Understanding Drought through Spatial-Temporal Learning

Xuwei Tan
Department of Computer Science and Engineering
The Ohio State University
tan.1206@osu.edu

Qian Zhao
School of Earth Sciences
The Ohio State University
zhao.4243@osu.edu

Yanlan Liu
School of Earth Sciences
The Ohio State University
liu.9367@osu.edu

Xueru Zhang
Department of Computer Science and Engineering
The Ohio State University
zhang.12807@osu.edu

Abstract

Drought is one of the most destructive and expensive natural disasters, severely impacting natural resources and risks by depleting water resources and diminishing agricultural yields. Under climate change, accurately predicting drought is critical for mitigating drought-induced risks. However, the intricate interplay among the physical and biological drivers that regulate droughts limits the predictability and understanding of drought, particularly at a subseasonal to seasonal (S2S) time scale. While deep learning has demonstrated potential in addressing complex climate forecasting challenges, its application to drought prediction has received relatively less attention. Therefore, in this work, we integrate predictive features and three drought indices from multiple remote sensing and reanalysis datasets across the contiguous United States and propose a dataset specifically for predicting the spatiotemporal variation of drought: DroughtSet. DroughtSet also provides the machine learning community with a new real-world dataset to benchmark time-series forecasting methods. Furthermore, We propose a spatial-temporal model to predict and interpret S2S droughts in the contiguous U.S. Our model learns from the spatial and temporal information of physical and biological features to predict three types of droughts simultaneously. Multiple strategies are employed to quantify the importance of physical and biological features for drought prediction. These results also give insight for researchers to better understand the predictability and sensitivity of drought to biological and physical conditions. We aim to contribute to the climate field by proposing a new tool to predict and understand the occurrence of droughts and provide AI community with a new benchmark to study deep learning in the climate science field.

1 Introduction

Drought is among the most disastrous and costly natural hazards, affecting water resources, agricultural yields, heat waves, and ecosystem carbon sink [1]. There are varying definitions and types for drought, e.g. lack of precipitation, soil moisture deficit, and reduced stream flow and groundwater. As the global temperatures continue to increase, droughts are setting in quicker, becoming more intense, and hotter [2, 3]. In particular, recent studies have highlighted the increasing frequency of flash droughts, which occur when a precipitation deficit is accompanied by abnormally high temperatures

Foundation Models for Science Workshop,38th Conference on Neural Information Processing Systems (NeurIPS 2024).

and low humidity, leading to high evapotranspiration rates that quickly deplete soil moisture. Flash droughts can have significant ecological and socioeconomic impacts. For example, precipitation deficits combined with record-high temperatures in 2012 led to rapid drought development across the central US within just two months, resulting in estimated losses exceeding $30 billion.

Accurate prediction of droughts is crucial for societal preparedness and risk mitigation strategies [4, 5, 6], but it still remains a significant challenge. Existing drought prediction models include data-driven [7, 8, 9], physically-based [10, 11], and hybrid models [12]. While some of these models consider the interaction of biological drivers, their representations are generally simplified. Tackling the drought prediction challenge requires a robust and interpretable data-driven method that systematically leverages datasets of relevant climate and vegetation features [13, 14]. However, most existing methods simplify or ignore dynamic interactions among different factors, which inhibit realizing the potential of artificial intelligence (AI) to improve drought prediction accuracy and advance mechanistic understanding of flash drought. We discuss these in Appendix 6.2 included in the supplementary material.

To address this research gap, we integrate climate, physical, and vegetation conditions that are related to droughts from various remote sensing and reanalysis datasets to create the DroughtSet, which focuses on the contiguous United States (CONUS). DroughtSet considers the large diversity of climatic and ecological settings and frequent occurrences of flash droughts in recent decades. Specifically, we collect and preprocess drought-related predictors (e.g. Elevation, Temperature, and precipitation) and three types of drought indices: agricultural droughts measured by normalized surface soil moisture [15], ecohydrological drought measured by Evaporative Stress Index (ESI) [16], and ecological drought measured by solar-induced chlorophyll fluorescence (SIF) [17]. These predictors consist of both static variables and dynamic variables with coordinates. Therefore, DroughtSet could be used to benchmark multivariate forecasting, spatiotemporal forecasting, and irregular forecasting (learning from static variables). We hope to accelerate future research in climate and earth system science and benchmark deep learning-based methods by releasing this dataset.

In addition, we propose a multi-task SPatial-temporal framework for drought prediction on DroughtSet, referred to as *SPDrought*, leveraging the spatial-temporal interconnections within and across climate and vegetation features. It accounts for geographical mutual influences by aggregating temporal features with neighbor locations and learns from both static and dynamic features to predict three types of drought indices. Furthermore, we employ the Integrated Gradient (IG) method, as described in [18], to interpret and quantify how these features influence drought development across different areas. These insights will serve as a data-driven benchmark, informing further research towards enhancing the mechanistic understanding and simulation of flash droughts in existing Earth system models. This, in turn, supports the development of strategies to mitigate associated risks under future climate. Our contributions can be summarized as follows:

- We introduce DroughtSet, a drought prediction dataset for the machine learning community. It serves as a complementary resource to existing climate datasets. DroughtSet is a collection of droughts indices and the corresponding climate, physical, and vegetation conditions, specifically focusing on the contiguous U.S. [1]
- To forecast drought, we propose *SPDrought*, a spatial-temporal drought prediction model that incorporates geographic neighbor features fusion. It jointly leverages both static and dynamic features to accurately predict three key drought indices.
- We leverage the Integrated Gradient to interpret our prediction and to understand the hidden relationship among climate, physical, and vegetation features and drought indices. It gives a new insight into the complex correlations of climate systems.

2 DroughtSet

In this section, we introduce DroughtSet, a spatiotemporal collection of climate, physical, vegetation conditions, and drought indices from multiple publicly available remote sensing and reanalysis datasets. We have carefully selected these variables based on their relevance and potential influence on the mechanisms of drought development.

[1] https://github.com/osu-srml/DroughtSet

2.1 Data collection and preprocessing

DroughtSet includes nationwide weekly climate-related data spinning from 2003 to 2013 (11 years, 572 weeks) across the United States. The area of focus is the contiguous United States, which covers an area of over 8,000,000 km^2. The details of drought/feature types, variables, and their sources are outlined in Table 1. To ensure consistency in geographical resolution, all variables have been resampled to a 4 km spatial scale. Therefore, these data are represented as a grid of 585×1386 pixels across the contiguous United States, with each pixel denoted as $P(i, j)$, where (i, j) is the spatial coordinates of the pixel on the map. Note that, 42% of this pixel area consists of the ocean, where droughts do not occur and no observable metrics are available. Therefore, we only use the remaining 58% pixels to predict and analyze the results. Furthermore, all temporal variables are aggregated to a weekly time scale, detailed in Section 2.2. For the static variables, both Elevation and Canopy height are numeric variables while Land cover is a categorical variable with 97 categories. We also include the mean and standard deviation of the drought variables as the static variables. In total, the dataset comprises $585 \times 1386 \times 11 \times 52 \times 3$ Drought indices, $585 \times 1386 \times 11 \times 52 \times 11$ temporal predictors, and $585 \times 1386 \times 9$ static predictors. All variables are presented in their original units without normalization. Note that NaN values exist in the datasets due to different temporal coverages of remote sensing-derived products .

Table 1: Variables to quantify three types of drought and predictive features.

Drought/Feature Type	Variables	Dynamic or Static	Dataset & Resolution
Soil moisture drought	Soil moisture across depths (SM)	Dynamic	NLDAS [19], hourly, 1/8° SMAP [20], daily, 9 km NLDAS or SMAP blended with in situ data [21], daily, 4 km
Ecohydrological drought	Evaporative Stress Index (ESI)	Dynamic	ALEXI [22], weekly, 0.25°
Ecological drought	Solar Induced Fluorescence (SIF)	Dynamic	CSIF [23], 4-day, 0.05°
Physical & climate features	Temperature, Radiation, VPD, Precipitation, Wind Speed, PET, PDSI, SP	Dynamic	gridMET [24], daily, 4 km ERA5 [25], hourly, 9 km
	Elevation	Static	SRTM [26], 30 m
Vegetation features	Vegetation Optical Depth (VOD)	Dynamic	VODCA [27], daily, 0.25°
	Leaf area index (LAI)	Dynamic	MODIS [28], 8-day, 500 m
	Canopy Height	Static	GLAD [29], 30 m
	Land Cover	Static	NLCD [30, 31], 30 m

2.2 Drought indices and predictors

DroughtSet focuses on three droughts: soil moisture drought measured by surface soil moisture (SM), ecohydrological drought measured by Evaporative Stress Index (ESI), and ecological drought measured by the solar-induced chlorophyll fluorescence (SIF), all normalized using their quantiles at each location. These drought indices are denoted as $D_{i,j}(t) = \left[d^1_{i,j}(t), \cdots, d^K_{i,j}(t)\right]$.

