Original Article

# Journal of Drought and Climate change Research (JDCR)

Winter 2025, Vol. 2, No. 8, pp 85-102

10.22077/jdcr.2024.7376.1061

# Efficiency of Machine Learning Techniques for Predicting Vapor Pressure Deficit in Arid and Semi-Arid Regions (Case Study: South Khorasan Province)

**Elham Ghochanian Haghverdi[1*], Hossein Khozeymeh Nezhad[2], AliReza Moghri Friz[3], Omid Khorashadizadeh[4]**

1- Department of Water Science and Engineering, Faculty of Agriculture, University of Birjand, Birjand, Iran.
2- Department of Water Science and Engineering, Faculty of Agriculture, University of Birjand, Birjand, Iran.
3- Department of Soil and Water Research Department, South Khorasan Agriculture and Natural Resources Research and Education Center, AREEO, Birjand, Iran.
4- Regional Water Company of south Khorasan, Water Resources Management Company, Birjand, Iran.
*** corresponding author: e.ghochanian@areeo.ac.ir**

**Keywords**:
Vapor pressure deficit, GAM, Climate change, Machine learning, Drought

**Received:**
02 March 2024

**Revised:**
28 April 2024

**Accepted:**
11 May 2024

## Abstract

Climate change, as one of the global challenges of the present century, has profound impacts on water resources and agriculture. Increase in temperature and decrease in rainfall in arid and semi-arid regions have made the optimal water resource management a top priority. In countries facing climate change and drought, accurate estimation of evapotranspiration plays a vital role in water resource management and ensuring food security. One of the key factors affecting evapotranspiration is the vapor pressure deficit (VPD), which significantly impacts the accuracy of related calculations. This study focuses on predicting the vapor pressure deficit using advanced machine learning techniques. The methods employed include Linear Regression (LR), Generalized Additive Model (GAM), Random Subspace (RSS), Random Forest (RF), and M5 Pruned model (M5P). In this study, monthly average data, including temperature, humidity, precipitation, and vapor pressure deficit, were extracted from the Japanese 55-year Reanalysis (JRA-55) database for the period from 1958 to 2023. The analysis on the vapor pressure deficit data in Birjand, Sarayan, Qaen, and Tabas showed that the annual average VPD increased by 6 Pa, 10 Pa, 4 Pa, and 5 Pa, respectively. In the next step, the extracted data for temperature, precipitation, and humidity were

**How to cite this article:**
Ghochanian Haghverdi, E., Khozeymeh Nezhad, H., Moghri Friz, A. & Khorashadizadeh, O. (2024). Efficiency of Machine Learning Techniques for Predicting Vapor Pressure Deficit in Arid and Semi-Arid Regions (Case Study: South Khorasan Province). *Journal of Drought and Climate change Research* (JDCR), 2(8), 85-102. 10.22077/jdcr.2024.7376.1061

used as input variables, and VPD was used as the target variable in machine learning algorithms. Model performance was evaluated using root mean square error (RMSE), mean absolute error (MAE), Pearson correlation coefficient (CC), and Kling-Gupta efficiency (KGE). Results showed that the GAM model outperformed other models in all regions. The evaluation indices for each region were as follows: Birjand [RMSE=0.308, MAE=0.247, KGE=0.914, and CC=0.920], Sarayan [RMSE=0.401, MAE=0.303, KGE=0.937, and CC=0.951], Qaen [RMSE=0.072, MAE=0.055, KGE=0.987, and CC=0.997] and Tabas [RMSE=0.230, MAE=0.184, KGE=0.920, and CC=0.942]. Predictions showed that, over the next 10 years, the annual average VPD in the studied regions will significantly increase. This increment is as follows: Birjand 9 Pa, Sarayan 10 Pa, Qaen 7 Pa, and Tabas 5 Pa. This increase signifies serious challenges for water resources and an increase in water consumption. Eventually, this study suggests the GAM model as an effective tool for future research, especially for use in the development of smart irrigation systems, which play a crucial role in sustainable water resource management.

