Drought Prediction Across the Continental United States through Statistical and Machine Learning Methods

Thee Ngamsangrat [1] **Yujie Cai** [1] **Jim Zhang** [1] **Michael Fein** [1]

Abstract

Drought has become one of the leading causes of humanitarian disasters. Drought leads to drinking water shortages, wild fires, loss in crop and livestock, and more. Post-drought aid is time consuming and due to the inability to perform long-term drought prediction, by the time aid has arrived, irreparable damage has already been done. In this paper we attempt to extend drought prediction accuracy past currently accepted capabilities using statistical and machine learning approaches based on a number of feature variables. We hope to use this tool, at the very least, to be able to more accurately allocate emergency response resources before the onset of more severe drought. If possible, we also hope to describe certain actions, taken by humans, that may have a relationship with drought.

1. Introduction

1.1. Understanding Drought

Climate change has been widely accepted as the leading cause of a measured increase in frequency of extreme weather events (Sharafati Shahid, 2020). Droughts are one of the most costly and environmentally damaging climatic events. Due to nature's dependency on hydraulic cycles, droughts can lead to the long-term destruction of large-scale ecosystems through positive feedback loops. Understanding and better predicting when and where drought will occur is of utmost importance to aiding those effected by drought and in understanding how to potentially mitigate oncoming droughts.

From a scientific perspective, droughts are difficult to measure with unclear beginnings and ends (Vincente-Serrano et al., 2010). Defining drought, while not the main topic of this work, is an important aspect of predicting its onset as well as its effects. For our purposes, we will use the U.S. Drought Monitor, which seeks to synthesize expert vetted information to classify various types of drought across the United States (National Drought Mitigation Center, 2022). The USDM makes use of the Palmer Drought Severity Index, the Standardized Precipitation Index (SPI), and other climatic inputs as well as indicators of soil moisture. As an example, the figure below shows the most recent drought classification across the Continental United States (CONUS).

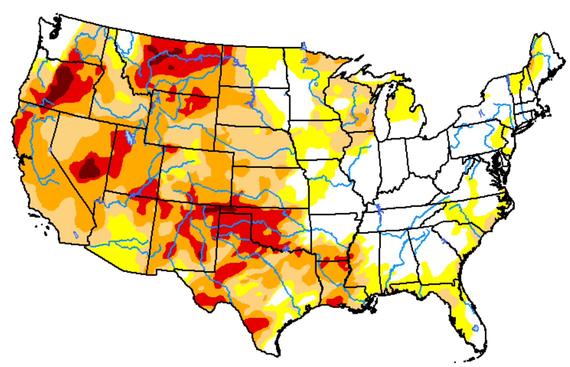

Figure 1. USDM Drought Severity classification across the Continental United Sates as of February 22nd, 2022.

Due to its relative acceptance as an accurate categorization of drought severity and its aggregation of different drought indices, we use the USDM as our response variable throughout this paper.

[1]Institute of Applied Computer Science, Harvard, Cambridge, United States of America, equal contribution. Correspondence to: 297r Capstone Project <Quantum Black>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

1.2. Available Data

For our purposes of drought prediction, we will be using data from a number of other sources in an effort to improve longer-term predictability of drought onset. These data sources include:

- Sentinel L2 Satellite Images
- Soil Moisture Active Passive (SMAP) Satellite Measurements
- SET Evapotranspiration Data

Using this data we hope to align changes over time to landscapes, urban development, soil moisture, agricultural land use, etc. with USDM drought categorization to achieve higher accuracy in drought prediction and timing. For example, the figure below demonstrates how SMAP can change over time:

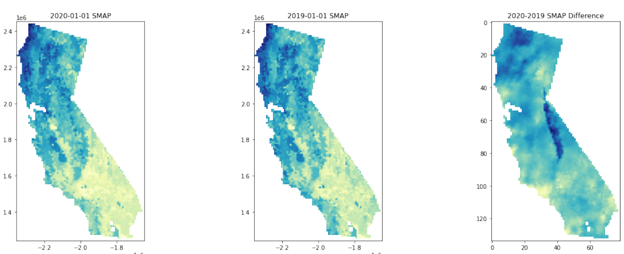

Figure 2. Difference in SMAP Measurements from 2019 to 2020 accross the state of California.

Using Sentinel Satellite images such as those pictured below gives us many options for creating feature-sets in our project. The data can be easily accessed and is relatively granular in terms of detail.

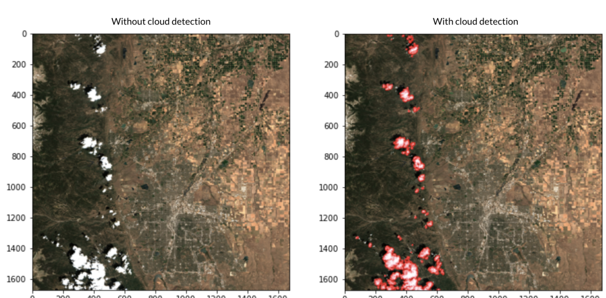

Figure 3. Sample of Satellite images with and without cloud masking.