- **Soil Moisture**: Surface soil moisture reflects flash drought intensity and controls the propagated impacts on the downstream drought types.
- **Evaporative Stress Index:** ESI quantifies ecohydrological drought, which is controlled by plant stomatal response to moisture deficit and could regulate drought intensification through atmospheric feedback [32].
- **Solar-induced chlorophyll Fluorescence:** SIF, a surrogate highly correlated with gross primary productivity, reflects the response of photosynthetic activity and has been shown to be sensitive to water stresses [33].

These metrics are of focus here for flash drought because, unlike other commonly used drought severity indices such as the Standardized Precipitation Index and Palmer Drought Severity Index (PDSI) that are typically used to capture interannual drought, these metrics have been demonstrated to respond quickly and have wide-ranging implications on water resources, ecosystem carbon sink strength, and agricultural productivity [15, 32, 33, 34, 35, 36]. In our work, these drought indices serve as prediction targets and our objective is to accurately predict potential future droughts. Figure 1 visualizes the evaporative stress, using the 2012 central US drought as an example, which illustrates the pattern of a drought index.

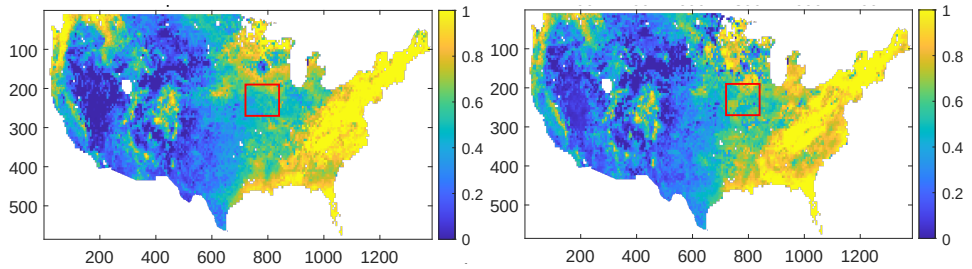


Figure 1: A S2S drought in July 2012. The left figure represents evaporative stress for the 26th week and the right figure for the 28th week. The area outlined by the red box was affected by the drought in 2012, where increased evaporative stress are evident.

To predict the three types of drought, we select attributes including climate and physical conditions, vegetation dynamics, and drought indices themselves in lagged time steps as predictive features for drought prediction. These features are considered as predictors for droughts and are categorized into temporal dynamic attributes and spatial static attributes that consist of numeric and categorical data. These features are described in the following.

Physical and climate conditions:

- **Elevation**: Elevation is a static numeric variable. It influences local climate conditions which can affect drought due to variations in temperature and precipitation patterns.
- **Air Temperature**: Higher air temperatures can increase evaporation rates, accelerating the speed of drought onset. We aggregate temperature by averaging the daily temperature for one week.
- **Precipitation**: Precipitation determines the amount of water input. Precipitation deficit directly leads to drought. We aggregate it by summing daily precipitation for one week.
- **Radiation**: Solar radiation drives evapotranspiration and higher radiation can increase water loss from soil and vegetation. We use the downward shortwave radiation and aggregate it by averaging the daily radiation for one week.
- **Vapor Pressure Deficit (VPD)**: VPD characterizes the degree of atmospheric moisture deficit and directly influences the evapotranspiration rates. We aggregate it by averaging the daily VPD for one week.
- **Wind Speed**: Wind speed influences evapotranspiration rates, thereby contributing to drought development. We aggregate it by averaging the daily value for one week.
- **Potential Evapotranspiration (PET)**: PET represents the amospheric water demand. We aggregate it by averaging the daily PET for one week.
- **PDSI**: It quantifies the severity of meteorological drought in an area using temperature and precipitation data, which primarily characterizes long-term drought conditions.
- **Surface Pressure (SP)**: SP is the atmospheric pressure at Earth's surface. Changes in surface pressure can regulate weather patterns, atmospheric moisture supply, and thus drought development.
- **SM Root**: Different from surface soil moisture, soil moisture at the root zone measures the amount of water available in the soil where most plant roots are located.

Vegetation dynamics:

- **Biomass dynamics measured by Leaf Area Index (LAI)**: LAI represents the leaf area per ground unit area. We aggregate it by linearly interpolating the raw data with an 8-day temporal resolution.
- **Vegetation Optical Depth (VOD)**: VOD represents vegetation water content. We aggregate it by averaging the daily VOD for one week.
- **Canopy Height**: Canopy height is a static categorical variable, representing ecosystem structure.

Auxiliary data:

- **Land Cover**: It is a static categorical variable, including categories such as forests, water bodies, and grasslands.

With these climate-related drought indices and attributes, each pixel has N static features $S_{i,j} = \left[s_{i,j}^1, \cdots, s_{i,j}^N\right]$ (i.e., land cover, elevation, canopy height, long-term averages of drought indices and their standard deviations to capture variability) and M dynamic features $X_{i,j}(t) = \left[x_{i,j}^1(t), \cdots, x_{i,j}^M(t)\right]$. The goal is to train a machine learning model $h \in \mathcal{H}$ from existing static and dynamic features $\{S_{i,j}, D_{i,j}(t), X_{i,j}(t)\}_{i\in[I],j\in[J],t\in[T]}$ that can simultaneously predict multiple drought indices $\{D_{i,j}(T+\tau)\}_{\tau\geq 1}$ for any location (i,j) in the future. Because both drought indices $D_{i,j}(t)$ and dynamic features $X_{i,j}(t)$ are time-varying and jointly used for predictive tasks, we combine them and define $U_{i,j}(t) = [D_{i,j}(t), X_{i,j}(t)]$.

In addition, as DroughtSet consists of both static features and temporal features for each location with geographic coordinates, it offers a versatile platform for benchmarking various forecasting methods. This dataset can be utilized in univariate forecasting tasks, which focus on directly predicting drought indices from single variables. It also supports multivariate forecasting, where multiple variables are used jointly to predict drought indices. Furthermore, DroughtSet is ideal for spatiotemporal forecasting, which leverages both spatial and temporal information to enhance prediction accuracy. Lastly, it can be employed in irregular forecasting tasks that jointly use static and temporal features, providing a comprehensive tool for advanced drought prediction models.

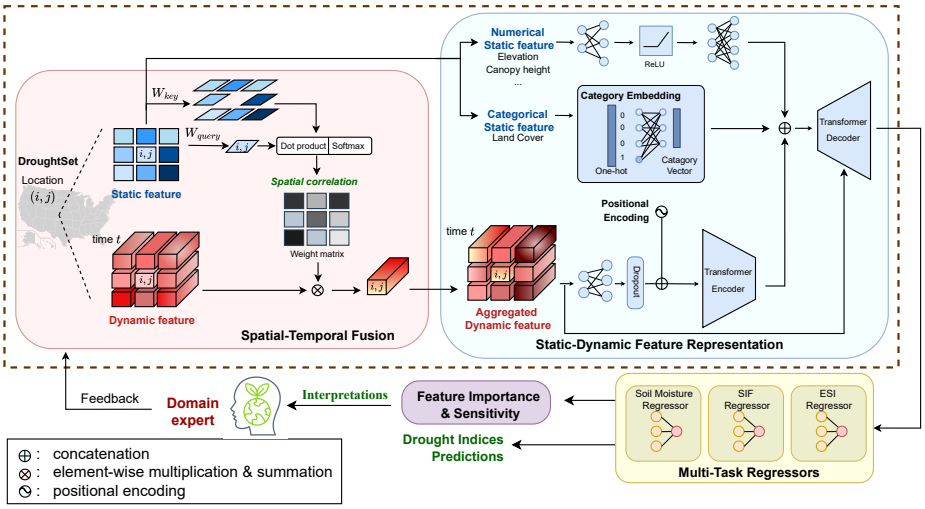


Figure 2: *SPDrought* architecture for Forecasting Drought Indices: the spatial-temporal fusion module first exploits the spatial correlation of data with its neighbors using static features and leverages the learned correlation to aggregate the dynamic features; the static-dynamic feature representation exploits both spatial and temporal patterns with three network modules. Such representation is shared among multi-task regressors for generating multiple drought indices predictions. Subsequently, we analyze how individual features at various timestamps influence the final predictions using our interpretation method. Domain experts are encouraged to provide feedback on variable selection and model design, which can further refine the model and uncover deeper relationships among variables.