## Introduction

Agriculture consumes the most water, using 70 % of all freshwater withdrawals on average. However, in some underdeveloped nations, this percentage can reach 95 % (Wada & Bierkens., 2014). To ensure future food security, improving irrigation efficiency is essential to produce more crops using less water (Smidte et al., 2016). Even with advanced biotechnology and conventional breeding techniques, achieving significant yields is only possible with sufficient water and proper crop and soil management (Ahmar et al., 2020). Vapor pressure deficit (VPD) is a critical parameter in precise agricultural water management, impacting moisture flow from the surface to the atmosphere and the water balance at national and global levels (Kimball et al., 1997; Qiu & Katul., 2020). It affects plant physiology and significantly influences plant water requirements and evapotranspiration (Grossiord et al., 2020; Qiu et al., 2019; Yuan et al., 2019). Multiple studies show that VPD significantly impacts evaporation and transpiration (Zhang et al., 2018). The amount of vapor pressure deficit represents the difference between saturation and actual pressures (Rawson et al., 1977). According to the Clausius-Clapeyron equation (Iribarne & Godson, 1981; Bolton, 1980), the saturation vapor pressure of water increases by approximately 7% for each degree Kelvin increases in atmospheric temperature. If the increase in saturation vapor pressure does not correspond with the actual atmospheric water vapor concentration, the vapor pressure deficit

will increase. Relative humidity, which measures the difference between the actual vapor pressure of water and the saturation pressure, fluctuates significantly in humid areas and inland regions (Pierce et al., 2013). Although the long-term trend of the global average relative humidity at the earth's surface is low. (Willett et al., 2008; Dai, 2006) a significant decrease has been observed after 2000 (Simmons et al., 2010; Willett et al., 2014), indicating a severe increase in the vapor pressure deficit at the surface. Recent research has shown that the increase in vapor pressure deficit, rather than changes in precipitation, significantly impacts crop yields, plant transpiration, and evaporation (Konings et al., 2017; Ding et al., 2018; Restaino et al., 2016; Carnicer et al., 2013). Furthermore, the increase in vapor pressure deficit alters plant stomatal activity, reducing vegetation cover over vast areas of land (Fletcher et al., 2007). VPD has a significant impact on plant production in forested areas and lands with trees and shrubs, whereas soil moisture plays a more critical role in rangelands (Sun et al., 2024). Currently, the earth is experiencing atmospheric drying across the globe due to increased vapor pressure deficiency, and it is predicted that this condition will worsen with the decline in global climate quality. (Dai et al., 2018; Ficklin & Novick, 2017; Liu & Sun., 2017; Wang et al., 2012). This effect is mainly due to the increase in saturated vapor pressure, which raises global temperatures, and the decrease in actual vapor pressure, which affects various hydrological phenomena (Ficklin & Novick., 2017). As a result of the increased vapor pressure deficit in agricultural and non-agricultural plants (Otieno et al., 2012), we are witnessing a decrease in plant productivity, which in turn leads to an increase in adverse climatic events such as droughts. In arid and semi-arid regions, the lack of atmospheric vapor pressure reduces the quality and productivity of agricultural products during the spring and summer growing seasons, especially in greenhouses. According to previous studies, increased humidity reduces leaf area and thickness (Devi et al., 2018). If the vapor pressure deficit is high, it delays leaf growth; if the atmosphere is humid, leaf wilting and disruption of xylem function may occur during the tree's growth period (Sellin et al., 2019). Actual VP ($e_a$) and standard VP ($e_s$) are widely used to determine the level of vapor pressure deficit. The present study's findings are essential for guiding future research in these areas and attracting policymakers' attention to the impact of vapor pressure deficit on vegetation and the overall hydrological cycle. Most studies have used vapor pressure deficit as one of the input variables similar to the model of the present study to develop machine learning-based models for predicting various hydro-climatic variables and