The data set also includes meta-data which attempts to classify portions of the image as cloud cover. Potential options for better understanding drought through satellite images include masking for land use (urban vs. non-urban vs. agricultural) and using the cloud cover mask as a potential predictor. We can combine both datasets with historical evaporation and run-off data from SET.

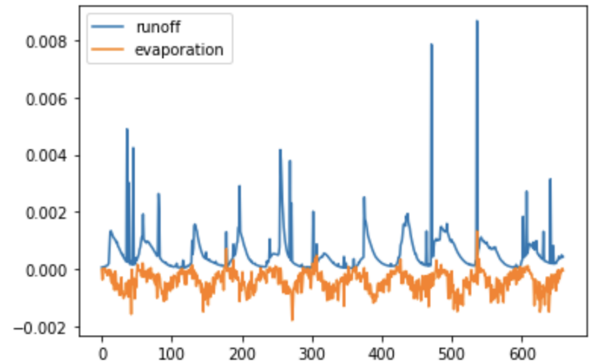

Figure 4. Sample of evaporation and runoff data for a specific location.

The main difficultly in using these various data sets is understanding how to most effectively merge them. For example, mapping satellite images to SMAP images raises issues along the border of maps as well as how to extract irregular polygons of the correct scale. In our efforts to predict drought, we plan to attempt both a statistical auto-regressive model with multiple X predictors as well as a machine learning black-box model.

2. Related Works

Previous work in the study of drought and arid climate forecasting has formed a major consensus – drought is the least understood natural disaster due to its convoluted causal factors and occurrence mechanisms (Sundararajan et al., 2021). This consensus declares that drought prediction based on numerical simulations and equation solving, similar to the existing weather forecasting, is impossible. However, with the flourishing of machine learning and deep neural networks, researchers have made some limited but encouraging progress in this direction that was once considered impossible.

2.1. Staggering complexity in droughts' causal mechanism

The factors that shape drought climates are extremely convoluted and usually involve the interaction of multiple factors (Salehi-Lisar Bakhshayeshan-Agdam 2016) . Most existing works analyze droughts in a case-by-case fashion (Marengo et al., 2021), usually limiting the scope of work on one drought case that happened within a time window narrower than a year.

Current literature suggests that causes of droughts are rooted deeply in the specific topographical and geographical conditions. Zhang et al. (2020) suggests that the reason for the stagnant development of drought forecasting is that the field has not agreed on a unified theoretical framework to explain the causal mechanism of droughts happening across the world, or even within the same continent.

2.2. Drastic variety in droughts' response to global warming

Previous works culminating in the survey work of Vicente-Serrano et al. (2020) recounted a snowballing problem that makes the drought forecasting even more unrealistic – droughts react very differently to the global warming context.

Some persistent and cyclical droughts have been paradoxically mitigated by the melting of the polar ice caps, leading to a number of sporadic misallocations of drought alleviation supplies (Scheff, 2018). Meanwhile, Bhaga et al. (2020) warned that global climate change has worsened the already severe droughts in sub-saharan Africa. Even assuming the existence of a stable mechanism for the occurrence of drought, global climate change completely disrupts it, making drought prediction more difficult.

2.3. Some promising works in drought forecasting

Despite the central flaw of unclear occurrence mechanisms and the disruption of the occurrence cycle by global climate change, scholars have made some promising progress in their attempts to predict drought.

Most of the existing work in drought forecasting can be summarized into two types: reduced-form approach and structural approach.

The reduced-form approach adapts and applies the State-of-the-Art (SOTA) machine learning models to drought forecasting, maximizing accuracy in a machine learning optimization setting without being overly concerned with the theoretical interpretability of the drought prediction models. Usually, this approach trains such a model:

$$\hat{Y_{t+h}} = f(\sum_{j=1}^{n}\sum_{i=1}^{t} X_{j,i})$$

where there are n features and t observable history and it seeks to learn a projection f and predict h steps into the future. Reduced-form studies do not care about how exactly features X interact in a causal way; they simply tackle this as a optimization task that aims to minimize the loss.

During training, these works usually use a mean squared error (MSE) loss for continuous drought responses:

$$Loss_{MSE} = \frac{\sum_{i=1}^{t}(y_i - \hat{y}_i)^2}{t}$$

or a categorical cross entropy (CCE) loss for index-level drought responses, in the form of:

$$Loss_{CCE} = -\sum_{i=1}^{t} y_i \cdot \log \hat{y}_i$$

The structural approach is its conceptual counterpart: the structural approach starts from the theoretical model and strives for the correctness and rationality of the theories and causal mechanisms at each step, machine learning methods or probabilistic modeling are auxiliary tools that help achieving better numerical performance.