3 Method

Next, we introduce a comprehensive framework that utilizes both spatial and temporal information to predict drought indices. Our approach considers regional spatial similarity to aggregate information for robust prediction and introduces climate attribute-specific representation functions to learn from the hidden pattern of both static and time-series data.

Spatial-Temporal Fusion. Since climate information from proximate geographical locations often exhibits mutual influences, we hypothesize that data from neighboring locations may contain useful information that can help enhance the accuracy and reliability of the prediction. The key challenge is to exploit the spatial correlation and strategically leverage the learned correlation to enhance

prediction at the target location. Intuitively, the target location may benefit more from those neighbors that are sufficiently correlated, e.g., sharing similar topography or land cover. Let neighborhood $\mathcal{N}_{i,j} = \{(\bar{i}, \bar{j}) \mid (|\bar{i} - i| \leq d, |\bar{i} - j| \leq d, (\bar{i}, \bar{j}) \neq (i, j))\}$, where d is a distance threshold. Inspired by scaled dot-product attention mechanism [37], we exploit the spatial correlation between any target location (i, j) (known as a query in attention mechanism) and neighbors in $\mathcal{N}_{i,j}$ (known as keys) based on the following:

$$A = \text{softmax}\left(\frac{(S_{i,j}W_{\text{query}})(S_{\mathcal{N}_{i,j}}W_{\text{key}})^T}{R_{i,j} \times \sqrt{N}}\right) \tag{1}$$

where $S_{i,j}$ are N static features at (i, j), $S_{\mathcal{N}_{i,j}}$ is a matrix with each column the static features corresponding to one neighbor in $\mathcal{N}_{i,j}$. $W_{query} \in \mathbb{R}^{N\times N}$ and $W_{key} \in \mathbb{R}^{N\times N}$ are two linear transformation matrices that are learned to exploit spatial correlation. $R_{i,j}$ is a vector with each element the Euclidean distance between (i, j) and neighbors in $\mathcal{N}_{i,j}$, which leverages the prior spatial information to refine correlation learning process. For simplicity, we consider a 5×5 square area in this study, where d is set to 2 in this paper. To avoid division by zero, we manually set the distance to itself as 0.8. The spatial correlation weight A can then be used to aggregate the regional time-varying attributes: $\tilde{U}_{\bar{i},\bar{j}}(t) = \Sigma_{\bar{i},\bar{j}\in\mathcal{N}_{(i,j)}} A_{\bar{i},\bar{j}} \cdot U_{\bar{i},\bar{j}}(t)$.

Spatial-Temporal Representation and Multi-task Learning. Given $\{S_{i,j}, \widetilde{U}_{i,j}(t)\}$, we next learn the representations of the climate data, which combine static and dynamic feature representations generated by separated networks:

- Static feature representation: Given a set of static features $S_{i,j} = \left[s^1_{i,j}, \cdots, s^N_{i,j}\right]$, we aim to obtain higher-level representation that encapsulates the underlying patterns among them. Because categorical (land cover type) and numerical (elevation, canopy height, long-term averages, and standard deviations of SM, SIF, and ESI) features have inherent differences in semantic meanings, we shall generate their representations differently. We apply two layers of MLP linked by the ReLU function to learn the representations of numerical features and adopt embedding approaches for categorical features to generate their representations, which we denote as $f^s_{i,j}$.
- Dynamic feature representation: To learn the temporal patterns, especially the long-term dependencies of climate data, we first adapt *Transformer* [37] encoder to generate temporal representations. Before the *Transformer*, we expand the initial temporal features $\widetilde{U}_{i,j}(t)$ of our data (including K drought indices and M dynamic features with a total dimension of 14) via linear transformation W and project the dimensions to 48. This linear transformation also facilitates learning the interconnections among these distinct dynamic features. After integrating the *positional encoding* $\text{PE}(t)$, we generate temporal representation $f^t_{i,j}(t) = \text{TransformerEncoder}\left(\widetilde{U}_{i,j}(t) \cdot W + \text{PE}(t)\right)$. We then concatenate the static representations to dynamic feature representations at each time stamp. The concatenated representations are fed into the *Transformer decoder* to generate representations $\{F_{i,j}(t')\}_{t'\in\{T+1\ldots T+26\}}$ for the next 26 weeks .

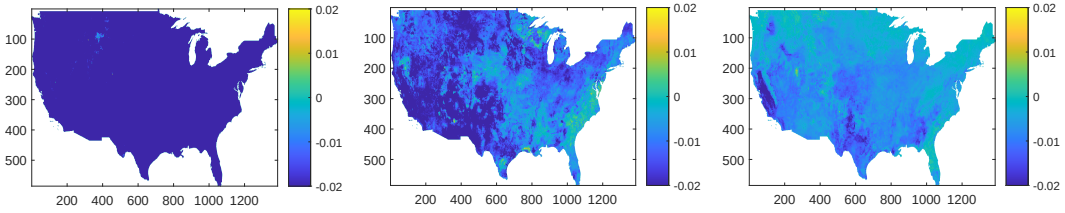
With the representation $\{F_{i,j}(t')\}_{t'\in\{T+1\ldots T+26\}}$ for the prediction weeks, we employ three *task-specific* regressors to map the representation of the next 26 weeks to drought indices. Specifically, let $\widehat{d}^k_{i,j}(t') = \text{Regressor}_k\left(F_{i,j}(t')\right)$ be the prediction of k-th drought index $d^k_{i,j}(t')$ after t' weeks. We jointly train all the parameters by minimizing the total loss between predictions $\widehat{d}^k_{i,j}(t')$ and ground-truth $d^k_{i,j}(t')$ for all drought indices at all locations (we use a batch of locations to update the model at every iteration in implementation) under mean absolute error loss function $\mathcal{L}$:

$$\min \sum_{k\in[K]} \sum_{t'\in\{T+1\ldots T+26\}} \sum_{i\in[I],j\in[J]} \mathcal{L}\left(\widehat{d}^k_{i,j}(t'), d^k_{i,j}(t')\right) \tag{2}$$

Drought Understanding. To identify which attributes and time steps contribute most to the final predictions, we leverage integrated gradient [18] to investigate how their contributions to final predictions change over time (i.e., how *sensitive* the predictions are to these features). It finds the feature sensitivities by integrating the gradients of the model's output with respect to the input along a straight path from a "baseline" to the input. Then we quantify the importance of static variables by looking at the features that cause the larger performance change and are more important for prediction tasks.

Table 2: Average mean absolute error over three runs of experiments

MAE ($\times 10^{-3}$)	SPDrought	Transformer	Informer	PatchTST	DLinear	iTransformer	TimesNet	LSTM
Soil Moisture	$21.39_{\pm 0.14}$	$34.56_{\pm 0.24}$	$38.08_{\pm 0.14}$	$36.32_{\pm 0.19}$	$47.61_{\pm 0.04}$	$32.34_{\pm 0.09}$	$25.96_{\pm 0.46}$	$31.36_{\pm 0.40}$
Evaporative Stress Index	$4.40_{\pm 0.02}$	$5.99_{\pm 0.07}$	$6.37_{\pm 0.06}$	$6.37_{\pm 0.00}$	$6.82_{\pm 0.01}$	$6.06_{\pm 0.01}$	$5.11_{\pm 0.02}$	$5.83_{\pm 0.09}$
Solar-induced chlorophyll Fluorescence	$12.21_{\pm 0.23}$	$16.00_{\pm 0.23}$	$17.71_{\pm 0.33}$	$21.36_{\pm 0.19}$	$20.99_{\pm 0.03}$	$15.47_{\pm 0.07}$	$14.11_{\pm 0.03}$	$15.35_{\pm 0.02}$
Total	$38.01_{\pm 0.35}$	$56.56_{\pm 0.05}$	$62.16_{\pm 0.43}$	$64.05_{\pm 0.34}$	$75.41_{\pm 0.05}$	$53.87_{\pm 0.16}$	$45.18_{\pm 0.44}$	$52.54_{\pm 0.40}$



(a) Influence of Surface Pressure (b) Influence of Radiation (c) Influence of PET

Figure 3: Interpretation on **Soil Moisture** prediction. Surface Pressure, Radiation, and PET are the most significant variables influencing surface soil moisture.

