Title: The use of soil moisture and Standardized Evaporative Stress Ratio (SESR) anomalies for increased lead time of the development flash drought and heat waves

Authors/Affiliation: Dr. Eric Hunt (Atmospheric and Environmental Research, Inc.), Dr. Amy McGovern (University of Oklahoma), Dr. Jordan Christian (University of Oklahoma), Dr. Lauren E.L. Lowman (Wake Forest University), Mr. Thomas Connor (Atmospheric and Environmental Research, Inc.)

Focal Area: We will be focused on pathway number 2 via predictive modeling of drought through the use of AI-techniques that will involve a hierarchy of models.

Science challenge: Drought and rapid intensification of drought (i.e., flash drought) can have significant impacts on the natural environment and the economy, particularly when coupled with a period(s) of extreme heat (Otkin et al., 2018). The development and/or intensification of drought falls squarely in the sub-seasonal to seasonal (S2S) time frame, which dictates that drought development, intensification, and propagation could be an output of a subseasonal or seasonal forecast. Current S2S forecasts have poor skill for forecasting the development of extreme drought and heat waves.

Rationale: Seasonal projections or forecasts of drought at the S2S timescale historically have had poor performance. Several factors can contribute to the lack of skill. Besides erroneous forecasts or projections of precipitation anomalies, one of the known issues with a poor seasonal forecast of drought is inaccuracy in the initialization of soil moisture. For example, DeAngelis et al. (2020) showed that accurate initialization of soil moisture did provide improvements in temperature and precipitation biases in a hindcast of the summer of 2012 drought over the central United States. One main issue with forecast skill of drought could simply be the rapid way in which many droughts develop in the warm season, which in recent years has been referred to as flash droughts. Flash droughts are defined in Otkin et al. (2018) as a rapid onset and intensification of drought and are characterized by abnormally high temperatures, increased wind speeds, greater incoming solar radiation, and rapid depletion of soil moisture that leads to a marked decline in vegetation health.

We contend that accurate S2S projections of drought and flash drought are unlikely to be achieved at this time with dynamical models, however, there is potential for a data driven Machine Learning (ML) model to find skill (Hwang, 2018 arXiv:1809.07394). We seek to explore the relationship between the many variables that contribute to the development and persistence of drought. Variables such as soil moisture index, geopotential height anomalies, sea surface temperatures, teleconnection indices (Madden Julian Oscillation, for example), and many more are involved in cultivating an environment that is more or less conducive to the development of drought over a particular region at the S2S timescales. We seek to build ML models that explore the complex and highly nonlinear variable space and elicit signal sufficient enough to produce drought projections with accuracies exceeding current capabilities.

We will leverage our prior knowledge around variables that we know to be significant in the early detection of emerging drought. One such variable is the Standardized Evaporative Stress Ratio (SESR; Christian et al., 2019), which is simply the standardized ratio between evapotranspiration (ET) and potential evapotranspiration (ET_P), where potential ET is determined using the Penman-

DOI: https://doi.org/10.2172/1769783
DOE BER, AI4ESP 2021 #AI4ESP1064 (https://www.ai4esp.org)

Monteith equation. The use of SESR is advantageous for both the identification of existing drought and the onset of flash drought ((Edris et al., 2020) as it incorporates, either directly or indirectly, numerous variables previously identified as critical in flash drought development including air temperature, wind speed, vapor pressure deficit, latent and sensible heat flux, soil moisture, and precipitation (Hunt et al., 2014; Otkin et al. 2014; Otkin et al. 2016; Hobbins et al. 2016; Ford et al. 2015, Otkin et al., 2019; Christian et al., 2019a,b; Christian et al., 2020; Hunt et al., in review). However, it is also possible that SESR anomalies may be less affected by erroneous forecasts of precipitation, which implies that it may be able (in some cases) to project drought without having a correct forecast of precipitation.

The primary benefit of the proposed approach is by considering the anomalies of SESR in addition to anomalies of precipitation, soil moisture, and teleconnections we will get a full picture of the "supply and demand balance" between the soil and the atmosphere and of the drivers that affect said balance. Another benefit of our proposed approach will allow us to determine if highly negative anomalies of SESR are reasonably correlated with extreme heat. Machine learning will help to determine the overall strength of the spatial and temporal relationships between variables, which in turn will lead to intelligence to help improve forecasts at the S2S timescale. A secondary benefit is the meteorological variables can be derived from a wide variety of gridded datasets and could also be calculated at different DOE Atmospheric Radiation Measurement sites. Significant barriers are not expected in terms of developing training datasets of the SMI and SESR to be used in neural networks for forecasting of flash drought.

Narrative: We will first generate a soil moisture index (SMI) and calculate SESR anomalies at a pentad timestep (i.e., 5 days) over a 40-year period of record over CONUS and other regions of the world. The SMI is based on Hunt et al. (2009) and is calculated as follows in Equation 1:

$$SMI = \frac{\theta - \theta_{WP}}{\theta_{FC} - \theta_{WP}} \qquad (1)$$

where θ is the observed water content, θ_{WP} is the minimum water content, and θ_{FC} is the maximum water content for a location. The calculation of SESR is given in Equation 2:

$$SESR = \frac{ESR_{ijp} - \overline{ESR_{ijp}}}{\sigma_{ESR_{ijp}}} \qquad (2)$$

where SESR is the z-score of the Evaporative Stress Ratio (ESR; calculated as ET/ET_p), $\overline{ESR}$ is the mean ESR for a specific pentad (p) using all years available in the gridded dataset and is done separately for each pentad, and σ_{ESR} is the standard deviation of ESR for a specific pentad and grid point (i,j) using all years available in the gridded dataset.

