atmosphere

MDPI

Article

Unravelling the Drought Variance Using Machine Learning Methods in Six Capital Cities of Australia

Wenjing Yang [1,2,3], Shahab Doulabian [4], Amirhossein Shadmehri Toosi [3] and Sina Alaghmand [5,*]

1 Department of Hydraulic Engineering, Tsinghua University, Beijing 100082, China; yangwj20@mails.tsinghua.edu.cn
2 China Institute of Water Resources and Hydropower Research (IWHR), Beijing 100038, China
3 College of Science and Engineering, Flinders University, Adelaide, SA 5042, Australia; amirh.shadmehr@flinders.edu.au
4 Department of Civil Engineering, Shahrood University of Technology, Shahrood 3619995161, Iran; doulabian@gmail.com
5 Department of Civil Engineering, Monash University, Melbourne, VIC 3800, Australia
* Correspondence: sina.alaghmand@monash.edu

Abstract: Understanding and projecting drought, especially in the face of climate change, is crucial for assessing its impending risks. However, the causes of drought are multifaceted. As the environmental research paradigm pivots towards machine learning (ML) for predictions, our investigation contrasted multiple ML techniques to simulate the Standardized Precipitation Evapotranspiration Index (SPEI) from 2009 to 2022, utilizing various potential evapotranspiration (PET) methods. Our primary focus was Australia, the world's driest inhabited continent. Given the challenges with ML model interpretation, SHAP (SHapley Additive exPlanations) values were employed to decipher SPEI variations and to gauge the relative importance of precipitation (Prec) and PET in six key Australian cities. Our findings revealed that while different PET methods resulted in distinct mean values, their trends remained consistent. Post the Millennium Drought, Australia experienced several drought events. SPEI discrepancies based on PET methods were minimal in humid regions like Brisbane and Darwin. However, for arid cities, the Priestley–Taylor equation-driven SPEI differed notably from other methods. Ridge regression was the most adept at mirroring SPEI changes among the assessed ML models. Furthermore, the SHAP explainer discerned that PET-related climate variables had a greater impact on SPEI in drier cities, whereas in humid cities, Prec was more influential. Notably, the research emphasised CO_2's role in influencing drought dynamics in humid cities. These insights are invaluable for enhancing drought mitigation strategies and refining predictive models. Such revelations are crucial for stakeholders aiming to improve drought prediction and management, especially in drought-prone regions like Australia.

Keywords: meteorological drought; SPEI; potential evapotranspiration; SHAP; machine learning

check for updates

Citation: Yang, W.; Doulabian, S.; Shadmehri Toosi, A.; Alaghmand, S. Unravelling the Drought Variance Using Machine Learning Methods in Six Capital Cities of Australia. *Atmosphere* **2024**, *15*, 43. https://doi.org/10.3390/atmos15010043

Academic Editors: Yulong Zhong, Shuang Zhu, Dongdong Kong and Peng Yang

Received: 6 December 2023
Revised: 26 December 2023
Accepted: 28 December 2023
Published: 29 December 2023

1. Introduction

Drought is a multi-faceted environmental challenge that has captivated the attention of hydrologists, environmental scientists, and policymakers worldwide in recent years. Among its various forms—including meteorological, hydrological, agricultural, and socioeconomic—the meteorological drought, marked by decreased rainfall, has emerged as a prominent concern [1,2]. All forms of drought are interrelated, leading to consequences ranging from agriculture to socio-economic issues [3]. These include water shortages, reduced crop outputs, ecosystem disruptions, and broader socioeconomic impacts [3,4].

Various drought indices have been promulgated to monitor and assess drought conditions [5,6]. The standardized precipitation index (SPI), introduced by McKee [7], is the universal meteorological drought index put forward by the World Meteorological Organization (WMO) due to its consistent representation of dry/wet states, ease of calculation and versatility across various timescales. Although the SPI has been shown to help detect different drought types that affect other systems and regions [8,9], it fails to detect drought conditions determined not by a lack of precipitation but by a higher-than-normal atmospheric evaporative demand, especially for studies of extreme heat waves and climate change related to global warming [10,11]. The Standard Precipitation Evapotranspiration Index (SPEI) stood out due to the consideration of water supply (precipitation, Prec) and water demand (potential evapotranspiration, PET) [12,13]. Li [14] compared the SPI and SPEI in China and found that the SPEI revealed an overall increase in drought severity, area, and frequency from 1998 to 2015. By contrast, SPI does not show this phenomenon since Prec does not exhibit a significant change overall. Nguvava [15] found that SPEI projected more intense and frequent droughts over East Africa, which is more robust than the SPI. Given PET increases in a warming climate, SPEI may be more suitable for monitoring droughts under climate change.

Machine learning (ML) methods have been increasingly employed for forecasting various environmental and climatic indices [2,16,17]. ML is progressively reshaping the forecasting landscape within drought research, demonstrating efficacy in predicting SPEI values [18,19]. However, ML is classified as a "black box" model, which is difficult to interpret. Hence, most studies have primarily used ML models for SPEI modelling and prediction while rarely addressing the mechanisms behind drought occurrence. The emerging SHapley Additive exPlanations (SHAP) method, proposed by Lundberg and Lee [20], is a unified framework used to understand the output of a model and the contribution of each feature to that output in ML, and to interpret the predictions of ML, which has been used in many fields [21,22]. SHAP is based on cooperative game theory, specifically the concept of Shapley values. It is a powerful tool for interpretability, allowing users to grasp not only the overall importance of features but also their individual impact on each prediction. This enhancement in understanding aids in demystifying and fostering trust in complex machine learning models. Therefore, SHAP could be helpful in explaining the reason for the change in SPEI.

Australia, the world's driest inhabited continent [23], faces unique vulnerability to drought [24,25], especially under climate change [26]. Recent droughts in Australia have had severe impacts [27]. The Australian "Black Summer" bushfires of 2019/2020 are a stark example of the devastating effects of prolonged drought conditions. These bushfires, which burned approximately 18 million hectares, represent the worst fire season in the recorded history of southeast Australia [28–30]. The ecological consequences of Australia's recent bushfires have been profound, especially since they followed extensive and broad-scale drought and landscape drying [27]. Additionally, eastern Australia has been experiencing severe drought in the lead-up to and during the current fire season. In particular, much of northeastern New South Wales (NSW) has had the lowest rainfall on record and above-average temperatures over the six months to 30 November 2019 [31]. Therefore, accurate forecasting and mitigating drought are critical in Australia's ecological degradation and risk management.

Despite the extensive usage of SPEI, a consensus has not yet been reached regarding which equation should be used to estimate PET [32]. Numerous studies have compared the effects of different PET models on SPEI in both arid and humid regions, and they analysed the differences in SPEI for future and historical data [32]. These studies employed models such as Penman [33], Abtew [34], Hargreaves [35], Jensen and Haise [36], FAO56 Penman–Monteith [37], Thornthwaite [38], and Priestley and Taylor [39]. However, few studies use the PET model to consider the CO_2 effect developed by Yang [40]. Elevated CO_2 levels can result in decreased stomatal conductance, which in turn reduces evapotranspiration in regions that are not water limited.

