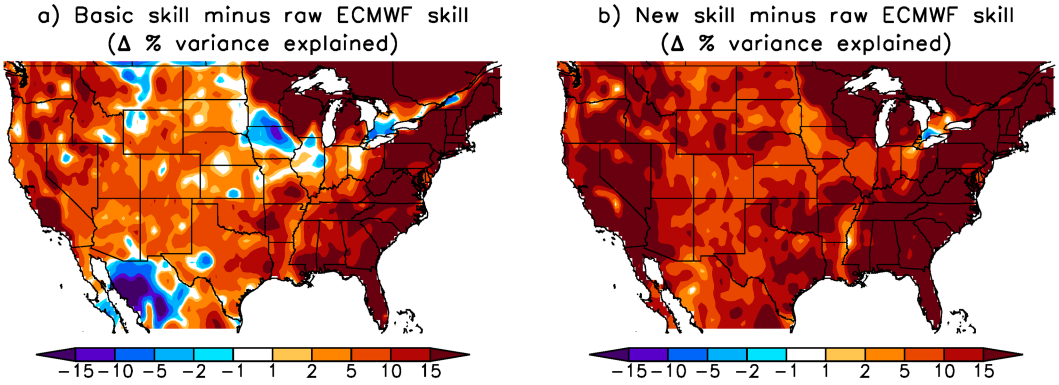


FIG. A1. (a) Percent variance explained for 0–100-cm soil moisture forecasts using the basic method minus percent variance explained by ECMWF soil moisture. Positive values mean the basic method is more skillful. (b) As in (a), but for the new method.

## 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, 333 pp., http://www.climasouth.eu/sites/default/files/FAO%2056.pdf.

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.

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

——, ——, ——, ——, and ——, 2007b: A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. *J. Geophys. Res.*, **112**, D11112, https://doi.org/10.1029/2006JD007507.

——, C. Hain, B. Wardlow, A. Pimstein, 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**, 2025–2044, https://doi.org/10.1175/2010JCLI3812.1.

——, ——, J. Otkin, X. Zhan, K. Mo, M. Svoboda, B. Wardlow, and A. Pimstein, 2013: An intercomparison of drought indicators based on thermal remote sensing and NLDAS-2 simulations with U.S. drought monitor classifications. *J. Hydrometeor.*, **14**, 1035–1056, https://doi.org/10.1175/JHM-D-12-0140.1.

——, and Coauthors, 2016a: The evaporative stress index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sens. Environ.*, **174**, 82–99, https://doi.org/10.1016/j.rse.2015.11.034.

——, and Coauthors, 2016b: Relationships between the evaporative stress index and winter wheat and spring barley yield anomalies in the Czech Republic. *Climate Res.*, **70**, 215–230, https://doi.org/10.3354/cr01411.

Apley, D. W., and J. Zhu, 2020: Visualizing the effects of predictor variables in black box supervised learning models. *J. Roy. Stat. Soc.*, **82B**, 1059–1086, https://doi.org/10.1111/rssb.12377.

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.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Brust, C., J. S. Kimball, M. P. Maneta, K. Jencso, and R. H. Reichle, 2021: DroughtCast: A machine learning forecast of the United States drought monitor. *Front. Big Data*, **4**, 773478, https://doi.org/10.3389/fdata.2021.773478.

Cao, J., H. Wang, J. Li, Q. Tian, and D. Niyogi, 2022: Improving the forecasting of winter wheat yields in northern China with machine learning–dynamical hybrid subseasonal-to-seasonal ensemble prediction. *Remote Sens.*, **14**, 1707, https://doi.org/10.3390/rs14071707.

Chen, J., and Coauthors, 2015: Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.*, **103**, 7–27, https://doi.org/10.1016/j.isprsjprs.2014.09.002.

Chen, T., and C. Guestrin, 2016: XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (SIGKDD), San Francisco, CA, Association for Computing Machinery, 785–794, https://doi.org/10.1145/2939672.2939785.