We will use the following reanalyses, forcing datasets, and options for generating the anomalies of the SMI and SESR: ERA-5 (Hersbach et al. 2020), MERRA, Version 2 (MERRA-2; Gelaro et al. 2017), and North American Land Data Assimilation System, version 2 (NLDAS-2; Xia et al. 2012a,b) forcing. In addition to the gridded data, we will calculate the SMI and SESR from the ARM sites where soil moisture data and associated data to calculate SESR are available. These stations could be clustered by ecoregion to determine if there is a difference in skill of predicting drought and flash drought in certain regions versus others.

After the pentad anomalies have been generated for both SMI and SESR, we will generate precipitation anomalies via a moving 1-month and 3-month Standardized Precipitation Index (SPI; McKee et al., 1993) that will coincide with the pentad measurements of SESR and the SMI. We will also use monthly and seasonal middle and upper tropospheric anomalies (e.g., an anomalous ridge at 500 mb) and teleconnection data (e.g., Madden Julian Oscillation) hat can be compared to the other variables. Finally, we will assimilate soil moisture and leaf area index (LAI) data in order to constrain uncertainty in SMI and SESR forecasts (Mocko et al. 2021). Using an Ensemble Kalman Filter (EnKF) based approach to assimilate soil moisture and LAI will allow for direct quantification of the uncertainty in the drought anomalies (Reichle et al. 2008; De Lannoy and Reichle, 2016).

We will designate drought regions across the dataset via applying a threshold value to the SESR and categorize drought based on severity (e.g., drought class 0 through 5). This is the dataset that will comprise the target for our ML models (i.e., the dependent variable). The independent variables will include indices of soil moisture, precipitation, and vegetation, and the data will be assembled so that the target (drought category) will be predicted by the model using independent variables that are a month, or more, older than it. We will deploy a number of ML approaches, starting with a simple classification decision tree, moving through more complex approaches involving an ensemble of gradient boosted decision trees (the xgboost algorithm, for example) and into highly complex Convolutional Neural Networks (CNN), Long Short Term Memory neural networks (LSTM) and hybrid version of those networks (CNN/LSTM). The advantage of training the simple decision tree first is that it establishes a baseline performance with results that are interpretable and easy to understand. The more complex model architectures would then be evaluated against the simple benchmark to ensure that the design is providing actual lift.

References:

Christian, J. I. et al. A methodology for flash drought identification: application of flash drought frequency across the United States. *Journal of Hydrometeorology* **20**, 833–846 (2019).

Christian, J. I., Basara, J. B., Otkin, J. A. & Hunt, E. D. Regional characteristics of flash droughts across the United States. *Environmental Research Communication* **1**, 125004 (2019).

Christian, J. I., Basara, J. B., Hunt, E. D., Otkin, J. A. & Xiao, X. Flash drought development and cascading impacts associated with the 2010 Russian heatwave. *Environ. Res. Lett.* **15,** 094078 (2020).

De Lannoy, G. and Reichle, R., 2016. Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrology and Earth System Sciences* **20**, 4895-4911 (2016).

Edris, S., Basara, J., Christian, J. , Wakefield, R., Illston, B., Sarkey, S. Evaluation of flash drought criteria components. *Advances in Monitoring, Analysis, and Prediction of Flash Drought and Related Heatwaves*, 101st annual American Meteorological Society. (2021)

Ford, T. W., McRoberts, D. B., Quiring, S. M. & Hall, R. E. On the utility of in situ soil moisture observations for flash drought early warning in Oklahoma, USA. *Geophysical Research Letters* **42,** 9790–9798 (2015).

Gelaro, R. *et al.* The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate* **30,** 5419–5454 (2017).

Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146,** 1999–2049 (2020).

Hobbins, M., Wood, A., Streubel, D., & Werner, K. What drives the variability of evaporative demand across the conterminous United States? *Journal of Hydrometeorology* **13,** 1195–1214 (2012).

Hunt, E. D. *et al.* Monitoring the effects of rapid onset of drought on non-irrigated maize with agronomic data and climate-based drought indices. *Agricultural and Forest Meteorology* **191,** 1–11 (2014).

Mocko, D. M., Kumar, S. V., Peters-Lidard, C. D., & Wang, S . Assimilation of vegetation conditions improves the representation of drought over agricultural areas, *Journal of Hydrometeorology* (2021)

Otkin, J. A., Anderson, M. C., Hain, C. & Svoboda, M. Examining the Relationship between Drought Development and Rapid Changes in the Evaporative Stress Index. *Journal of Hydrometeorology* **15,** 938–956 (2014).

Otkin, J. A. *et al.* Assessing the evolution of soil moisture and vegetation conditions during the 2012 United States flash drought. *Agricultural and Forest Meteorology* **218-219,** 230–242 (2016).

Otkin, J. A. *et al.* Flash droughts: a review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bulletin of the American Meteorological Society* **99,** 911–919 (2018).

Otkin, J. A. *et al.* Assessing the evolution of soil moisture and vegetation conditions during a flash drought–flash recovery sequence over the South-Central United States. *Journal of Hydrometeorology* **20,** 549–562 (2019).

Reichle, R.H., Crow, W.T., Koster, R.D., Sharif, H.O. and Mahanama, S.P.P., Contribution of soil moisture retrievals to land data assimilation products. *Geophysical Research Letters* **35**, (2008).

Xia, Y., M.B. Ek, D. Mocko, C.D. Peters-Lidard, J. Sheffield, J. Dong, and E.F. Wood, 2014: Uncertainties, Correlations, and Optimal Blends of Drought Indices from the NLDAS Multiple Land Surface Model Ensemble. *J. Hydrometeor.*, **15**, 1636–1650.