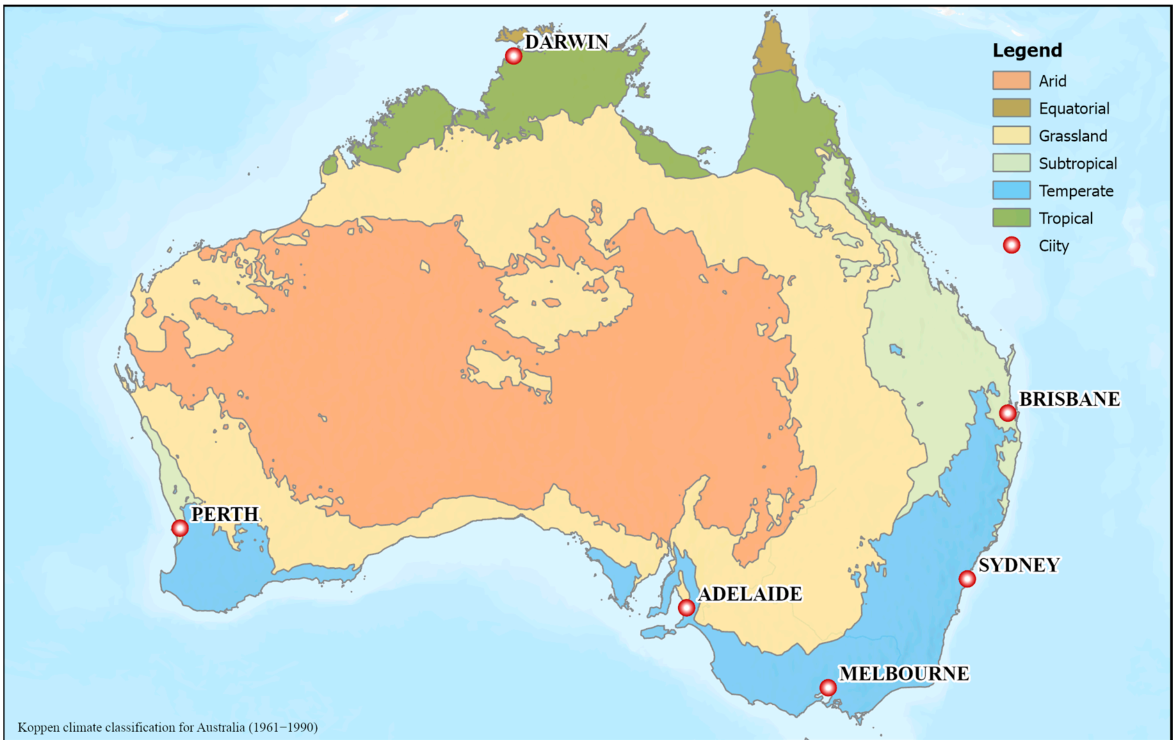
Recent research has utilised ML techniques to simulate SPI [41] and SPEI [42] based on the SPI/SPEI from the preceding month. However, they rely on the prior month's SPI/SPEI. In this study, meteorological variables during the previous 12 months were input in ML methods to predict the following month's SPEI. This approach facilitates a deeper understanding of how microclimatic factors influence meteorological drought.

We used the six capitals in Australia as a case study. This research aims to (1) compare different PET methods for calculating SPEI; (2) evaluate different ML methods for simulating SPEI using meteorological variables; (3) explain the reason for SPEI changing, which is caused by Prec or meteorological variables related to PET in primary cities in Australia. By bridging modern ML strategies with SHAP, this research offers valuable hydrological insights that enhance our understanding of water systems.

2. Data and Methods

2.1. Study Area

Australia is a continent marked by its susceptibility to extreme climatic events like drought and flooding [23]. Australia's unique geographical and climatic characteristics make it a focal point for studying drought, with profound implications for its ecosystems, agriculture, and urban centres [24,25]. The significance of drought in Australia is underscored by its enduring struggle with the consequences of climate change, a factor that intensifies the region's susceptibility to drought [26]. This study focuses on six major cities in Australia: Adelaide, Brisbane, Darwin, Melbourne, Perth, and Sydney. Figure 1 illustrates the locations of the selected meteorological stations in each major city, along with the mean annual rainfall and temperature data. It also delineates the climate zones where these stations are situated, showcasing the diversity of climate conditions under investigation.



Station	Latitude	Longitude	Elevation (m)	Mean Annual Rainfall (mm)	Mean Annual Max temperature (°C)	Mean Annual Min temperature (°C)
ADELAIDE AIRPORT	34.95	138.52	2	441.3	21.6	11.5
BRISBANE AERO	27.39	153.13	5	1015.1	25.4	15.8
DARWIN AIRPORT	12.42	130.89	30	1725.6	32.1	23.2
MELBOURNE AIRPORT	37.67	144.83	113	539.2	19.9	9.6
PERTH AIRPORT	31.93	115.98	15	760.7	24.6	12.2
SYDNEY AIRPORT AMO	33.95	151.17	6	1092.2	22.4	13.6

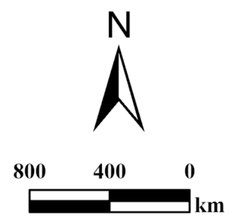


Figure 1. Map showing the six stations used in this study.

Adelaide, situated on the southern coast, experiences a Mediterranean climate characterised by its mild, wet winters and hot, dry summers [43]. Further north on the eastern coast is Brisbane, which lies in Queensland with a subtropical climate resulting in warm, humid conditions. Darwin, at the top end of Australia, serves as the capital of the Northern Territory. With its tropical savanna climate, it is marked by a pronounced wet season followed by a dry season. To the southeast, Melbourne has a temperate oceanic climate, which leads to variable weather patterns, with rainfall moderately distributed throughout the year [44]. On the opposite coast, Perth lies bordered by the Indian Ocean. Its semi-arid conditions manifest in hot, dry summers and mild, wet winters. Lastly, on the southeastern coast, Sydney enjoys a subtropical climate characterised by warm summers, mild winters, and consistent rainfall [45]. These cities represent diverse regions across the continent and compress a range of climate conditions, which can offer a lens to examine the impacts and challenges of drought [46].

2.2. Flowchart

Figure 2 shows the flowchart of SPEI simulation and SHAP. Details can be found in the text from Sections 2.3–2.7.

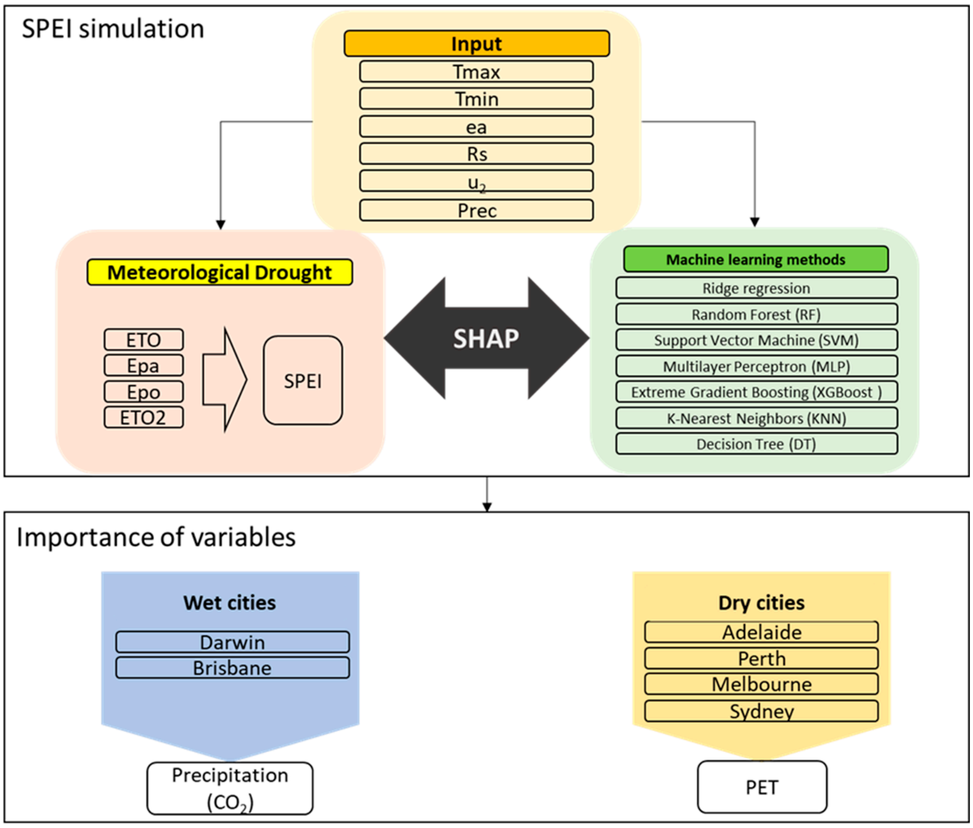


Figure 2. Flowchart of Standardized Precipitation Evapotranspiration Index (SPEI) simulation and the SHapley Additive exPlanations (SHAP) methodology. The inputs are maximum and minimum air temperatures (Tmax, Tmin), actual vapor pressure (ea), solar radiation (Rs), 2 m wind speed (u2), and precipitation (Prec). SPEI is calculated using four models: FAO56 Penman–Monteith Reference Crop Model (ETO), Penman Open-Water Model (Epa), Priestley–Taylor Evaporation (Epo), and the FAO56-CO2 Model (ETO2). SHAP method was used for determining the importance of variables in wet and dry cities.