For example, a structural model [1] would start by specifying the curve between temperature and moisture, and then, how the moisture factors in precipitation, and ultimately, drought. It will be in the shape such as:

Temperature-moisture curve:

$$p_t = a - bq_t$$

From moisture to precipitation:

$$\max E\left[\sum_{t=0}^{\infty} \delta^t (p_t - c_t) q_t (p_t)\right]$$

Drought-precipitation inter-plays:

$$\hat{q}_t = q_t + e_t$$
$$\hat{p}_t = p_t + v_t$$

From this, further structural equations can be derived, such as:

$$\hat{q}_t = \frac{a - c_t}{2b} + e_t$$
$$\hat{p}_t = \frac{a + c_t}{2} + v_t$$

Researchers build model structures based on domain knowledge and uses statistical and computational methods to optimize the prediction performance.

[1] Note that the structural equations here are for illustrative purpose only and do not represent a real structural drought model

Quite some of the reduced-form work, such as Poornima and Pushpalatha (2019), Dikshit and Pradhan (2021) uses recurrent neural networks and long short term memory (RNN LSTM) to study the time series drought. Dikshit and Pradhan. (2021) , one of the most recent scientometric analyses in the field of drought prediction confirms that reduced-form machine learning is achieving steadily increasing accuracy in drought prediction.

Structural modeling, on the other hand, has made relatively slow progress. As summarized above, the causes of drought are complex and variegated, and the conclusions and theoretical models from different geographical areas lack interchangeability and generalizability. Hao et al. (2018) points out that most of the breakthroughs in structural modeling have been happening in utilizing factors in large-scale climate indices, local climate variables, and land initial conditions. Advances in sensor technology and computing power have made such breakthroughs possible.

Ideologically, there is no advantage or disadvantage to these two types of prediction methods. Reduced form faces lighter constraints and can usually produce more accurate predictions. This is of great importance for pragmatic drought relief. Structural modeling, on the other hand, although slower in development and less accurate in prediction, can reveal to humans the factors that actually cause drought. Knowing the causal factors enables us to proactively make changes to avoid and prevent droughts.

Work Cited

Bhaga, Dube, T., Shoko, C. (2021). Satellite monitoring of surface water variability in the drought prone Western Cape, South Africa. *Physics and Chemistry of the Earth. Parts A/B/C, 124*, 102914. https://doi.org/10.1016/j.pce.2020.102914

Dikshit, Pradhan, B. (2021). Interpretable and explainable AI (XAI) model for spatial drought prediction. *The Science of the Total Environment, 801, 149797–149797*. https://doi.org/10.1016/j.scitotenv.2021.149797

Dikshit, Pradhan, B. (2021). Explainable AI in drought forecasting. *Machine Learning with Applications, 6,* 100192. https://doi.org/10.1016/j.mlwa.2021.100192

Hao, Singh, V. P., Xia, Y. (2018). Seasonal Drought Prediction: Advances, Challenges, and Future Prospects. *Reviews of Geophysics (1985), 56*(1), 108–141. https://doi.org/10.1002/2016RG000549

Jimenez, Marengo, J. A., Alves, L. M., Sulca, J. C., Takahashi, K., Ferrett, S., Collins, M. (2021). The role of ENSO flavours and TNA on recent droughts over Amazon forests and the Northeast Brazil region. *International Journal of Climatology, 41*(7), 3761–3780. https://doi.org/10.1002/joc.6453

National Drought Mitigation Center. USDM Dataset. (2022). Retrieved form https://droughtmonitor.unl.edu/Data.aspx

Poornima, Pushpalatha, M. (2019). Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units. *Atmosphere, 10*(11), 668. https://doi.org/10.3390/atmos10110668

Salehi-Lisar, Bakhshayeshan-Agdam, H. (2016). Drought Stress in Plants: Causes, Consequences, and Tolerance. In Drought Stress Tolerance in Plants, Vol 1 (pp. 1–16). *Springer International Publishing*. https://doi.org/10.1007/978-3-319-28899-4_1

Scheff. (2018). Drought Indices, Drought Impacts, CO2, and Warming: a Historical and Geologic Perspective. *Current Climate Change Reports, 4*(2), 202–209. https://doi.org/10.1007/s40641-018-0094-1

Sharafati, Nabaei, S., Shahid, S. (2020). Spatial assessment of meteorological drought features over different climate regions in Iran. *International Journal of Climatology, 40*(3), 1864–1884. https://doi.org/10.1002/joc.6307

Sundararajan, Garg, L., Srinivasan, K., Bashir, A. K., Kaliappan, J., Ganapathy, G. P., Selvaraj, S. K., Meena. (2021). A Contemporary Review on Drought Modeling Using Machine Learning Approaches. *Computer Modeling in Engineering Sciences, 128*(2), 447–487. https://doi.org/10.32604/cmes.2021.015528

Vincente-Serrano, Begueria, S., Lopez-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming. *Journal of Climate, 23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Zhang, Qu, Y., Zhang, X., Wu, X., Zhou, X., Ren, B., Zeng, J., Wang, Q. (2021). Spatiotemporal variability of annual drought severity, duration, and frequency from 1901 to 2020. *Climate Research*. https://doi.org/10.3354/cr01680