4 Experiments

4.1 Experimental setup

Training details. All experiments are conducted on a server equipped with multiple NVIDIA V100 GPUs, Intel Xeon(R) Platinum 8260 CPU, and 256GB memory. The code is implemented with Python 3.9 and PyTorch 1.10.0.

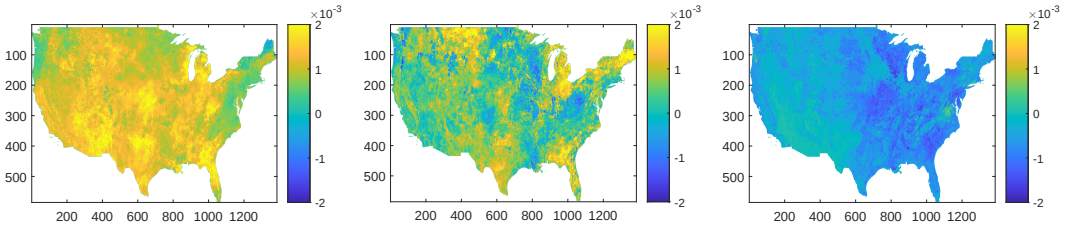
In this study, we split the pixels by 5×5 pixel block to avoid similar neighboring pixels and randomly select 80% blocks as training pixels and the remaining 20% for testing. Each pixel in our dataset has 572 weeks of temporal features and drought indices. We divide these 572 weeks into multiple windows for training and analysis. Each window consists of 100 weeks (approximately 2 years) designated as the training period, followed by 26 weeks (approximately half a year) designated as the prediction period. We then slide this window forward by 26 weeks (half a year) at a time, creating a total of 18 overlapping windows. To mitigate the impact of missing values (NaN) in the dataset, we impute the yearly average value for each week to maintain seasonal trends. Training is skipped for any NaN values in drought indices. Additionally, before training, we normalize each predictor and drought index by dividing it by its maximum value, scaling all values to a range between 0 and 1.

During the training, we sample a batch of pixels randomly and shuffle the order of these windows to sequentially update the model. We train the model over 30 epochs where an epoch is defined as each training pixel being visited and trained once. After filtering out ocean locations, where most variables are NaN, the number of effective training pixels totals 380,801. The test set comprises 93,220 effective pixels. We set the batch size to 32, and employ Adam optimizer with a learning rate of 1e-4. The mean absolute error is used as the loss function. For categorical variable land cover, the embedding dimension is set at 4. For static numeric variables, the MLP uses a hidden dimension of 10 and an output dimension of 16. Temporal features are first processed through a linear layer with a dropout rate of 0.1, mapping the dimension from 14 to 48. Then three layers of Transformer encoders and two layers of decoders with dimensions of 256 and 2 attention heads are used to learn from the projected temporal features.

Baselines We consider state-of-the-art deep learning methods for time-series forecasting as baselines to evaluate our method. Note that these methods are mostly designed for time-series features without considering static features, including Transformer [37], Informer [38], PatchTST [39], DLinear [40], iTransformer [41], TimesNet [42], and LSTM [43]. We introduce the details of each baseline in Appendix 6.1.

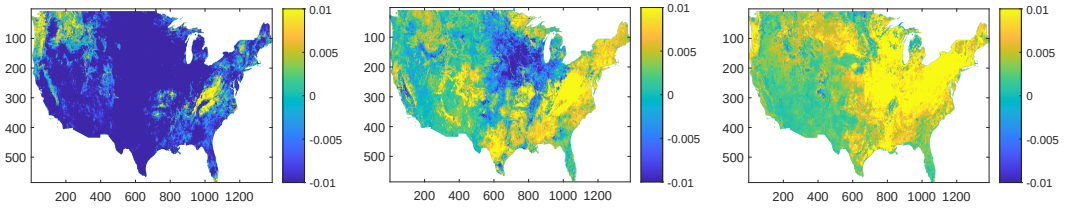
4.2 Results

Performance Comparison. We first compare *SPDrought* with five widely recognized time-series forecasting baseline models. Table 2 presents the average mean absolute error across three runs



(a) Influence of Radiation (b) Influence of Pressure (c) Influence of SIF

Figure 4: Interpretation on **Evaporative Stress Index**. Radiation and Pressure show a positive influence on the evaporative stress index while SIF reflects a minor negative influence.



(a) Influence of Pressure (b) Influence of Radiation (c) Influence of SM root

Figure 5: Interpretation on **Solar-induced Fluorescence**. Both radiation and root-zone soil moisture directly influence the rate of photosynthesis, which in turn affects the SIF signal.

on DroughtSet. The results demonstrate that *SPDrought* has superior forecasting performance at forecasting 26 weeks of three drought indices at test locations compared with baselines. This outcome underscores its effectiveness in capturing the dynamics of the variables under study—Soil Moisture, Evaporative Stress Index, and Solar-induced Chlorophyll Fluorescence.

Among the baselines, *DLinear* has previously shown robust performance, outperforming several transformer-based methods [40]. *DLinear* decomposes the time series and uses two linear layers for trend and abnormality respectively. However, *DLinear* encounters challenges with drought indices forecasting tasks because it uses prediction variables (drought indices) independently rather than leveraging all predictors and indices together. In contrast, transformer-based methods typically account for patterns among variables, resulting in better performance in our tasks. It highlights the importance of learning the interplay of predictors to improve forecasting performance in drought prediction challenges. We also compare our model with a vanilla *Transformer*. Except for the main difference in using static features, the *Transformer* baseline uses embedding for temporal input tokenization as same as other methods, while our model considers the linear transformation for the temporal features to learn the representation across predictors.

Drought Interpretation. To examine the relative contribution of features to a drought, we study the flash drought in July 2012 in the US. Specifically, we leverage integrated gradient to analyze the prediction in July 2012 and use gradient information to compare the contribution of features to the drought. Here, we evaluate the influence of each variable on soil moisture prediction by comparing the integrated gradient value in Figure 3. We select and present the top 3 significant influence variables. Pressure, Radiation, and PET show strong negative contributions (negative gradients) to soil moisture, which means a higher value of these metrics may lead to a decrease in soil moisture and indicate a potential drought. For example, high-pressure surface climate typically leads to drier weather, reduces rainfall, which in turn leads to drier soil conditions. In contrast, low-pressure systems are often associated with increased cloud cover and precipitation, which can enhance soil moisture [44]. Our model effectively captures the negative influence of pressure on soil moisture. Similar to radiation, high radiation heats the soil surface, increasing soil evaporation and thus reducing soil moisture. Thus, our model is able to capture the relationship between radiation and soil moisture. Figure 4 highlights the significant predictors influencing the Evaporative Stress Index, including Radiation, Pressure, and SIF. Nationwide, higher levels of radiation and surface pressure are indicative of increased evaporative stress [44, 45], which aligns with domain knowledge. Through comparative analysis, it

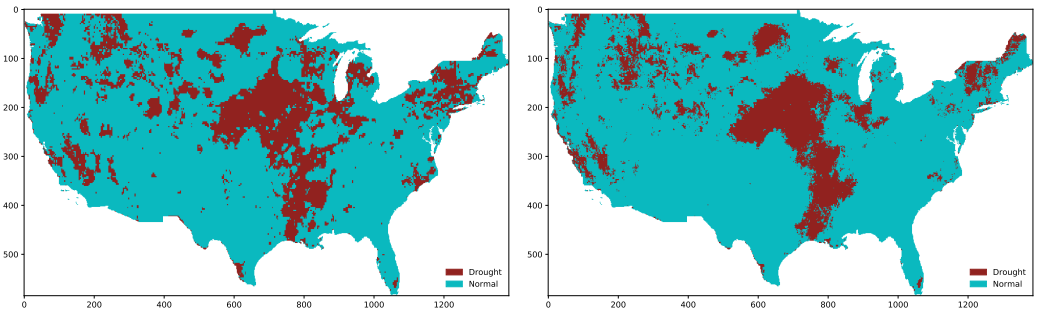
can be found that radiation should be prioritized when analyzing the drought type forecasted by the evaporative stress. Figure 5 shows pressure, radiation, and root zone soil moisture largely influence solar-induced fluorescence. These observed results are consistent with established meteorological principles. The results on interoperability reveal the relative importance of these predictors to each drought indices, which contributes to discipline-specific insights to understand the development and propagation of droughts.