2.3. Data

The primary datasets utilised in this research were collected from the Bureau of Meteorology (http://www.bom.gov.au/; accessed on 20 December 2023), including maximum and minimum air temperatures (Tmax, Tmin), the maximum and minimum relative humidity (RHmax, RHmin), precipitation (Prec), solar radiation (Rs), and wind speed at the height of 10 m (u_{10}). Actual vapor pressure (ea) was derived from RHmax and RHmin. All variables were averaged to a monthly scale. Monthly Atmospheric CO_2 Concentration was sourced from the records maintained at the Mauna Loa Observatory, Global Monitoring Laboratory (GML) (https://gml.noaa.gov/ccgg/trends/; accessed on 20 December 2023).

2.4. PET Calculation

Potential evapotranspiration (PET) is a critical parameter to assess in drought research, indicating the potential water loss from the soil surface due to evaporation and plant transpiration [47]. In this study, various methods are employed to calculate PET (Table 1).

Table 1. Inputs for Potential Evapotranspiration (PET) Model.

PET Method	Abbr	Input	Reference
Penman Open-Water Model	Epa	Tmax, Tmin, ea, Rs, u_2	(Penman, 1948) [33]
Priestley–Taylor Evaporation	Epo	Tmax, Tmin, ea, Rs	(Priestley & Taylor, 1972) [39]
FAO56 Penman–Monteith Reference Crop Model	ETO	Tmax, Tmin, ea, Rs, u_2	(R. Allen et al., 1998) [37]
FAO56-CO_2 Model	ETO2	Tmax, Tmin, ea, Rs, u_2, CO_2	(Yang et al., 2019) [40]

The Penman Open-Water Model (E_{pa}), introduced by Penman [33], is a simplified adaptation of the Penman–Monteith model. This model estimates evaporation for open-water surfaces, denoted as E_{pa} (mm d^{-1}), under the assumption of surface resistance (r_s) = 0.

$$E_{pa} = \frac{\Delta R_n + \frac{\rho C_p D}{r_a}}{\Delta + \gamma\left(1 + \frac{r_s}{r_a}\right)} \tag{1}$$

where Δ is the slope of the saturation vapour pressure curve at the air temperature (kPa °C^{-1}); γ is the psychrometric constant (kPa °C^{-1}); ρ is the air density (kg m^{-3}); C_p is the specific heat of the air (J kg^{-1} K^{-1}); D is the water vapour deficit (kPa); R_n is the net radiation (mm d^{-1}) calculated from Allen [37]; albedo was set as default value 0.23. The aerodynamic conductance, r_a (m s^{-1}), in this model is determined using the Rome wind function [48] as:

$$\frac{1}{r_\text{a}} = \frac{2.6(1 + 0.54u_2)patm}{0.622\rho * 3600 * 24} \tag{2}$$

Here, $patm$ represents the air pressure (kPa); u_2 is the wind speed at 2 m (m s^{-1}) derived from u_{10} using the relation: $u_2 = u_{10}\ (2/10)^{1/7}$ [49].

Priestley–Taylor Evaporation (E_{po}), developed by Priestley and Taylor [39], is a method which offers an empirical approach to ascertain PET. Its magnitude is mainly impacted by the net radiation with sufficient water supply because the atmospheric conditions are adapted to saturated surfaces. The formula is as follows:

$$E_{po} = \alpha \frac{\Delta R_n}{\Delta + \gamma} \tag{3}$$

where R_n is the net radiation (mm d^{-1}); α is the Priestley–Taylor coefficient, it is often taken as 1.26 for water surfaces but may vary depending on the surface and conditions being evaluated.

The FAO56 Penman–Monteith Reference Crop Model (ETO): This model is a standardised version of the Penman–Monteith model. It is designed for an idealised reference crop surface with specific parameters, such as r_s = 70 s m^{-1}, a vegetation height of 0.12 m, and a surface albedo of 0.23 [37]. Rn is net radiation at the crop surface in the unit of MJ m^{-2} d^{-1}. The model calculates *ETO* (mm d^{-1}) using the following equation:

$$ETO = \frac{0.408\Delta R_n + \gamma\frac{900}{\mathrm{Ta}+273}u_2 D}{\Delta + \gamma(1 + 0.34u)} \tag{4}$$

The FAO56-CO$_2$ Model (*ETO*2): Proposed by [40], this modification incorporates the influence of changing atmospheric CO$_2$ concentrations on the FAO-56 reference crop formula. The unique term ((2.4 × 10^{-4}([CO_2] − 300)) captures the effect of varying CO_2 levels on r_s.

$$ETO2 = \frac{0.408\Delta R_n + \gamma\frac{900}{\mathrm{Ta}+273}u_2 D}{\Delta + \gamma\{1 + u[0.34 + 2.4 \times 10^{-4}([CO_2] - 300)]\}} \tag{5}$$

These approaches offer a comprehensive evaluation of PET, capturing details from aerodynamic radiative balances to atmospheric CO_2 fluctuations.

2.5. SPEI Calculation

The SPEI uses the monthly difference between precipitation and PET, which is a simple water balance methodology that is calculated at different times [12]. P-PET is modified using a 3-parameter log–logistic distribution, subsequently transformed into standard deviations for mean values. The SPEI aims to depict deviations in climatological drought by considering the equilibrium between water availability and atmospheric water demand. This study calculates SPEI at 12 months to analyse and monitor long-term drought events. A drought event is identified when the SPEI value is negative [50].

2.6. Machine Learning Methods

In environmental science, a selection of ML algorithms has surged, primarily due to their adeptness at discerning and forecasting complex patterns in voluminous datasets. We applied several ML models using the PyCaret package [51] in Python 3.8 to predict SPEI. The first method is the ridge regression. It is a linear regression technique used in ML and statistics and is an extension of ordinary least squares (OLS) regression. Ridge regression is used to mitigate the problem of multicollinearity and overfitting in linear regression models. Next, support vector machine regression (SVM) was employed, a progression of support vector machines designed for regression tasks. Using the core principles intrinsic to the SVM classification approach, it aims to discern a line or, in multidimensional scenarios, a hyperplane that optimally segregates the data. The random forest (RF) approach was also harnessed, typifying an ensemble learning methodology. Here, a 'forest' comprising numerous decision trees is assembled, and the final prediction is derived by selecting the mode of the individual trees' outputs.

Furthermore, the K-Nearest Neighbors (KNN) approach, a non-parametric technique, was employed for both regression and classification tasks. It derives its input from the k closest training exemplars in the feature space. Our study also included Extreme Gradient Boosting (XGB), renowned for its proficient implementation of gradient boosting designed to maximise speed and model efficiency. In addition, Multi-layer Perceptron (MLP) is a type of artificial neural network (ANN) used in ML and deep learning. It is a feedforward neural network that consists of multiple layers of interconnected nodes or neurons. MLPs are known for their ability to model complex relationships and solve various problems. Lastly, decision tree (DT) is a non-linear and non-parametric algorithm that makes decisions based on a tree-like graph structure. Each internal node of the tree represents a decision or a test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label or a regression value.

The dataset was split into training and testing sets. In total, 70% of the data will be used for training, and the remaining 70% will be used for testing. A ten-time-fold cross-validation approach was used to assess model stability using a training dataset. The advantage of this cross-validation approach is that it provides a more reliable estimate of model performance and reduces the impact of randomness in the performance evaluation. Each data point is used for both validation and training, and the model is validated multiple times on different subsets of data, helping to gain a more thorough understanding of its performance. We employed three key metrics to assess our model's performance: Mean Absolute Error (MAE), Root Mean Square Error (RMSE/RMSD), and the Coefficient of Determination (R^2).

2.7. Machine Learning Exploratory Analysis

SHAP is a renowned Python library designed to elucidate the outputs of ML models [20]. It offers a consolidated framework for assessing feature importance in prediction-making, thereby enhancing the interpretability and trustworthiness of model determinations. This study evaluated multiple ML models for their efficacy in simulating the SPEI over 12 months. These models utilised inputs of climatic variables (Tmax, Tmin, Prec, ea, Rs, u_2) and CO_2 during the previous 12-month period. After model evaluation, the optimal model for each site was identified. Subsequently, we integrated the chosen models with the SHAP library to comprehensively explain the ML model outputs.

3. Results and Discussion

3.1. SPEI Change Based on Different PET Methods

Figure 3 illustrates the variation in PET and precipitation across each site using different PET methods, highlighting that while the precise PET values varied depending on the method utilised, a consistent trend was observed across these methodologies. Notably, among the PET models, Epa consistently recorded higher values across all regions than other models because it assumes water is continually available, and rs = 0. On the other hand, the value of Epo was lower than the other models in dry areas (except notably in Darwin and Brisbane) because Epo ignores horizontal heat advection, which is of importance in dry regions. Despite the variations in absolute values, all models exhibited analogous trends. This unity in trend patterns suggests that although there might be discrepancies in PET calculations across different methods, the overarching climatic patterns remain predominantly consistent.