estimating the water needs of vegetation, which emphasizes the importance of improving their prediction accuracy and precision. (Feng et al., 2019; Huntington et al., 2020; khosravi et al., 2021; Mokhtar et al., 2021; Emami et al., 2022). In South Khorasan, more than 90 percent of water is used for agriculture. Therefore, accurately estimating the evapotranspiration for agricultural products is essential to achieve sustainability under climate change conditions and Iran's limited water share. Monthly vapor pressure deficits can be helpful tools for scheduling irrigation, as they reflect the atmospheric demand for water by plants. The goal of irrigation scheduling is to use the appropriate amount of water at the right time to meet crops' water needs and optimize yield while minimizing water consumption. As mentioned, the vapor pressure deficit is the difference between the amount of moisture in the air and the maximum amount of moisture the air can hold. This index is influenced by temperature and relative humidity and can be calculated using meteorological data. High values of vapor pressure deficit indicate a high atmospheric demand for water. This means that plants release water into the atmosphere more quickly and may require more water. Farmers can monitor their area's vapor pressure deficit values and adjust their irrigation practices to utilize the predicted monthly vapor pressure deficit for irrigation scheduling. For example, when the vapor pressure deficit is high, farmers may need to irrigate more to ensure the plants have enough water. Conversely, farmers can reduce water consumption when the vapor pressure deficit is low while maintaining optimal crop growth. The predicted monthly vapor pressure deficit can also be used to schedule irrigation in advance. Farmers can predict periods of high atmospheric demand for water by monitoring the forecast of vapor pressure deficit and schedule irrigation accordingly. This method can help farmers optimize water consumption and prevent over- or under-irrigation, which may lead to reduced crop yields and water wastage. The predicted monthly vapor pressure deficit can be a helpful tool for irrigation scheduling, allowing farmers to optimize water use and increase crop yields while minimizing water waste. To obtain accurate parameters such as soil moisture, soil temperature, and climatic parameters, the JRA-55 database can be used. In Iran, no studies or research have yet been conducted in the field of agriculture using JRA-55 data. (Mollasharifi et al., 2019) in their study of the impact of the North Atlantic Oscillation on the relationship between the North Atlantic storm tracks and the Mediterranean, have used NCPE/NCAR and JRA-55 data. In another study, (Azarm et al., 2022) have used the JRA-55 database for climatology of Bandal

events. However, this database has not yet been used in the fields of agriculture, water resources and climate change, which highlights the importance of this study. Considering the negative impacts of climate change in South Khorasan and the limitations of water resources in this province, the development of new methods and the use of machine learning algorithms instead of traditional irrigation methods must receive more attention from researchers to maintain sustainable agriculture and improve water use efficiency.

In this study, the prediction of the VPD parameter using machine learning algorithms was examined, marking the beginning of a path for future research aimed at developing intelligent irrigation systems to mitigate the effects of climate change and adapt drought conditions. The aim of this research is to assess the long-term predictive capabilities of several algorithms such as Linear Regression (LR), General Additive Model (GAM), Random SubSpace (RSS), Random Forest (RF), and M5P for Vapor Pressure Deficit (VPD) values in four regions of South Khorasan, including Tabas, Birjand, Qaen, and Sarayan, facing serious water resource limitations.

Additionally, selecting the best machine learning model for predicting VPD based on statistical indicators such as accuracy, high performance, and low statistical errors is another goal. Moreover, this paper aims to familiarize water resource researchers with the JRA-55 database system, particularly in the context of agricultural approaches and vapor pressure deficit.