Assessment of Drought Using Soil Moisture Percentiles In this section, we use the soil moisture as an example to assess drought. We employ a percentile-based approach [46] using soil moisture data. Each measurement of weekly soil moisture is compared against a multi-year average for the same calendar week, derived from historical data to represent typical moisture levels. We calculate the deviation of current soil moisture levels from these averages. We then use the 30th percentile as the threshold in our analysis, values below this percentile are indicative of drought conditions. Our method is compared with baselines in terms of accuracy and precision and the results on the test set are reported in Table 3.

Table 3: Evaluation of drought prediction by soil moisture. The standard deviations are reported in Appendix.

	SPDrought	**Transformer**	**Informer**	**PatchTST**	**DLinear**	**iTransformer**	**TimesNet**	**LSTM**
Accuracy	86.26	76.09	72.16	62.24	62.85	77.18	81.54	77.45
Precision	76.80	59.94	53.40	36.94	37.96	61.74	68.98	62.18

We visualize our prediction using the drought in 2012. In Figure 6, we use soil moisture ground truth to indicate the drought and present Figure 6.a. Figure 6.b shows the drought indicated by the soil moisture prediction from *SPDrought*. In addition, we conduct ablation studies on static variables and model components, which are reported in Appendix 6.3.



(a) Drought indicated by Soil Moisture ground truth (b) Drought indicated by Soil Moisture Prediction

Figure 6: Drought Prediction using Soil Moisture

5 Conclusion

This paper introduces DroughtSet, a specialized time-series forecasting dataset designed for predicting drought indices. It integrates vegetation and climate predictors, incorporating static and dynamic features. Based on DroughtSet, we also propose a framework, *SPDrought*, which leverages spatial-temporal interactions to accurately predict drought indices, and interpret the prediction results to advance our understanding of drought development and propagation.

Limitation and Future Work. This paper focuses exclusively on the contiguous U.S., as the dataset is only collected for this region. Therefore, the trained model is not suitable for direct deployment in other regions because of the geographical differences. However, our method is not limited to CONUS and is expected to be effective in other regions, provided that suitable data covering both static and dynamic predictors is available. In this study, we primarily examine the spatial influence of physical and climate conditions, as well as vegetation dynamics, on drought indices. In future case studies, we

will conduct a comprehensive analysis of the temporal interplay and dependencies related to flash droughts.

Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under award IIS-2202699 and IIS-2416895, by OSU President's Research Excellence Accelerator Grant, and grants from the Ohio State University's Translational Data Analytics Institute and College of Engineering Strategic Research Initiative.

References

[1] Benjamin I Cook, Justin S Mankin, and Kevin J Anchukaitis. Climate change and drought: From past to future. *Current Climate Change Reports*, 4:164–179, 2018.

[2] Kevin E Trenberth, Aiguo Dai, Gerard Van Der Schrier, Philip D Jones, Jonathan Barichivich, Keith R Briffa, and Justin Sheffield. Global warming and changes in drought. *Nature Climate Change*, 4(1):17–22, 2014.

[3] Kumar P Tripathy, Sourav Mukherjee, Ashok K Mishra, Michael E Mann, and A Park Williams. Climate change will accelerate the high-end risk of compound drought and heatwave events. *Proceedings of the National Academy of Sciences*, 120(28):e2219825120, 2023.

[4] Jason A Otkin, Mark Svoboda, Eric D Hunt, Trent W Ford, Martha C Anderson, Christopher Hain, and Jeffrey B Basara. 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(5):911–919, 2018.

[5] Christopher J White, Henrik Carlsen, Andrew W Robertson, Richard JT Klein, Jeffrey K Lazo, Arun Kumar, Frederic Vitart, Erin Coughlan de Perez, Andrea J Ray, Virginia Murray, et al. Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological applications*, 24(3):315–325, 2017.

[6] Christopher J White, Daniela IV Domeisen, Nachiketa Acharya, Elijah A Adefisan, Michael L Anderson, Stella Aura, Ahmed A Balogun, Douglas Bertram, Sonia Bluhm, David J Brayshaw, et al. Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, 103(6):E1448–E1472, 2022.

[7] Brunella Bonaccorso, Antonino Cancelliere, and Giuseppe Rossi. Probabilistic forecasting of drought class transitions in sicily (italy) using standardized precipitation index and north atlantic oscillation index. *Journal of Hydrology*, 526:136–150, 2015.

[8] João Filipe Santos, Maria Manuela Portela, and Inmaculada Pulido-Calvo. Spring drought prediction based on winter nao and global sst in portugal. *Hydrological Processes*, 28(3): 1009–1024, 2014.

[9] Niko Wanders and Yoshihide Wada. Decadal predictability of river discharge with climate oscillations over the 20th and early 21st century. *Geophysical Research Letters*, 42(24):10–689, 2015.

[10] Niko Wanders and Eric F Wood. Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, 11(9):094007, 2016.

[11] Aihui Wang, Theodore J Bohn, Sarith P Mahanama, Randal D Koster, and Dennis P Lettenmaier. Multimodel ensemble reconstruction of drought over the continental united states. *Journal of Climate*, 22(10):2694–2712, 2009.

[12] Zhiyong Wu, Hao Yin, Hai He, and Yuan Li. Dynamic-lstm hybrid models to improve seasonal drought predictions over china. *Journal of Hydrology*, 615:128706, 2022.

[13] A AghaKouchak, B Pan, O Mazdiyasni, M Sadegh, S Jiwa, W Zhang, CA Love, S Madadgar, SM Papalexiou, SJ Davis, et al. Status and prospects for drought forecasting: Opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philosophical Transactions of the Royal Society A*, 380(2238):20210288, 2022.

[14] Zengchao Hao, Vijay P Singh, and Youlong Xia. Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, 56(1):108–141, 2018.

[15] Xing Yuan, Yumiao Wang, Peng Ji, Peili Wu, Justin Sheffield, and Jason A Otkin. A global transition to flash droughts under climate change. *Science*, 380(6641):187–191, 2023.

[16] Jason A Otkin, Martha C Anderson, Christopher Hain, and Mark Svoboda. Examining the relationship between drought development and rapid changes in the evaporative stress index. *Journal of Hydrometeorology*, 15(3):938–956, 2014.

[17] Koushan Mohammadi, Yelin Jiang, and Guiling Wang. Flash drought early warning based on the trajectory of solar-induced chlorophyll fluorescence. *Proceedings of the National Academy of Sciences*, 119(32):e2202767119, 2022.

[18] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[19] Youlong Xia, Kenneth Mitchell, Michael Ek, Justin Sheffield, Brian Cosgrove, Eric Wood, Lifeng Luo, Charles Alonge, Helin Wei, Jesse Meng, et al. Continental-scale water and energy flux analysis and validation for the north american land data assimilation system project phase 2 (nldas-2): 1. intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, 117(D3), 2012.

[20] Narendra N Das, Dara Entekhabi, R Scott Dunbar, Andreas Colliander, Fan Chen, Wade Crow, Thomas J Jackson, Aaron Berg, David D Bosch, Todd Caldwell, et al. The smap mission combined active-passive soil moisture product at 9 km and 3 km spatial resolutions. *Remote sensing of environment*, 211:204–217, 2018.

[21] Ning Zhang, Steven M Quiring, and Trent W Ford. Blending noah, smos, and in situ soil moisture using multiple weighting and sampling schemes. *Journal of Hydrometeorology*, 22(7): 1835–1854, 2021.

[22] S Hook and G Hulley. Ecostress land surface temperature and emissivity daily l2 global 70 m v001 [data set]. *NASA EOSDIS Land Processes DAAC. doi*, 10, 2019.

[23] Yao Zhang, Joanna Joiner, Seyed Hamed Alemohammad, Sha Zhou, and Pierre Gentine. A global spatially contiguous solar-induced fluorescence (csif) dataset using neural networks. *Biogeosciences*, 15(19):5779–5800, 2018.

[24] John T Abatzoglou. Development of gridded surface meteorological data for ecological applications and modelling. *International journal of climatology*, 33(1):121–131, 2013.

[25] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9):4349–4383, 2021.

[26] NASA JPL. NASA Shuttle Radar Topography Mission Global 1 arc second. `https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003`, 2013. URL `https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003`. Accessed 2024-05-17.