Moreover, there are considerable differences in annual rainfall patterns, particularly when contrasting cities such as Darwin and Adelaide. During specific observed periods (2009–2022), Darwin recorded an impressive annual precipitation rate, nearing 2000 mm, while Adelaide's precipitation is notably limited at around 400 mm. Darwin's rainfall figures fluctuated between 1070 mm in 2019 and a maximum of 2725 mm in 2011, averaging 1827 mm annually for the studied period. Moreover, Sydney recorded its minimum annual rainfall of 697 mm in 2019, and it rose significantly to 2214 mm in 2022, with a 13-year average standing at 1106 mm. Overall, the data underscore Australia's vast disparities in precipitation patterns, which have profound implications for the nation's regional hydrological and ecological dynamics.

Figure 4 displays the SPEI for 12-month periods derived from different PET models for each location. The shaded regions between the grey vertical dots denote drought periods at each location. The start and end of these droughts are easily identifiable. Each drought's duration is distinctly represented by colours corresponding to each PET model. Australia experienced a severe long-term drought period from 1997 to 2009, known as the Millennium Drought [52]. As shown in Figure 4, except for Darwin, all other cities have negative SPEI values from 2009 to 2010. They recovered for a few years, and then various droughts impacted Australia. In Adelaide, there was a prolonged drought from 2017 to 2019. Brisbane, Sydney, and Melbourne also experienced the same drought period. This drought event matches real-world observations. Eastern Australia has been experiencing severe drought, with the historically lowest precipitation and elevated temperatures over six months in 2019 [31]. On the other hand, Darwin endured a prolonged drought from 2018 to 2020. Perth faced several short and less severe drought periods after the Millennium Drought.

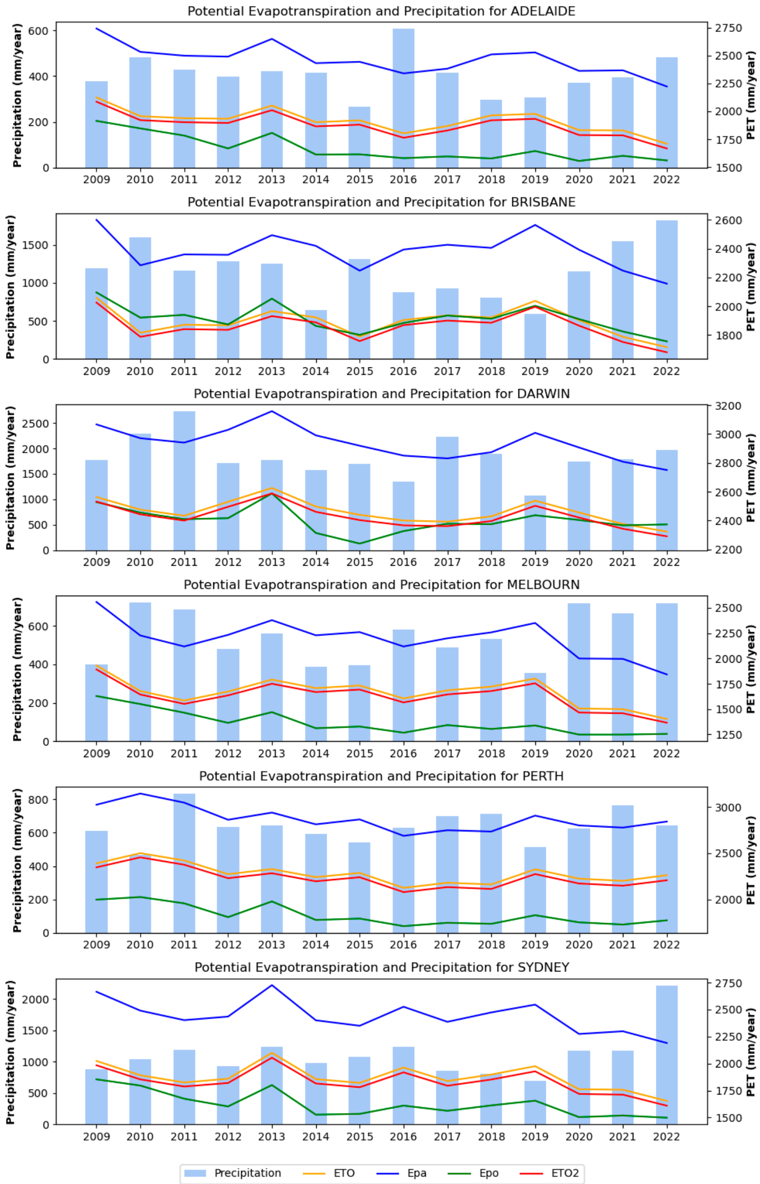


Figure 3. Variation in PET and precipitation at each site.

In both Brisbane and Darwin, using different PET methods reveals minimal distinctions in SPEI. However, a more nuanced pattern emerges when we examine Adelaide, Melbourne, and Sydney, with distinctions in SPEI. SPEI_Epo tends to predict a higher value for drought conditions around 2010 but a lower value around 2018. This disparity can be attributed to the absence of the aerodynamics term (advection) in the Epo equation, originally formulated by Priestley and Taylor [39]. This omission significantly impacts PET calculations in arid regions.

3.2. SPEI Simulation by ML Models

Figure 5 provides a visual comparison of various ML models based on their ability to understand the time series changes in SPEI over 12-month intervals. The figure uses box plots for each model to represent the distribution of metrics during ten-time-fold cross-validation. A notable feature of these box plots is the width of the bars, which signifies the stability of the model during cross-validation. A wider box indicates lower stability, meaning the model's performance varied more extensively across different validation sets. Conversely, a narrower box denotes higher stability and consistent performance across validation sets.