## Materials and Methods

### Study area

South Khorasan Province is one of the easternmost provinces of Iran, covering an area of approximately 150,000 square kilometers, which accounts for 22.6% of the country's total area. The province is located between 57 degrees and 1 minute to 60 degrees and 57 minutes of eastern longitude, and between 30 degrees and 32 minutes to 34 degrees and 36 minutes of northern latitude. According to the latest administrative divisions, the province currently includes 12 counties, 32 cities, 28 districts, and 66 rural districts. Due to the province's location between deserts and scattered mountain ranges, the climate in the southern and southwestern regions is hot and dry, while the climate in the northern and northeastern mountainous areas is semi-arid, mild, and cold. The mountain ranges, with their northwestern-southeastern orientation, contribute to the diversity and extreme fluctuations in the region's climate. The province has 35 study areas for water resource management, of which 8 are classified as critical prohibited areas, 18 as critical areas, and 9 as accessible areas. According to recent statistics, the long-term average

annual rainfall in the province is about 113 millimeters, indicating a severe shortage of water resources. This climatic condition and the limitation of water resources significantly affect agriculture and the livelihood of the local population. In this province around 88% of groundwater resources are used for agricultural purposes. This comes from three primary sources: wells (73%), qanats (23%), and springs (4%). The province has been dealing with over two decades of continuous drought and significant reductions in rainfall, which have severely affected both surface and groundwater resources. This heavy reliance on groundwater underscores the critical need for sustainable water management and policies to combat the ongoing water scarcity crisis

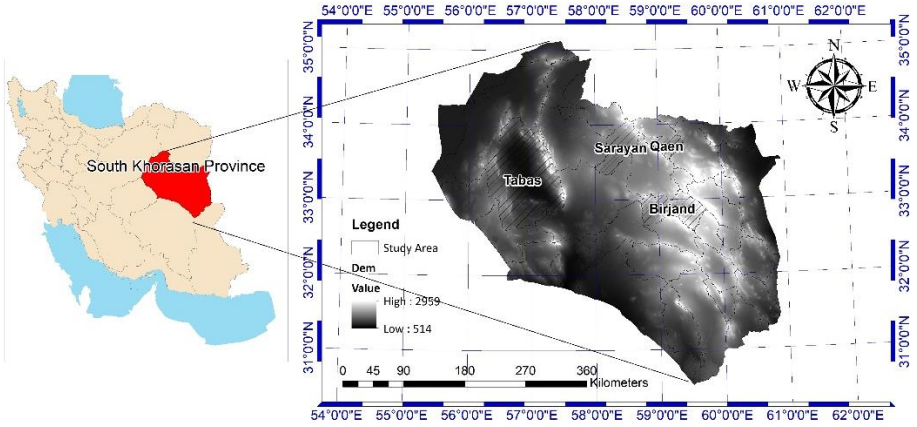


**Fig 1. Study area**

**Introducing the Japanese 55-year Reanalysis**

JRA-55 (Japanese 55-Year Reanalysis) is a global atmospheric reanalysis dataset developed by the Japan Meteorological Agency (JMA). Spanning the period from 1958 to the present, it is a crucial resource for studying climate variability and long-term changes. JRA-55 was designed to enhance previous versions, such as JRA-25, by addressing issues like model biases and gaps in observational data. Key features of JRA-55 include:

1. **Data Sources:** It incorporates diverse observational data from sources like weather stations, satellites, and aircraft, ensuring a consistent and comprehensive atmospheric dataset.
2. **Spatial and Temporal Coverage:** With more than five decades of data, JRA-55 provides detailed climate information applicable to various fields, including climate research, weather forecasting, and hydrology.
3. **Improvements:** Enhancements in areas such as temperature analysis, long wave radiation schemes, and the representation of oceanic fronts make

JRA-55 more accurate than earlier reanalysis products.

4. **Applications:** Widely used in atmospheric studies, climate modeling, and global trend assessments, JRA-55 is particularly valuable for research on climate variability and hydrological cycles.

This dataset supports climate scientists, meteorologists, and researchers in better understanding climate systems, evaluating the impacts of climate change, and informing decisions related to environmental policies and resource management. For more detailed information, visit the official JRA-55 website.