Christian, J. I., J. B. Basara, J. A. Otkin, E. D. Hunt, R. A. Wakefield, P. X. Flanagan, and X. Xiao, 2019: 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.

Cortes, C., and V. Vapnik, 1995: Support-vector networks. *Mach. Learn.*, **20**, 273–297, https://doi.org/10.1007/BF00994018.

DeAngelis, A. M., H. Wang, R. D. Koster, S. D. Schubert, Y. Chang, and J. Marshak, 2020: Prediction skill of the 2012 U.S. Great Plains flash drought in subseasonal experiment (SubX) models. *J. Climate*, **33**, 6229–6253, https://doi.org/10.1175/JCLI-D-19-0863.1.

Deo, R. C., and M. Şahin, 2015: Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia. *Atmos. Res.*, **153**, 512–525, https://doi.org/10.1016/j.atmosres.2014.10.016.

Dikshit, A., and B. Pradhan, 2021: Interpretable and explainable AI (XAI) model for spatial drought prediction. *Sci. Total Environ.*, **801**, 149797, https://doi.org/10.1016/j.scitotenv.2021.149797.

——, ——, and A. M. Alamri, 2020: Temporal hydrological drought index forecasting for New South Wales, Australia using machine learning approaches. *Atmosphere*, **11**, 585, https://doi.org/10.3390/atmos11060585.

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.

Fernández-Delgado, M., M. S. Sirsat, E. Cernadas, S. Alawadi, S. Barro, and M. Febrero-Bande, 2019: An extensive experimental survey of regression methods. *Neural Networks*, **111**, 11–34, https://doi.org/10.1016/j.neunet.2018.12.010.

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. *Computer Analysis of Images and Patterns*, A. Gagalowicz and W. Phillips, Eds., Springer, 407–414.

Friedman, J. H., 1991: Multivariate adaptive regression splines. *Ann. Stat.*, **19** (1), 1–67, https://doi.org/10.1214/aos/1176347963.

——, 2001: Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, **29**, 1189–1232, https://doi.org/10.1214/aos/1013203451.

Hastie, T., and R. Tibshirani, 1986: Generalized additive models. *Stat. Sci.*, **1**, 297–310, https://doi.org/10.1214/ss/1177013604.

He, S., X. Li, T. DelSole, P. Ravikumar, and A. Banerjee, 2021: Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances. *Proc. Conf. AAAI Artif. Intell.*, **35**, 169–177, https://doi.org/10.1609/aaai.v35i1.16090.

——, ——, L. Trenary, B. A. Cash, T. DelSole, and A. Banerjee, 2022: Learning and dynamical models for sub-seasonal climate forecasting: Comparison and collaboration. *Proc. Conf.*

*AAAI Artif. Intell.*, **36**, 4495–4503, https://doi.org/10.1609/aaai.v36i4.20372.

Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. *Neural Comput.*, **9**, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Hoell, A., T. W. Ford, M. Woloszyn, J. A. Otkin, and J. Eischeid, 2021: Characteristics and predictability of midwestern United States drought. *J. Hydrometeor.*, **22**, 3087–3105, https://doi.org/10.1175/JHM-D-21-0052.1.

Hoerling, M., J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager, 2014: Causes and predictability of the 2012 Great Plains drought. *Bull. Amer. Meteor. Soc.*, **95**, 269–282, https://doi.org/10.1175/BAMS-D-13-00055.1.

Hunt, E. D., 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. 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. *Agric. For. Meteor.*, **191**, 1–11, https://doi.org/10.1016/j.agrformet.2014.02.001.

Hwang, J., P. Orenstein, J. Cohen, K. Pfeiffer, and L. Mackey, 2019: Improving subseasonal forecasting in the western U.S. with machine learning. *Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Anchorage, AK, Association for Computing Machinery, 2325–2335, https://doi.org/10.1145/3292500.3330674.