[27] Leander Moesinger, Wouter Dorigo, Richard De Jeu, Robin Van Der Schalie, Tracy Scanlon, Irene Teubner, and Matthias Forkel. The global long-term microwave vegetation optical depth climate archive (vodca). *Earth System Science Data*, 12(1):177–196, 2020.

[28] R Myneni, Y Knyazikhin, and T Park. Mcd15a3h modis/terra+ aqua leaf area index/fpar 4-day l4 global 500 m sin grid v006, nasa eosdis land processes daac [data set], 2015.

[29] Peter Potapov, Xinyuan Li, Andres Hernandez-Serna, Alexandra Tyukavina, Matthew C Hansen, Anil Kommareddy, Amy Pickens, Svetlana Turubanova, Hao Tang, Carlos Edibaldo Silva, et al. Mapping global forest canopy height through integration of gedi and landsat data. *Remote Sensing of Environment*, 253:112165, 2021.

[30] Collin G Homer, Joyce A Fry, and Christopher A Barnes. The national land cover database. Technical report, US Geological Survey, 2012.

[31] Collin Homer, Jon Dewitz, Suming Jin, George Xian, Catherine Costello, Patrick Danielson, Leila Gass, Michelle Funk, James Wickham, Stephen Stehman, et al. Conterminous united states land cover change patterns 2001–2016 from the 2016 national land cover database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:184–199, 2020.

[32] Hanh Nguyen, Matthew C Wheeler, Jason A Otkin, Tim Cowan, Andrew Frost, and Roger Stone. Using the evaporative stress index to monitor flash drought in australia. *Environmental Research Letters*, 14(6):064016, 2019.

[33] Mingzhu He, John S Kimball, Yonghong Yi, Steve Running, Kaiyu Guan, Kelsey Jensco, Bruce Maxwell, and Marco Maneta. Impacts of the 2017 flash drought in the us northern plains informed by satellite-based evapotranspiration and solar-induced fluorescence. *Environmental Research Letters*, 14(7):074019, 2019.

[34] Jordan I Christian, Jeffrey B Basara, Eric D Hunt, Jason A Otkin, Jason C Furtado, Vimal Mishra, Xiangming Xiao, and Robb M Randall. Global distribution, trends, and drivers of flash drought occurrence. *Nature communications*, 12(1):6330, 2021.

[35] Trent W Ford, Jason A Otkin, Steven M Quiring, Joel Lisonbee, Molly Woloszyn, Junming Wang, and Yafang Zhong. Flash drought indicator intercomparison in the united states. *Journal of Applied Meteorology and Climatology*, 62(12):1713–1730, 2023.

[36] RD Koster, SD Schubert, H Wang, SP Mahanama, and Anthony M DeAngelis. Flash drought as captured by reanalysis data: Disentangling the contributions of precipitation deficit and excess evapotranspiration. *Journal of Hydrometeorology*, 20(6):1241–1258, 2019.

[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[38] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[39] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

[40] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

[41] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2023.

[42] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The eleventh international conference on learning representations*, 2022.

[43] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

[44] Gordon Bonan. *Climate change and terrestrial ecosystem modeling*. Cambridge University Press, 2019.

[45] Jason A Otkin, Yafang Zhong, David Lorenz, Martha C Anderson, and Christopher Hain. Exploring seasonal and regional relationships between the evaporative stress index and surface weather and soil moisture anomalies across the united states. *Hydrology and Earth System Sciences*, 22(10):5373–5386, 2018.

[46] Linying Wang and Xing Yuan. Two types of flash drought and their connections with seasonal drought. *Advances in Atmospheric Sciences*, 35(12):1478–1490, 2018.

[47] Wentao Yu, Jing Li, Qinhuo Liu, Jing Zhao, Yadong Dong, Cong Wang, Shangrong Lin, Xinran Zhu, and Hu Zhang. Spatial–temporal prediction of vegetation index with deep recurrent neural networks. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.

[48] Ali Danandeh Mehr, Amir Rikhtehgar Ghiasi, Zaher Mundher Yaseen, Ali Unal Sorman, and Laith Abualigah. A novel intelligent deep learning predictive model for meteorological drought forecasting. *Journal of Ambient Intelligence and Humanized Computing*, 14(8):10441–10455, 2023.

[49] Mohd Imran Khan and Rajib Maity. Development of a long-range hydrological drought prediction framework using deep learning. *Water Resources Management*, 38(4):1497–1509, 2024.

[50] Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572, 2019.

[51] Young-Jae Park, Minseok Seo, Doyi Kim, Hyeri Kim, Sanghoon Choi, Beomkyu Choi, Jeongwon Ryu, Sohee Son, Hae-Gon Jeon, and Yeji Choi. Long-term typhoon trajectory prediction: A physics-conditioned approach without reanalysis data. In *The Twelfth International Conference on Learning Representations*, 2023.

[52] Sungwon Park, Karandeep Singh, Arjun Nellikkattil, Elke Zeller, Tung Duong Mai, and Meeyoung Cha. Downscaling earth system models with deep learning. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 3733–3742, 2022.

[53] Yuxuan Liang, Yutong Xia, Songyu Ke, Yiwei Wang, Qingsong Wen, Junbo Zhang, Yu Zheng, and Roger Zimmermann. Airformer: Predicting nationwide air quality in china with transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14329–14337, 2023.

[54] Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. Climax: A foundation model for weather and climate. In *International Conference on Machine Learning*, pages 25904–25938. PMLR, 2023.

[55] Lujun Zhang, Taereem Kim, Tiantian Yang, Yang Hong, and Qian Zhu. Evaluation of subseasonal-to-seasonal (s2s) precipitation forecast from the north american multi-model ensemble phase ii (nmme-2) over the contiguous us. *Journal of Hydrology*, 603:127058, 2021.

[56] Kingtse C Mo and Dennis P Lettenmaier. Prediction of flash droughts over the united states. *Journal of Hydrometeorology*, 21(8):1793–1810, 2020.

[57] Chrisgone Adede, Robert Oboko, Peter Waiganjo Wagacha, and Clement Atzberger. A mixed model approach to vegetation condition prediction using artificial neural networks (ann): case of kenya's operational drought monitoring. *Remote Sensing*, 11(9):1099, 2019.

[58] Ahmad Fatehi Marj and Allard MJ Meijerink. Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *International Journal of Remote Sensing*, 32(24):9707–9719, 2011.

[59] Aya Ferchichi, Ali Ben Abbes, Vincent Barra, and Imed Riadh Farah. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. *Ecological Informatics*, 68:101552, 2022.

[60] Amogh Gyaneshwar, Anirudh Mishra, Utkarsh Chadha, PM Durai Raj Vincent, Venkatesan Rajinikanth, Ganapathy Pattukandan Ganapathy, and Kathiravan Srinivasan. A contemporary review on deep learning models for drought prediction. *Sustainability*, 15(7):6160, 2023.

[61] Norbert A Agana and Abdollah Homaifar. A deep learning based approach for long-term drought prediction. In *SoutheastCon 2017*, pages 1–8. IEEE, 2017.

[62] Foyez Ahmed Prodhan, Jiahua Zhang, Fengmei Yao, Lamei Shi, Til Prasad Pangali Sharma, Da Zhang, Dan Cao, Minxuan Zheng, Naveed Ahmed, and Hasiba Pervin Mohana. Deep learning for monitoring agricultural drought in south asia using remote sensing data. *Remote Sensing*, 13(9):1715, 2021.

[63] Amandeep Kaur and Sandeep K Sood. Deep learning based drought assessment and prediction framework. *Ecological Informatics*, 57:101067, 2020.

[64] Tuong Quang Vo, Seon-Ho Kim, Duc Hai Nguyen, and Deg-Hyo Bae. Lstm-cm: a hybrid approach for natural drought prediction based on deep learning and climate models. *Stochastic Environmental Research and Risk Assessment*, 37(6):2035–2051, 2023.

[65] Thomas Lees, Gabriel Tseng, Clement Atzberger, Steven Reece, and Simon Dadson. Deep learning for vegetation health forecasting: a case study in kenya. *Remote Sensing*, 14(3):698, 2022.

[66] Abhirup Dikshit and Biswajeet Pradhan. Interpretable and explainable ai (xai) model for spatial drought prediction. *Science of the Total Environment*, 801:149797, 2021.

[67] Amobichukwu C Amanambu, Joann Mossa, and Yin-Hsuen Chen. Hydrological drought forecasting using a deep transformer model. *Water*, 14(22):3611, 2022.