Temperature, humidity, precipitation, and vapor pressure deficit data were extracted monthly from the JRA-55 database in NC format files. Then, using the NC files containing monthly statistics for one year (12 months), the Make NetCDF Raster Layer tool in ArcMap software was used to convert these files into raster format. Subsequently, using the Zonal Statistics as Table tool and the boundary layer of the study areas, including Qaen, Birjand, Sarayan, and Tabas, the average data within the borders of these areas were calculated. (Fig 2 shows the process of extracting data from JRA-55 using ArcMap software).

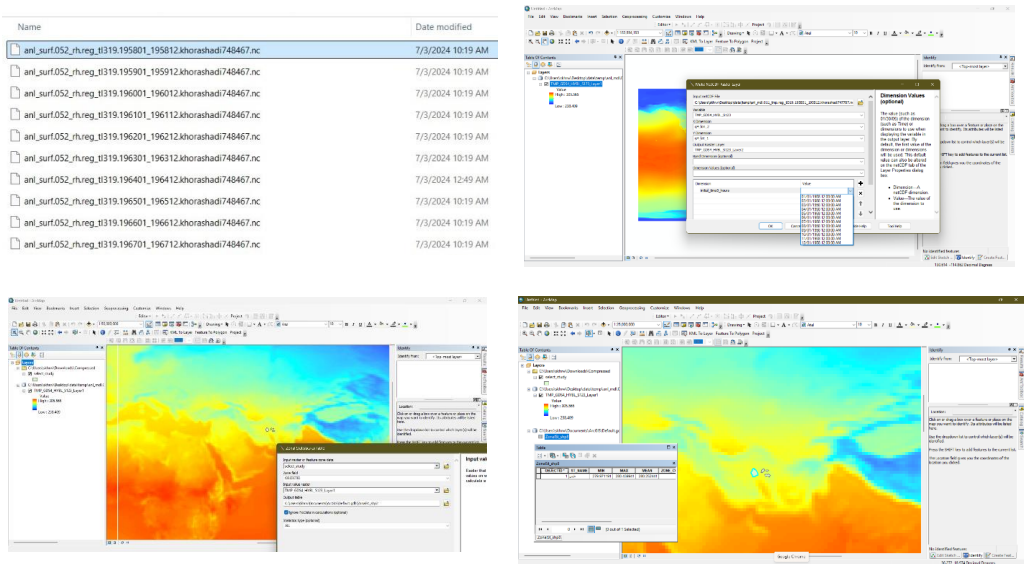


**Fig 2. The process of extracting data from the JRA-55 database using ArcMap software.**

## Machine learning Model description

The present study examines the performance of machine learning models in estimating monthly vapor pressure deficit (VPD) in four study areas of South Khorasan Province. MATLAB software was used to run the models. The data for this region is divided into two sections. The first group consists of 70% for training, and the second group consists of 30% for testing. This study analyzes five machine learning models: linear

regression, generalized additive model, random subset, random forest, and M5P model. The following sections provide a complete explanation of these models.

**1. Linear regression (LR)**

Linear regression is one of the simplest and most widely used statistical modeling methods, which models the linear relationship between a dependent variable and one or more independent variables. In multiple linear regression, where there is more than one independent variable, the model is expanded as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (1)$$

$x_1, x_2, \ldots, x_p$ : Independent variables

are regression coefficients for each independent variable indicate the effect of changes in each of these variables on y.

**2. General additive model (GAM)**

GAM is a combination of generalized linear models with additive models. These models have been presented to create a unified method for statistical models such as linear, logistic, and Poisson regression. The general idea of this method is that instead of assuming various functions in the mentioned regressions, knowing the data distribution—which generally comes from a logical assumption of the problem—one can obtain their estimate for the model. This model consists of three components:

1- A distribution for the target variable y, which is usually chosen from the exponential family with dispersion parameter.

$$f_Y(yI\theta,\tau) = h(y,c)\exp\left(\frac{b(\theta)^T T(y) - A(\theta)}{d(\tau)}\right) \quad (2)$$

Pre-dispersion is used for modeling high and low variances

$$\eta = X\beta \quad (3)$$

Pre-dispersion is used for modeling high. The link function g is a strictly monotonic function that connects the two above components with low variances.