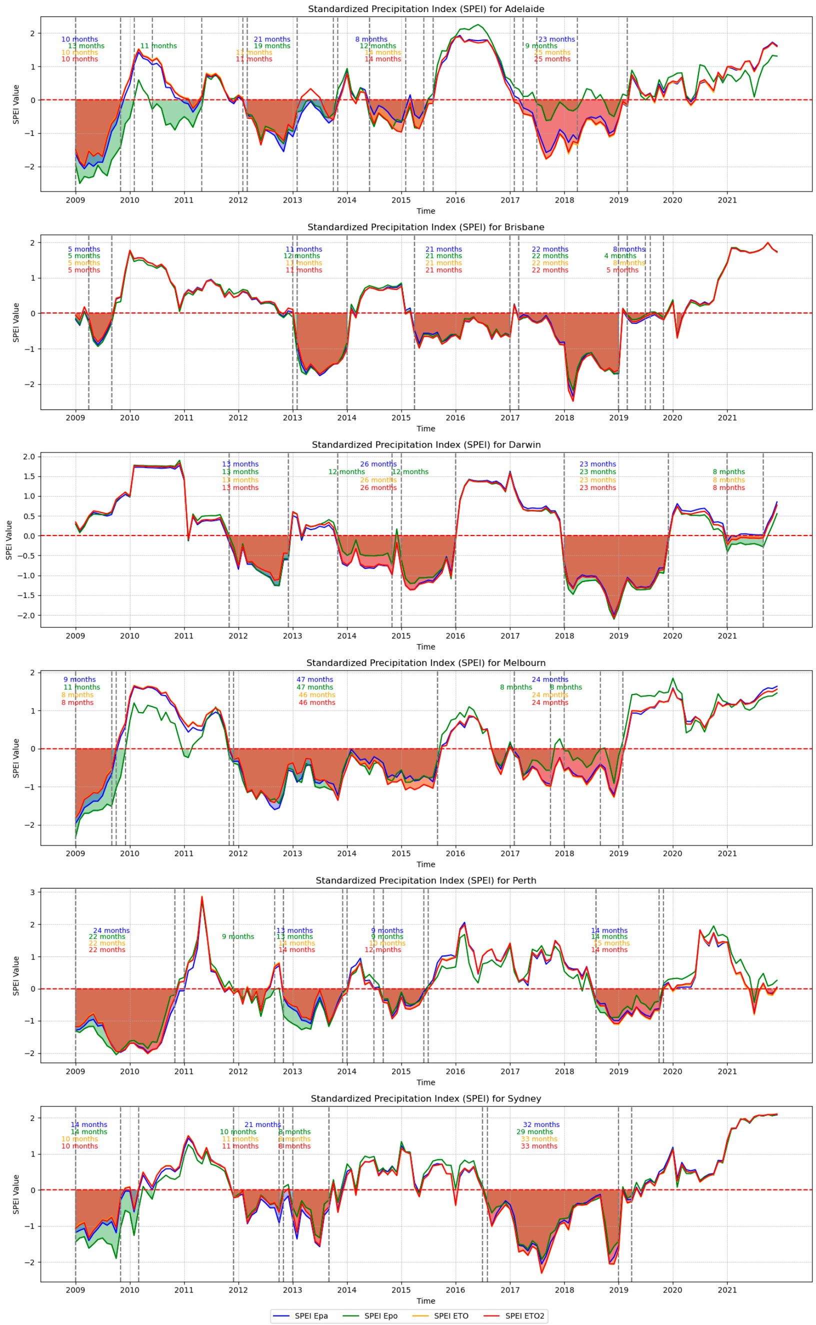


Figure 4. SPEI_12 derived from different PET models at each site. Grey vertical dots indicate the start and end of drought periods. The duration of each period is represented by text with the corresponding colour for each model.

From Figure 5, the performance of the ridge regression model stands out as markedly better, displaying both higher accuracy and greater stability than the other models. This is evident from its narrower box width. The MLP is the second-best performance model, followed closely by the SVM. Model training and model testing results are consistent (Figure 6). Figure 6 presents a Taylor diagram, a graphical method used to assess the similarity between outputs from different models in terms of their statistical moments. These moments include the standard deviation, correlation coefficient, and RMSD. The ridge model demonstrates the best simulation across all sites. The diminished performance of other ML models could potentially be attributed to the limited length of the dataset used, as these models often require larger datasets to fine-tune their parameters more effectively.

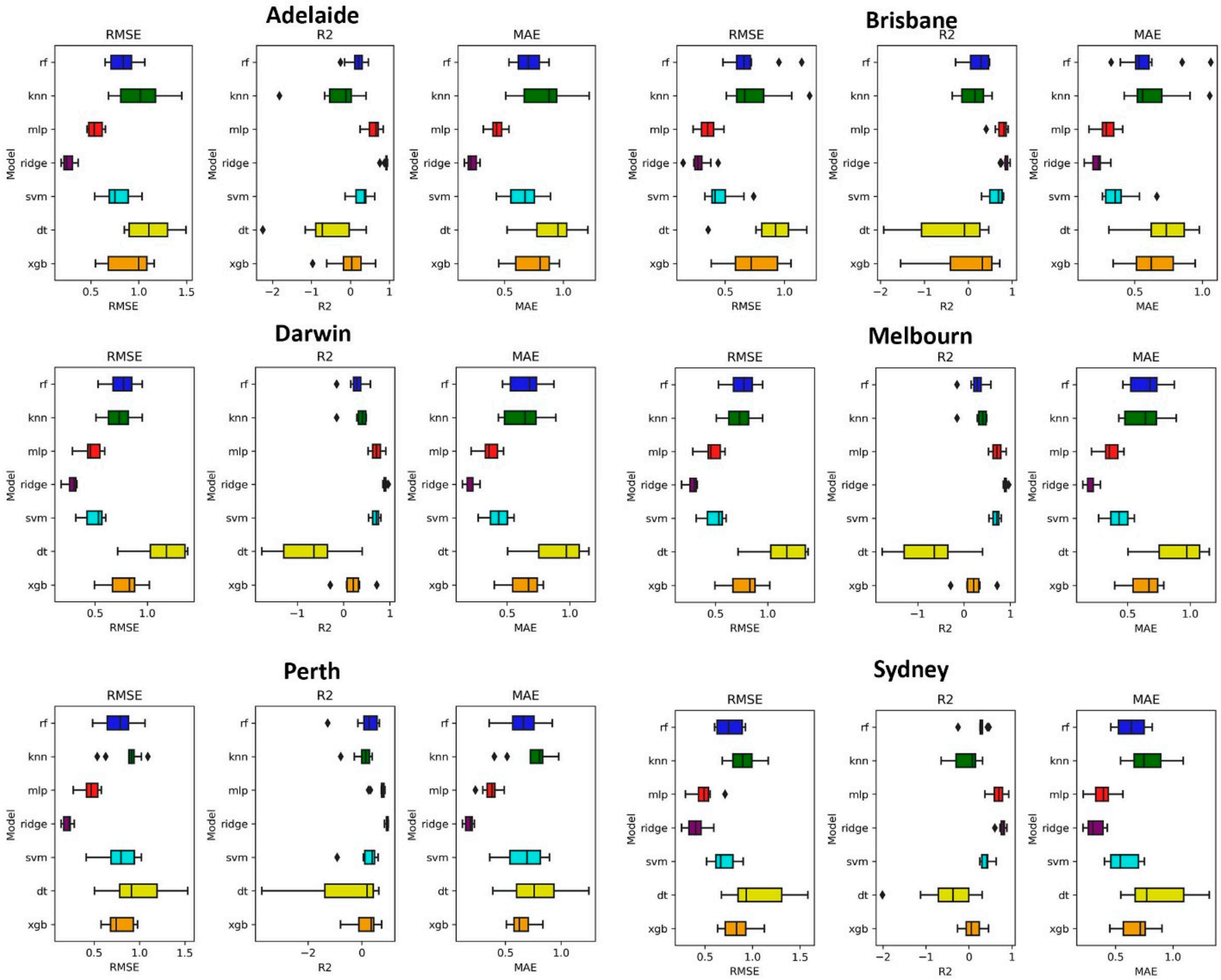


Figure 5. Model comparison of ten-time-fold cross-validation for train data at each site. Train dataset is 70% of the whole data.

3.3. Drivers of SPEI

Using SHAP values, we assessed the contribution of each feature to individual predictions within the best ML model (ridge). SHAP values illustrate how far each feature shifts the model's prediction from the average prediction. Figure 7 shows the top 20 most important variables influencing the model for each site. For instance, in Sydney, a higher Rs value (indicated by the pink colour) correlates with a negative impact on the model's output (a SHAP value of around −3). The number following the variable in the y-axis direction signifies its temporal relevance; "Rs.1" represents the last month's Rs value; "Rs.8" represents the Rs value from eight months prior.

Interestingly, only in Darwin is precipitation the dominant factor. Given Darwin's humid climate with an annual rainfall exceeding 1700 mm, the SPEI shows high sensitivity to fluctuations in precipitation. In Brisbane, an increase in rainfall leads to a positive influence on the model's output. In contrast, at other locations, variables associated with PET hold greater significance than precipitation. Furthermore, CO_2's impact on SPEI reveals a relatively higher influence in wet cities (Darwin and Brisbane) than in dry cities. The SPEI shows different sensitivity to Prec and PET based on the climatology (Figure 1). In dry cities (Adelaide, Perth, Melbourne, Sydney), the variation in PET exerts a more significant influence on SPEI than changes in precipitation. However, in wet cities (Darwin, Brisbane), variations in rainfall play a pivotal role in shaping SPEI. This observation aligns with findings from a prior study [53].