[68] Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang. Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 4365–4373, 2023.

[69] Shruti Kaushik, Abhinav Choudhury, Pankaj Kumar Sheron, Nataraj Dasgupta, Sayee Natarajan, Larry A Pickett, and Varun Dutt. Ai in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in big data*, 3:4, 2020.

[70] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

[71] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pages 27268–27286. PMLR, 2022.

A Appendix

A.1 Baselines

We ensured consistent representation dimensions and batch sizes across all baseline methods and adjusted model layers to maintain comparable training times. Additionally, other method-specific hyperparameters are adjusted to improve performance and ensure similar training times across all methods.

- Transformer [37]: A vanilla Transformer for the time-series forecasting task.
- Informer [38]: Informer introduces the ProbSparse self-attention mechanism for efficiently capturing long-range dependencies in time series forecasting.
- PatchTST [39]: PatchTST applies patching techniques to time series data, enhancing the transformer architecture's performance in capturing temporal patterns.
- DLinear [40]: DLinear is a simplified linear model which succeeds many transformer-based models.
- iTransformer [41]: iTransformer inverted the duties of the self-attention mechanism and the feed-forward network to achieve better performance.
- TimesNet [42]: TimesNet considers intraperiod and interperiod variations in 2D space for time series analysis.
- LSTM [43]: LSTM is widely used in drought prediction. We follow the same architecture as existing studies [47, 48, 49], with a convolution neural network as the feature extractor and an LSTM network to forecast drought indices.

To ensure comparable training times across different methods, we utilize a 3-layer encoder and 2-layer decoder for Transformer, PatchTST, and iTransformer, aligning with our approach. For TimesNet, we employ a 2-layer encoder and 1-layer decoder to maintain training durations similar to those of other baselines.

A.2 Related Work

A.2.1 AI in Climate

AI in climate science has received significant attention in recent years. These advancements have enabled researchers to enhance climate models, improve climate prediction accuracy, and gain deeper insights into the complex dynamics of the Earth's climate system. Currently, AI has been applied to predict El Niño-Southern Oscillation (ENSO) [50], Typhoon detection [51], climate modeling [52].

Traditional climate models rely on physical equations and numerical methods to simulate the climate system, which can be computationally intensive and limited by the resolution and accuracy of input data. In contrast, deep learning models excel in recognizing complex patterns in large datasets, offering a robust complementary approach to traditional methods. Especially, some deep learning models show strong performance in processing temporal features. For example, [51] successfully used a transformer model to predict typhoon trajectories without relying on reanalysis data. [53] proposed AirFormer for nationwide air quality prediction in China. [54] proposed a foundation model to forecast key climate variables. These examples underscore the growing efficacy and application of machine learning in climate science.

A.2.2 Drought Prediction

Drought Prediction is one of the important tasks in climate science. Traditional climate models for drought prediction, which rely on physical equations and historical data, often struggle with the chaotic nature of climate systems. For example, the current generation of Earth system models (ESMs) has large biases in predicting precipitation at a sub-seasonal scale and thus flash drought [55]. The Global Ensemble Forecast System based on process-based models, which holds the potential to implement operationally flash drought forecast guidance, also exhibits large prediction errors[56]. Thus, many studies have highlighted the effectiveness of data-driven models in predicting droughts and identifying their key indicators. With the ability to deal with multicollinearity and non-linear relations among predictive features, machine learning (ML) models were applied to predict flash

drought from weeks to months, measured by hydrological, meteorological, and agricultural metrics. These methods include support vector machines, random forests, decision trees, etc. For example, Adede et al. [57] predicts vegetation condition index using a simple ANN model. It is also applied for agricultural drought prediction using satellite images and climate indices Marj and Meijerink [58]. However, these simple models were built using traditional ML approaches and often require handcrafted feature engineering or cannot effectively learn the feature from the data, which cannot exploit complex intercorrelation among different features and usually has limited predictive power. To tackle this issue, recent studies have leveraged more advanced deep learning methods for flash drought predictions, which can learn hierarchical feature representations automatically from data and often outperform traditional ML methods [59, 60]. For example, deep neural networks have been applied in drought prediction [61, 62, 63]. Models specifically designed for time-series data, such as LSTM, have also been used for predicting natural drought index [64], and agricultural drought conditions [65]. Dikshit and Pradhan [66] further combine LSTM with the convolution neural network to predict the meteorological drought index in Eastern Australia and use SHapley Additive exPlanations to understand model outputs. Similar CNN-LSTM combined models are also used in Yu et al. [47], Danandeh Mehr et al. [48] and Khan and Maity [49], where Yu et al. [47] predicts vegetation Index, Danandeh Mehr et al. [48] consider meteorological drought forecasting andKhan and Maity [49] focus on predicting hydrological drought. Amanambu et al. [67] further adapts Transformer [37] to accurately forecast hydrological drought in the Apalachicola River.

Compared with other studies using limited features, we consider more extensive features including physical conditions, climate conditions, and vegetation dynamics based on the underlying mechanisms of drought development. In particular, we consider the interplay between physical drivers and vegetation dynamics and advance our understanding of how climate and vegetation features and their spatial-temporal interactions regulate droughts. Thus we could learn more comprehensive representation from both static and temporal data to improve the forecasting performance. In addition, our method forecasts three different types of drought through multi-task learning using the shared representation without the need for extra computations to train separate models.

A.2.3 Time-series Forecasting

Time series forecasting has been extensively studied across various domains, including climate science [53], traffic [68], and healthcare [69]. The complex and dynamic nature of time series data makes forecasting a challenging task. Depending on the forecasting length, time series tasks can be categorized into long-term and short-term forecasting. Additionally, based on data types, there are univariate, multivariate, and spatio-temporal forecasting. LSTM [43] has been widely used in many time-series forecasting tasks. Recently, due to the tremendous success of the Transformer in natural language processing and computer vision, it has also been widely adopted in time-series forecasting problems. Researchers have proposed many variants of Transformers, such as Informer [38], Autoformer [70], FEDformer [71], PatchTST [39], iTransformer [41]. Even though some transformers are proven not effective as linear-based methods like [40] in some tasks. The ability of Transformer to model global dependencies still makes them a popular choice for time series problems.

A.3 Addition Results

A.3.1 Variable Importance Comparison

In this section, we quantify the importance of each attribute to prediction from the perspective of prediction accuracy. We train models by excluding the predictive features one at a time and measure the change in model loss at the first epoch compared to a baseline model that uses all features. The differences shows in Figure 7 illustrate the contribution of each feature to the performance of the models. The results also prove that static features such as vegetation dynamics and climate conditions are valuable in spatiotemporal forecasting for drought prediction tasks.

A.3.2 Ablation Study on Static Features

In this section, we analyze the impact of each component within our model design. Initially, we assess the effectiveness of integrating static features into drought prediction tasks. To do this, we ablate the static features and compare *SPDrought* with the modified version (*SPDrought*(t)) that does not utilize static features. *SPDrought*(t) concatenates a full zero vector to the temporal representations before

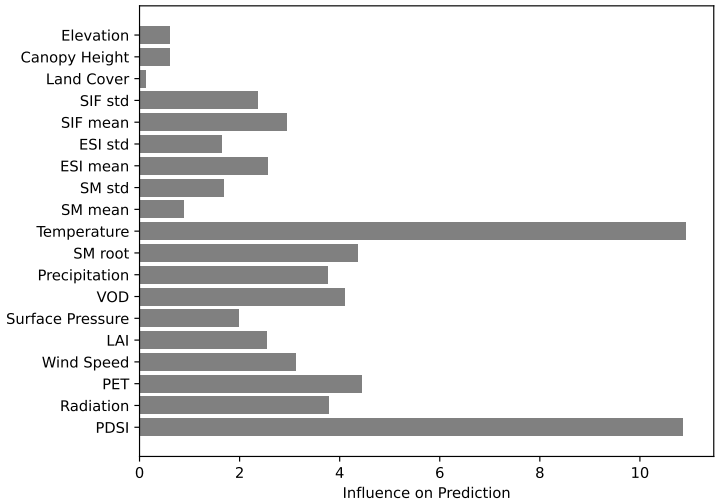


Figure 7: Relative Importance of Predictors

Table 4: Average MAE of the ablation study on multi-task learning

Model	Soil Moisture	Evaporative Stress Index	Solar-induced chlorophyll Fluorescence	Total
SPDrought(Single)	$19.39_{\pm 1.26}$	$3.35_{\pm 0.22}$	$10.50_{\pm 0.45}$	$33.24_{\pm 0.93}$
SPDrought(Multi)	$21.39_{\pm 0.14}$	$4.40_{\pm 0.02}$	$12.21_{\pm 0.23}$	$38.01_{\pm 0.35}$

the transformer decoder and is also trained for 30 epochs. Then, we also investigate the impact of the spatial-temporal feature fusion module on model performance. So, we also conduct an additional experiment where we remove this module (denoted as *SPDrought*(f)) and present results in Figure 8. The results show that combining static predictors in representation can improve the forecasting performance across all drought prediction tasks, and the spatial-temporal feature fusion module consistently improves the performance and yields more stable outcomes.