Jalali, A., P. Ravikumar, and S. Sanghavi, 2013: A dirty model for multiple sparse regression. *IEEE Trans. Inf. Theory*, **59**, 7947–7968, https://doi.org/10.1109/TIT.2013.2280272.

Januschowski, T., Y. Wang, K. Torkkola, T. Erkkilä, H. Hasson, and J. Gasthaus, 2022: Forecasting with trees. *Int. J. Forecasting*, **38**, 1473–1481, https://doi.org/10.1016/j.ijforecast.2021.10.004.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, 2017: Lightgbm: A highly efficient gradient boosting decision tree. *NIPS'17: Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach CA, Curran Associates Inc., 3149–3157, https://dl.acm.org/doi/10.5555/3294996.3295074.

Khan, N., D. A. Sachindra, S. Shahid, K. Ahmed, M. S. Shiru, and N. Nawaz, 2020: Prediction of droughts over Pakistan using machine learning algorithms. *Adv. Water Resour.*, **139**, 103562, https://doi.org/10.1016/j.advwatres.2020.103562.

Lawson, C. L., and R. J. Hanson, 1995: *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, 337 pp.

Lorenz, D. J., J. A. Otkin, M. Svoboda, C. R. Hain, M. C. Anderson, and Y. Zhong, 2017: Predicting the U.S. drought monitor 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.

——, ——, B. Zaitchik, C. Hain, and M. C. Anderson, 2021: Predicting rapid changes in evaporative stress index (ESI) and soil moisture anomalies over the continental United States. *J. Hydrometeor.*, **22**, 3017–3036, https://doi.org/10.1175/JHM-D-20-0289.1.

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-institution 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.

Otkin, J. A., M. C. Anderson, C. Hain, I. E. Mladenova, J. B. 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.

——, M. Shafer, M. Svoboda, B. Wardlow, M. C. Anderson, C. Hain, and J. Basara, 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.*, **218–219**, 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, 2018a: 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.

——, Y. Zhong, D. Lorenz, M. C. Anderson, and C. Hain, 2018b: Exploring seasonal and regional relationships between the evaporative stress index and surface weather and soil moisture anomalies across the United States. *Hydrol. Earth Syst. Sci.*, **22**, 5373–5386, https://doi.org/10.5194/hess-22-5373-2018.

——, and Coauthors, 2022: Getting ahead of flash drought: From early warning to early action. *Bull. Amer. Meteor. Soc.*, **103**, E2188–E2202, https://doi.org/10.1175/BAMS-D-21-0288.1.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

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, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

Servén, D., and C. Brummitt, 2018: pyGAM: Generalized additive models in Python. Zenodo, accessed 20 July 2021, https://doi.org/10.5281/zenodo.1208723.

Slater, L., and Coauthors, 2023: Hybrid forecasting: Blending climate predictions with AI models. *Hydrol. Earth Syst. Sci.*, **27**, 1865–1889, https://doi.org/10.5194/hess-27-1865-2023.

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.

Tibshirani, R., 1996: Regression shrinkage and selection via the LASSO. *J. Roy. Stat. Soc.*, **58A**, 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

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.

Wang, C., Z. Jia, Z. Yin, F. Liu, G. Lu, and J. Zheng, 2021: Improving the accuracy of subseasonal forecasting of China precipitation with a machine learning approach. *Front. Earth Sci.*, **9**, 659310, https://doi.org/10.3389/feart.2021.659310.

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.

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76**, 1077–1091, https://doi.org/10.1175/JAS-D-18-0269.1.

Zhu, Q., and Y. Wang, 2021: The diagnosis about spatiotemporal characteristics and driving factors of flash drought and its prediction over typical humid and semiarid basins in China. *J. Hydrometeor.*, **22**, 2783–2798, https://doi.org/10.1175/JHM-D-21-0062.1.