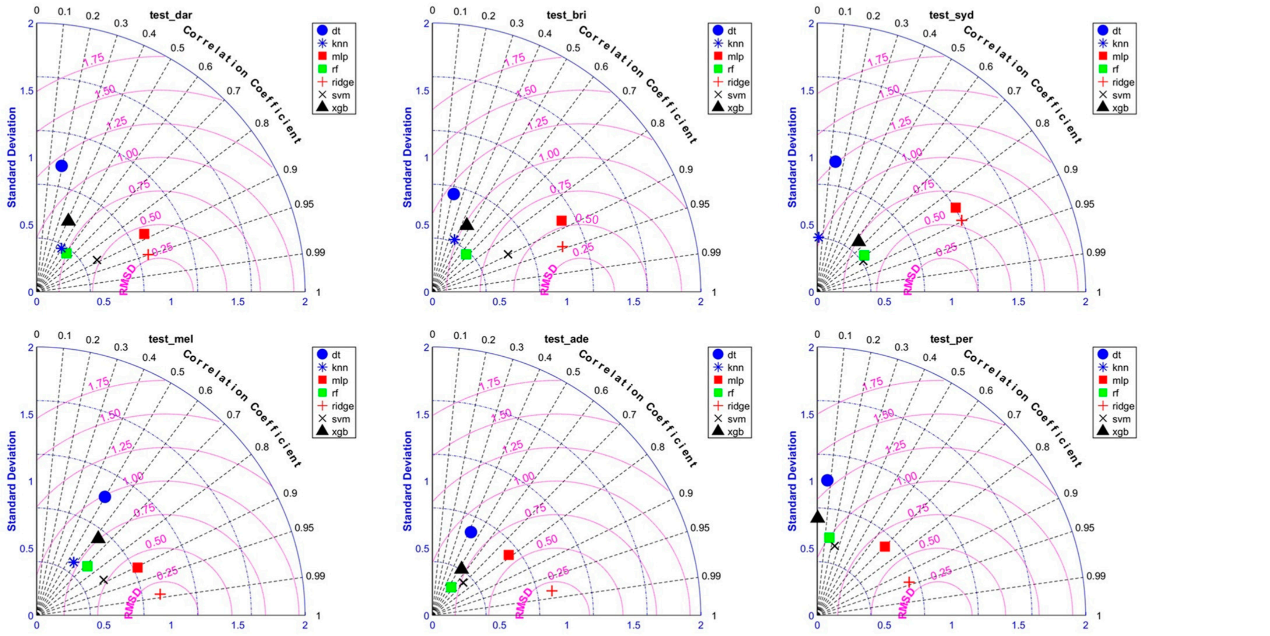


Figure 6. Model comparison for test dataset at each site. Test dataset is 30% of the whole data. RMSD is Root Mean Square Error.

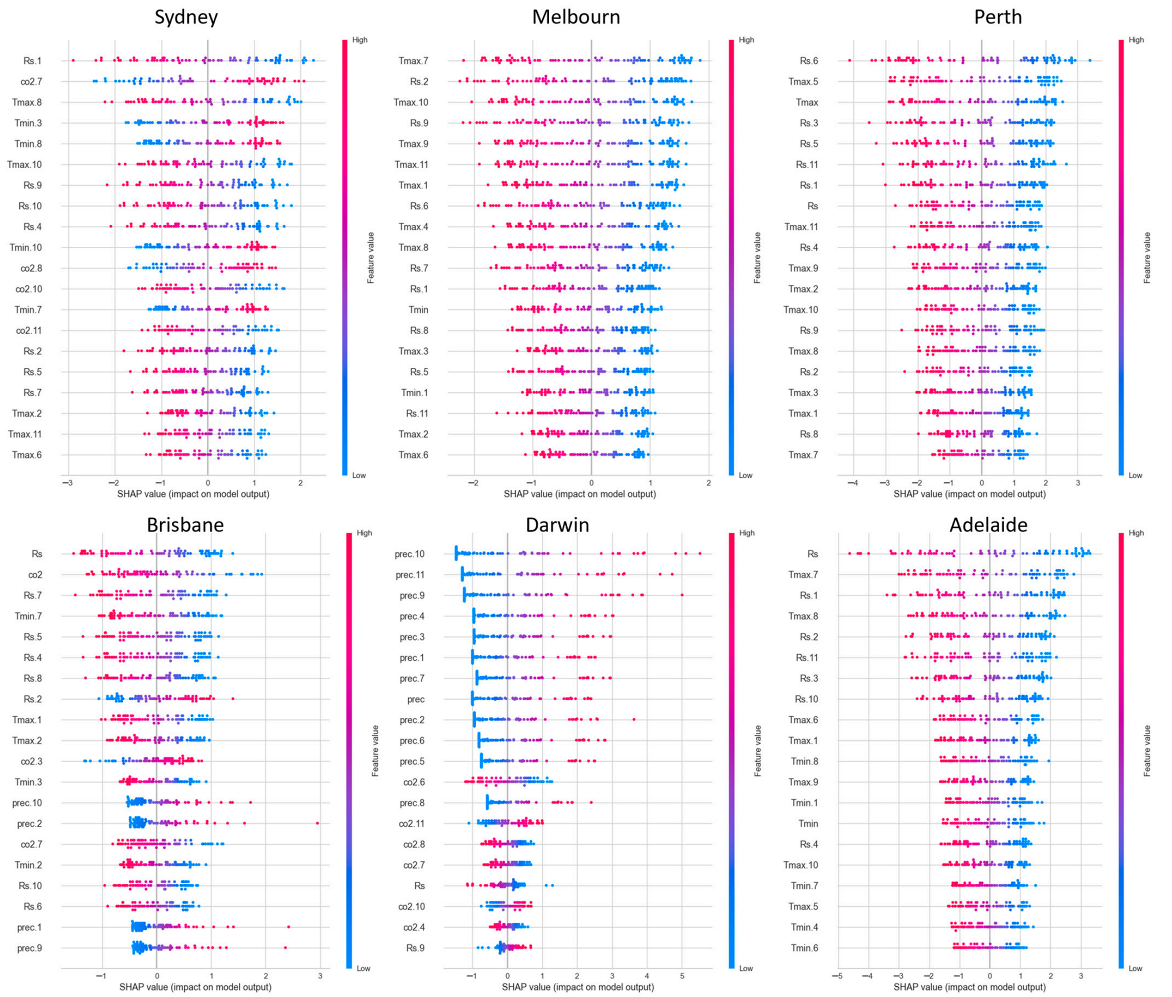


Figure 7. The impact of various features on model predictions.

4. Discussion

ML has rapidly gained prominence in climate research due to its ability to deliver accurate predictions, outperforming traditional models and statistical techniques in capturing complex interrelations between various climatic factors. Deo and Şahin [54] and Dikshit [55] have showcased the effectiveness of artificial neural networks and random forest models in forecasting the SPEI in Australia. However, they rely on the previous SPEI for forecasting the following month's SPEI. Our research corroborates the robustness of ML models in interpreting and predicting SPEI, notably producing consistent results even without incorporating the previous month's SPEI data. An important strength of this method is its ability to discern the time-lagged impacts of climatic variables on SPEI. For instance, in Darwin, SPEI is influenced by precipitation over a 12-month period (Figure 7), whereas in Sydney, Melbourne, Brisbane, and Adelaide, Rs shows no significant time lag effect, and Tmax or Tmin exhibits a lag of 5–8 months. This insight is particularly valuable for regions where temperature and precipitation are not synchronized, allowing for a more reasonable interpretation of climate lag on drought conditions.

Different PET methods can yield varying SPEI values for one site. SPEI calculated using the Epo equation (SPEI_Epo) often predicts a discrepant value from other PET methods, likely due to the exclusion of the aerodynamic term (advection) in the Epo equation, which significantly influences PET calculations in arid regions.

Concerning the impact of atmospheric moisture demand, which is lower in humid climates, it is presumed that on an annual basis, relative humidity remains relatively stable. This stability suggests that humid regions (Darwin) are less sensitive to variations in PET. Our findings align with other studies [53,56,57], highlighting that PET's influence can surpass that of precipitation in dry cities.

The ETO2 model stands out due to its ability to consider the changing CO_2, a crucial element given the increasing global CO_2 levels and their profound impact on plant physiology and water use [40]. By doing so, the ETO2 model provides a physiologically accurate, climate-contextual, and contemporary perspective on potential evapotranspiration, making it particularly valuable for analyses of drought assessment under a changing climate. The CO_2 influence is on reducing stomatal openings with higher CO_2 values (for the purpose of photosynthesis) [58], and thus limiting evaporative loss. Presumably it is less effective when the temperature is warm (i.e., less rainfall), for then a reduction in evaporative loss can cause the plant to overheat. In humid cities, there would be less evaporative demand, and so the plant can cut down evaporative loss without it having a thermal consequence [58].

5. Limitations

SPEI is characterized by its multi-time scale nature, and its applications vary with different time scales. Prior research indicates that SPEI9 and SPEI12 are typically used to represent the annual trend of regional drought [59]. However, SPEI's response to climatic variables may differ across various time scales. Our results, focusing on SPEI12, reflect long-term meteorological drought trends. Moreover, SPEI might not adequately capture local variations in drought conditions due to its reliance on regional climatic data, and it does not directly incorporate soil moisture or groundwater levels and vegetation conditions, which are vital in fully understanding meteorological drought impacts.