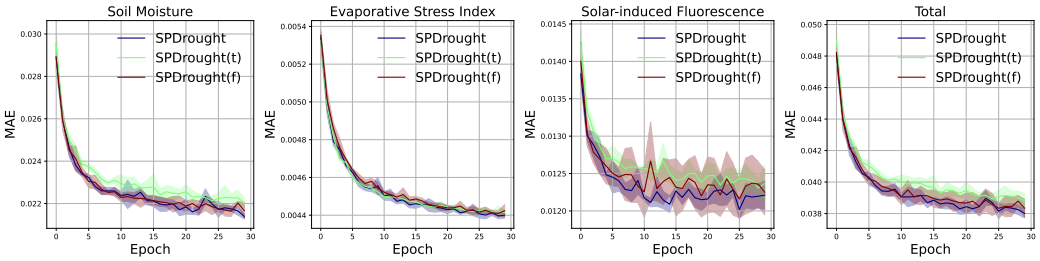


Figure 8: Ablation study on static features and spatial-temporal fusion module. We run the experiment three times and report the average MAE loss on the test set.

A.3.3 Ablation Study on Multi-task Training

Here, we compare multi-task training with single-task training and report the results in Table 4. The ablation study highlights the efficiency of the multi-task learning approach. While training the

Table 5: Ablation study on model parameters

MAE ($\times 10^{-3}$)	SPDrought	SPDrought(w/o Encoder)	SPDrought(w/o Decoder)	SPDrought(50)
Soil Moisture	$21.39_{\pm 0.14}$	$28.20_{\pm 0.17}$	$23.87_{\pm 0.33}$	$22.46_{\pm 0.14}$
Evaporative Stress Index	$4.40_{\pm 0.02}$	$5.51_{\pm 0.01}$	$4.72_{\pm 0.04}$	$4.51_{\pm 0.01}$
Solar-induced chlorophyll Fluorescence	$12.21_{\pm 0.23}$	$17.24_{\pm 0.58}$	$12.41_{\pm 0.13}$	$12.30_{\pm 0.15}$
Total	$38.01_{\pm 0.35}$	$50.95_{\pm 0.70}$	$41.01_{\pm 0.48}$	$39.27_{\pm 0.06}$

Table 6: Average MAE over three runs of experiments on drought indices over 26 weeks using the temporal split

MAE ($\times 10^{-3}$)	**SPDrought**	**Informer**	**PatchTST**	**DLinear**	**iTransformer**	**TimesNet**	**LSTM**
Soil Moisture	$47.77_{\pm 0.58}$	$51.97_{\pm 0.09}$	$48.27_{\pm 0.28}$	$48.66_{\pm 0.81}$	$49.55_{\pm 0.12}$	$48.48_{\pm 0.21}$	$53.61_{\pm 0.16}$
Evaporative Stress Index	$5.94_{\pm 0.02}$	$6.37_{\pm 0.06}$	$6.52_{\pm 0.04}$	$6.69_{\pm 0.10}$	$6.39_{\pm 0.03}$	$6.20_{\pm 0.03}$	$6.48_{\pm 0.04}$
Solar-induced chlorophyll Fluorescence	$22.26_{\pm 0.05}$	$25.91_{\pm 0.18}$	$29.16_{\pm 0.47}$	$29.28_{\pm 0.38}$	$25.28_{\pm 0.15}$	$25.27_{\pm 0.59}$	$26.04_{\pm 0.20}$
Total	$75.94_{\pm 0.65}$	$84.24_{\pm 0.19}$	$83.93_{\pm 0.78}$	$84.60_{\pm 1.23}$	$81.20_{\pm 0.13}$	$79.93_{\pm 0.81}$	$86.10_{\pm 0.23}$

SPDrought model on multiple tasks simultaneously (*SPDrought(Multi)*), it achieves comparable results to training on individual tasks (*SPDrought(Single)*) but in about one-third the time. This demonstrates that multi-task learning can significantly speed up the training process without a substantial drop in accuracy, making it a valuable strategy when time and computational resources are limited.

A.3.4 Ablation Study on Model Parameters

In this section, we explore the contribution of each component by removing the component from *SPDrought*. We conduct this study by creating variants of the model: *SPDrought* without the Transformer Encoder (*SPDrought(w/o Encoder)*), *SPDrought* without the Transformer Decoder (*SPDrought(w/o Decoder)*), and *SPDrought* with a reduced training window of 50 weeks, approximately one year (*SPDrought(50)*). These variants help us understand the role of each component in capturing temporal dependencies and learning drought patterns from historical data. As shown in Table 5, the Transformer encoder effectively helps to capture time dependence and thus significantly improves the overall performance. Furthermore, reducing the training window to 50 weeks (*SPDrought(50)*) slightly affects the model's accuracy. It suggests that SPDrought can effectively learn drought patterns even with limited historical data. However, extended historical data contributes to better model performance, highlighting the importance of a more comprehensive dataset for training.

A.3.5 Comparison on Temporal Splitting

In the previous comparison, we evaluate each method on test pixel regions. Here, we adopt a temporal split in the data, assessing the baseline methods and our approach for predicting drought indices over the next 26 weeks which are not seen during training in Table 6.

Table 7: Evaluation of drought prediction by soil moisture with standard deviations (Full version of Table 2).

	SPDrought	**Transformer**	**Informer**	**PatchTST**	**DLinear**	**iTransformer**	**TimesNet**	**LSTM**
Accuracy	$86.26_{\pm 0.11}$	$76.09_{\pm 0.11}$	$72.16_{\pm 0.34}$	$62.24_{\pm 4.63}$	$62.85_{\pm 0.01}$	$77.18_{\pm 0.04}$	$81.54_{\pm 0.91}$	$77.45_{\pm 0.26}$
Precision	$76.80_{\pm 0.17}$	$59.94_{\pm 0.19}$	$53.40_{\pm 0.57}$	$36.94_{\pm 7.70}$	$37.96_{\pm 0.02}$	$61.74_{\pm 0.07}$	$68.98_{\pm 1.51}$	$62.18_{\pm 0.43}$
p-value	-	1.49×10^{-4}	9.96×10^{-5}	1.17×10^{-2}	6.78×10^{-6}	1.68×10^{-5}	9.77×10^{-3}	5.57×10^{-4}

A.4 Data Source

We collect data from the following source:

- NLDAS [19]: NLDAS is provided by NASA, collected from `https://ldas.gsfc.nasa.gov/nldas`.
- SMAP [20]: SMAP is a public dataset provided by NASA, collected from `https://smap.jpl.nasa.gov/`.
- ALEXI [22]: ALEXI is provided under the U.S. Geological Survey (USGS), collected from `https://lpdaac.usgs.gov/products/eco4esialexiuv001/`.
- CSIF [23]: CSIF dataset is under CC BY 4.0, collected from `https://figshare.com/articles/dataset/CSIF/6387494`.
- ERA5 [25]: ERA5 is provided by the European Centre for Medium-Range Weather Forecasts under Copernicus license, data collected from `https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5`.

- SRTM [26]: NASA Shuttle Radar Topography Mission (SRTM) datasets are provided under the U.S. Geological Survey (USGS), collected from https://lpdaac.usgs.gov/products/srtmgl1v003/.
- VODCA [27]: VODCA is under CC BY 4.0.
- MODIS [28]: MODIS is provided by NASA, collected from https://modis.gsfc.nasa.gov/.
- GLAD [29]: Global Land Analysis & Discovery (CC BY), collected from https://glad.umd.edu/.
- NLCD [30]: Nation Land Cover Database is in the public domain, provided under USGS, collected from https://www.usgs.gov/centers/eros/science/national-land-cover-database.