Additionally, we did not consider the effect of vegetation change on SPEI. Increased forest cover or native forest restoration can restore deep soil moisture that sustains base flow during dry periods [60,61]. As a result, it can act as a buffer and reduce hydrological drought or agriculture drought, increasing water ecosystem services. However, other types of droughts are not the aim of this paper. This can be investigated in further research.

6. Conclusions

Following the Millennium Drought, Australia encountered numerous drought events. Given the importance of drought in Australia in this study, we assessed the Standardised Precipitation Evapotranspiration Index (SPEI), a meteorological drought index, using various potential evapotranspiration (PET) models. Our study evaluated the accuracy of multiple ML techniques in SPEI predictions in six major Australian cities from 2009 to 2022. The PET values computed by distinct methods (Penman Open-Water Model, Priestley–Taylor Evaporation, FAO56 Penman–Monteith Reference Crop Model, and FAO56-CO_2 Model) differed in mean value, while their fluctuations remained similar. Notably, the SPEI determined using different PET methods varied minimally in wet areas (Brisbane and Darwin).

In contrast, due to the omission of advection effects in the Priestley–Taylor equation, SPEI calculated by Epo showed visible discrepancies with other methods in drier cities such as Adelaide, Perth, Melbourne, and Sydney. Among the ML model evaluation, the ridge model excelled in simulating SPEI_CO_2. By integrating the ridge model with the SHAP explainer, we discerned the crucial variables influencing SPEI at each site. In drier cities, climate variables tied to PET significantly impacted SPEI, whereas Prec was more pivotal in wetter regions. Intriguingly, CO_2′s influence on SPEI was more pronounced in humid cities than in their drier counterparts. Future research should investigate the CO_2 effect on drought dynamics under diverse scenarios. This research underscores the potency of ML models in predicting and demystifying the intricate dynamics of meteorological drought. We emphasise understanding the mode's operational context and the importance of continuous validation, especially in geographically diverse areas.

Author Contributions: Conceptualization, W.Y., A.S.T. and S.D.; methodology, W.Y. and S.D.; software, W.Y. and S.D.; validation, W.Y., S.D. and S.A.; formal analysis, W.Y. and S.D.; investigation, W.Y. and S.D.; resources, W.Y. and A.S.T.; data curation, W.Y. and S.D.; writing—original draft preparation, A.S.T., S.D. and W.Y.; writing—review and editing, A.S.T., W.Y., S.D. and S.A.; visualization, A.S.T. and W.Y.; supervision, S.A.; project administration, S.A.; funding acquisition, S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available through the internet. Station data: http://www.bom.gov.au/ (accessed on 20 December 2023); CO_2 Concentration from GML: https://gml.noaa.gov/ccgg/trends/ (accessed on 20 December 2023).

Acknowledgments: We thank the Bureau of Meteorology (BoM) for their invaluable data, greatly enriching our study. Their dedication to providing high-quality datasets has been instrumental to our research. We appreciate their continued commitment to the scientific community.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dumitraşcu, M.; Mocanu, I.; Mitrică, B.; Dragotă, C.; Grigorescu, I.; Dumitrică, C. The assessment of socioeconomic vulnerability to drought in Southern Romania (Oltenia Plain). *Int. J. Disaster Risk Reduct.* **2018**, *27*, 142–154. [CrossRef]
2. Swain, S.; Mishra, S.K.; Pandey, A.; Kalura, P. Inclusion of groundwater and socioeconomic factors for assessing comprehensive drought vulnerability over Narmada River Basin, India: A geospatial approach. *Appl. Water Sci.* **2022**, *12*, 14. [CrossRef]
3. Vicente-Serrano, S.M.; Quiring, S.M.; Peña-Gallardo, M.; Yuan, S.; Domínguez-Castro, F. A review of environmental droughts: Increased risk under global warming? *Earth-Sci. Rev.* **2020**, *201*, 102953. [CrossRef]
4. Van Loon, A.F.; Gleeson, T.; Clark, J.; Van Dijk, A.I.; Stahl, K.; Hannaford, J.; Di Baldassarre, G.; Teuling, A.J.; Tallaksen, L.M.; Uijlenhoet, R. Drought in the Anthropocene. *Nat. Geosci.* **2016**, *9*, 89–91. [CrossRef]
5. Łabędzki, L.; Bąk, B. Meteorological and agricultural drought indices used in drought monitoring in Poland: A review. *Meteorol. Hydrol. Water Manag.* **2014**, *2*, 3–14. [CrossRef]
6. Yihdego, Y.; Vaheddoost, B.; Al-Weshah, R.A. Drought indices and indicators revisited. *Arab. J. Geosci.* **2019**, *12*, 69. [CrossRef]
7. McKee, T.B.; Doesken, N.J.; Kleist, J. The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology, Anaheim, CA, USA, 17–22 January 1993.
8. Deo, R.C.; Kisi, O.; Singh, V.P. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmos. Res.* **2017**, *184*, 149–175. [CrossRef]

9. Anshuka, A.; van Ogtrop, F.F.; Willem Vervoort, R. Drought forecasting through statistical models using standardised precipitation index: A systematic review and meta-regression analysis. *Nat. Hazards* **2019**, *97*, 955–977. [CrossRef]
10. Beguería, S.; Vicente-Serrano, S.M.; Reig, F.; Latorre, B. Standardised precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.* **2014**, *34*, 3001–3023. [CrossRef]
11. Cook, B.I.; Smerdon, J.E.; Seager, R.; Coats, S. Global warming and 21st century drying. *Clim. Dyn.* **2014**, *43*, 2607–2627. [CrossRef]
12. Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A multiscalar drought index sensitive to global warming: The standardised precipitation evapotranspiration index. *J. Clim.* **2010**, *23*, 1696–1718. [CrossRef]
13. Vicente-Serrano, S.M.; Beguería, S.; Lorenzo-Lacruz, J.; Camarero, J.J.; López-Moreno, J.I.; Azorin-Molina, C.; Revuelto, J.; Morán-Tejeda, E.; Sanchez-Lorenzo, A. Performance of drought indices for ecological, agricultural, and hydrological applications. *Earth Interact.* **2012**, *16*, 1–27. [CrossRef]
14. Li, L.; She, D.; Zheng, H.; Lin, P.; Yang, Z.-L. Elucidating diverse drought characteristics from two meteorological drought indices (SPI and SPEI) in China. *J. Hydrometeorol.* **2020**, *21*, 1513–1530. [CrossRef]
15. Nguvava, M.; Abiodun, B.J.; Otieno, F. Projecting drought characteristics over East African basins at specific global warming levels. *Atmos. Res.* **2019**, *228*, 41–54. [CrossRef]
16. Alawsi, M.A.; Zubaidi, S.L.; Al-Ansari, N.; Al-Bugharbee, H.; Ridha, H.M. Tuning ANN Hyperparameters by CPSOCGSA, MPA, and SMA for Short-Term SPI Drought Forecasting. *Atmosphere* **2022**, *13*, 1436. [CrossRef]
17. Padmanaban, K. A Novel Groundwater Resource Forecasting Technique for Cultivation Utilizing Wireless Sensor Network (WSN) and Machine Learning (ML) Model. *Turk. J. Comput. Math. Educ. (TURCOMAT)* **2021**, *12*, 2186–2192.
18. Poornima, S.; Pushpalatha, M. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Comput.* **2019**, *23*, 8399–8412. [CrossRef]
19. Shen, R.; Huang, A.; Li, B.; Guo, J. Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *79*, 48–57. [CrossRef]
20. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
21. Covert, I.C.; Lundberg, S.; Lee, S.-I. Explaining by removing: A unified framework for model explanation. *J. Mach. Learn. Res.* **2021**, *22*, 9477–9566.
22. Lubo-Robles, D.; Devegowda, D.; Jayaram, V.; Bedle, H.; Marfurt, K.J.; Pranter, M.J. Machine learning model interpretability using SHAP values: Application to a seismic facies classification task. In Proceedings of the SEG International Exposition and Annual Meeting, Virtual, 11–16 October 2020.
23. McLennan, W. *Year Book Australia 2000*; Australian Bureau of Statistics: Canberra, Australia, 2000.
24. O'Neill, C.; Chandler-Ho, S. Decreasing water budget of the Australian continent from Grace satellite gravity data. *arXiv* **2021**, arXiv:2101.11167.
25. King, A.D.; Pitman, A.J.; Henley, B.J.; Ukkola, A.M.; Brown, J.R. The role of climate variability in Australian drought. *Nat. Clim. Chang.* **2020**, *10*, 177–179. [CrossRef]
26. Yadav, A.; Das, S.; Bakar, K.S.; Chakrabarti, A. Understanding the complex dynamics of climate change in south-west Australia using Machine Learning. *Phys. A Stat. Mech. Its Appl.* **2023**, *627*, 129139. [CrossRef]
27. McDonald, T. Drought, fire, flood and COVID—Complex systems and disruption. *Ecol. Manag. Restor.* **2020**, *21*, 73. [CrossRef]
28. Collins, L.; Bennett, A.F.; Leonard, S.W.; Penman, T.D. Wildfire refugia in forests: Severe fire weather and drought mute the influence of topography and fuel age. *Glob. Chang. Biol.* **2019**, *25*, 3829–3843. [CrossRef] [PubMed]
29. Mariani, M.; Connor, S.; Fletcher, M.-S.; Romano, A.; Maezumi, S. Higher fuel loads and more fire follow removal of Indigenous cultural burning across southeast Australia. *Past Glob. Chang. Mag.* **2022**, *30*, 34–35. [CrossRef]
30. Bowman, D.; Williamson, G.; Yebra, M.; Lizundia-Loiola, J.; Pettinari, M.L.; Shah, S.; Bradstock, R.; Chuvieco, E. Wildfires: Australia needs national monitoring agency. *Nature* **2020**, *584*, 188–191. [CrossRef]
31. Nolan, R.H.; Boer, M.M.; Collins, L.; Resco de Dios, V.; Clarke, H.; Jenkins, M.; Kenny, B.; Bradstock, R.A. Causes and consequences of eastern Australia's 2019–20 season of mega-fires. *Glob. Chang. Biol.* **2020**, *26*, 1039–1041. [CrossRef]
32. Shi, L.; Feng, P.; Wang, B.; Liu, D.L.; Yu, Q. Quantifying future drought change and associated uncertainty in southeastern Australia with multiple potential evapotranspiration models. *J. Hydrol.* **2020**, *590*, 125394. [CrossRef]
33. Penman, H.L. Natural evaporation from open water, hare soil and grass. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **1948**, *193*, 120–145. [CrossRef]
34. Abtew, W. Evapotranspiration measurements and modeling for three wetland systems in South Florida. *JAWRA J. Am. Water Resour. Assoc.* **1996**, *32*, 465–473. [CrossRef]
35. Hargreaves, G.L.; Hargreaves, G.H.; Riley, J.P. Irrigation water requirements for Senegal River basin. *J. Irrig. Drain. Eng.* **1985**, *111*, 265–275. [CrossRef]
36. Jensen, M.E.; Haise, H.R. Estimating evapotranspiration from solar radiation. *J. Irrig. Drain. Div.* **1963**, *89*, 15–41. [CrossRef]
37. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao Rome* **1998**, *300*, D05109.
38. Thornthwaite, C.W. An approach toward a rational classification of climate. *Geogr. Rev.* **1948**, *38*, 55–94. [CrossRef]
39. Priestley CH, B.; Taylor, R.J. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Mon. Weather Rev.* **1972**, *100*, 81–92. [CrossRef]

40. Yang, Y.; Roderick, M.L.; Zhang, S.; McVicar, T.R.; Donohue, R.J. Hydrologic implications of vegetation response to elevated CO_2 in climate projections. *Nat. Clim. Chang.* **2019**, *9*, 44–48. [CrossRef]
41. Achite, M.; Elshaboury, N.; Jehanzaib, M.; Vishwakarma, D.K.; Pham, Q.B.; Anh, D.T.; Abdelkader, E.M.; Elbeltagi, A. Performance of machine learning techniques for meteorological drought forecasting in the wadi mina basin, Algeria. *Water* **2023**, *15*, 765. [CrossRef]
42. Lotfirad, M.; Esmaeili-Gisavandani, H.; Adib, A. Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran. *J. Water Clim. Chang.* **2021**, *13*, 383–406. [CrossRef]
43. Lucas, C.; Hennessy, K.; Mills, G.; Bathols, J. *Bushfire Weather in Southeast Australia: Recent Trends and Projected Climate Change Impacts*; Bushfire CRC: Melbourne, Australia, 2007.
44. Grant, S.B.; Fletcher, T.D.; Feldman, D.; Saphores, J.-D.; Cook, P.L.; Stewardson, M.; Low, K.; Burry, K.; Hamilton, A.J. Adapting urban water systems to a changing climate: Lessons from the millennium drought in southeast Australia. *Environ. Sci. Technol.* **2013**, *47*, 10727–10734. [CrossRef]
45. Köppen, W.; Geiger, R. *Handbuch der Klimatologie*; Gebrüder Borntraeger Berlin: Berlin, Germany, 1930; Volume 1.
46. Gannon, K.E.; Conway, D.; Pardoe, J.; Ndiyoi, M.; Batisani, N.; Odada, E.; Olago, D.; Opere, A.; Kgosietsile, S.; Nyambe, M. Business experience of floods and drought-related water and electricity supply disruption in three cities in sub-Saharan Africa during the 2015/2016 El Niño. *Glob. Sustain.* **2018**, *1*, e14. [CrossRef]
47. Milly PC, D.; Dunne, K.A. Potential evapotranspiration and continental drying. *Nat. Clim. Chang.* **2016**, *6*, 946–949. [CrossRef]
48. Brutsaert, W. *Evaporation Into the Atmosphere: Theory, History, and Applications*; Cornell University: New York, NY, USA; Springer: Berlin/Heidelberg, Germany, 1982. [CrossRef]
49. Brutsaert, W. *Hydrology: An Introduction*; Cornell University: New York, NY, USA, 2005. [CrossRef]
50. Pei, Z.; Fang, S.; Wang, L.; Yang, W. Comparative Analysis of Drought Indicated by the SPI and SPEI at Various Timescales in Inner Mongolia, China. *Water* **2020**, *12*, 1925. [CrossRef]
51. PyCaret. PyCaret Version 1.0.0; Python; 2020. Available online: https://pycaret.org (accessed on 20 December 2023).
52. Jiao, T.; Williams, C.A.; Rogan, J.; De Kauwe, M.G.; Medlyn, B.E. Drought impacts on Australian vegetation during the millennium drought measured with multisource spaceborne remote sensing. *J. Geophys. Res. Biogeosci.* **2020**, *125*, e2019JG005145. [CrossRef]
53. Vicente-Serrano, S.M.; Van der Schrier, G.; Beguería, S.; Azorin-Molina, C.; Lopez-Moreno, J.-I. Contribution of precipitation and reference evapotranspiration to drought indices under different climates. *J. Hydrol.* **2015**, *526*, 42–54. [CrossRef]
54. Deo, R.C.; Şahin, M. Application of the Artificial Neural Network model for prediction of monthly Standardised Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. *Atmos. Res.* **2015**, *161–162*, 65–81. [CrossRef]
55. Dikshit, A.; Pradhan, B.; Huete, A. An improved SPEI drought forecasting approach using the long short-term memory neural network. *J. Environ. Manag.* **2021**, *283*, 111979. [CrossRef]
56. Aminzade, J. Projections of Future Drought. In *Our Warming Planet*; World Scientific: Singapore, 2018; pp. 231–249. [CrossRef]
57. Rosenzweig, C.; Rind, D.; Lacis, A.; Manley, D. *Our Warming Planet*; World Scientific: Singapore, 2018. [CrossRef]
58. Zhang, X.; Zhang, Y.; Ma, N.; Kong, D.; Tian, J.; Shao, X.; Tang, Q. Greening-induced increase in evapotranspiration over Eurasia offset by CO_2-induced vegetational stomatal closure. *Environ. Res. Lett.* **2021**, *16*, 124008. [CrossRef]
59. Wang, Q.; Liu, X.; Wang, Z.; Zhao, L.; Zhang, Q.-P. Time scale selection and periodicity analysis of grassland drought monitoring index in Inner Mongolia. *Glob. Ecol. Conserv.* **2022**, *36*, e02138. [CrossRef]
60. Lara, A.; Jones, J.; Little, C.; Vergara, N. Streamflow response to native forest restoration in former Eucalyptus plantations in south central Chile. *Hydrol. Process.* **2021**, *35*, e14270. [CrossRef]
61. Li, Z.; Zhou, P.; Shi, X.; Li, Y. Forest effects on runoff under climate change in the Upper Dongjiang River Basin: Insights from annual to intra-annual scales. *Environ. Res. Lett.* **2020**, *16*, 014032. [CrossRef]