**3. Random Sub Space (RSS)**

The RSS model trains and integrates multiple categories in spaces with different features. This framework categorizes several subsets of data for training, which later serves as the basis for training and considers clustering and self-supervised approaches. While utilizing artificial neural networks, decision trees, or other algorithms, this framework demonstrates nonlinear relationships. This system's classification is developed according to the RSS framework and is related to specific data features. The output of all classifications is easily combined by the voting system in the model. This method improves the ability of each weak classifier. (Skurichina & Duin., 2002) presented the RSS algorithm as follows.

$$\gamma(S) = argmax \sum \delta_{sgn(c^d(s),y} \quad (4)$$

In this formula, $C^d(S)$ represents the classifier (d = 1,2…D); $\delta_{i,j}$ denotes the Kronecker delta, and y = (-1,1) signifies the class label or classification decision

**4. Random Forest (RF)**

Random forest is a type of ensemble learning algorithm that combines multiple decision trees for prediction. Each decision tree in the random forest is trained on a random subset of the training data and a random subset of the features. Then, the random forest is determined by aggregating the predictions of all the decision trees. This approach allows the random forest to be very accurate and resistant to the problem of overfitting (a common issue in machine learning where the model performs well on training data but poorly on new data).

$$d(x,y) = \sqrt{I - s(x,y)} \quad (5)$$

The RF technique can significantly enhance the estimation efficiency of the system with minimal possible errors and the least noise. RF can effectively work with large-scale datasets with high dimensions (Khosravi et al., 2019; Li et al., 2022)

**5.M5 Pruned (M5P)**

Reconstructing the M5 Quinlan algorithm helps in generating the M5P model tree. The M5P algorithm is reconstructed to the leaf nodes with a convolution decision tree and a linear regression function (Blaifi et al., 2018). The M5P model algorithm is based on a numerical prediction system. The linear regression model is stored in each leaf, which helps identify the corresponding point cluster value that reaches the leaf. (Shamshirband et al., 2020) identify the best feature split criterion for partitioning a specific section (T) of training data related to a specific node. The standard deviation of the batch related to T is measured as the error for that node. In addition, the reduction of potential error is determined by making estimates at each node. The selection of features used for splitting increases the potential error reduction of the associated node. Minimizing the expected error with standard deviation reduction (SDR) is performed as follows.

$$SDR = sd(T) - \sum \frac{|T_i|}{T} \times sd(T_i) \quad (6)$$

In this formula, $T_i$ corresponds to $T_1$, $T_2$, $T_3$, ... obtained from the division of nodes based on selected features.

Table 1 and Figure 3 shows the workflow diagram related to data collection, processing, and modeling of vapor pressure deficit based on machine learning algorithms from 1955 to 2023.

**Table 1. The parameters of the machine learning algorithm used for modeling VPD in the study regions.**

| Model name | Description of parameters | Cross-validation | Input | Target |
|---|---|---|---|---|
| Linear regression (LR) | Batch size-100, Debug = False, Eliminate Collinear Attribute =True, Minimal = False | K – fold;k=5 | Temp<br>Wind<br>Humidity<br>Precipitation | VPD |
| Random Subspace (RSS) | Batch size-100, Classifier = REPTree, random seed-1, subspace size = 0. 5, numbers of executions slots = 1, number of iterations= 10 | | | |
| M5 Pruned (M5P) | Batch size-100, Minimum number of instances = 4 | | | |
| General additive model (GAM) | Batch size-100, Name-value arguments=optimize Hyper parameters, parameters to optimize=auto, | | | |

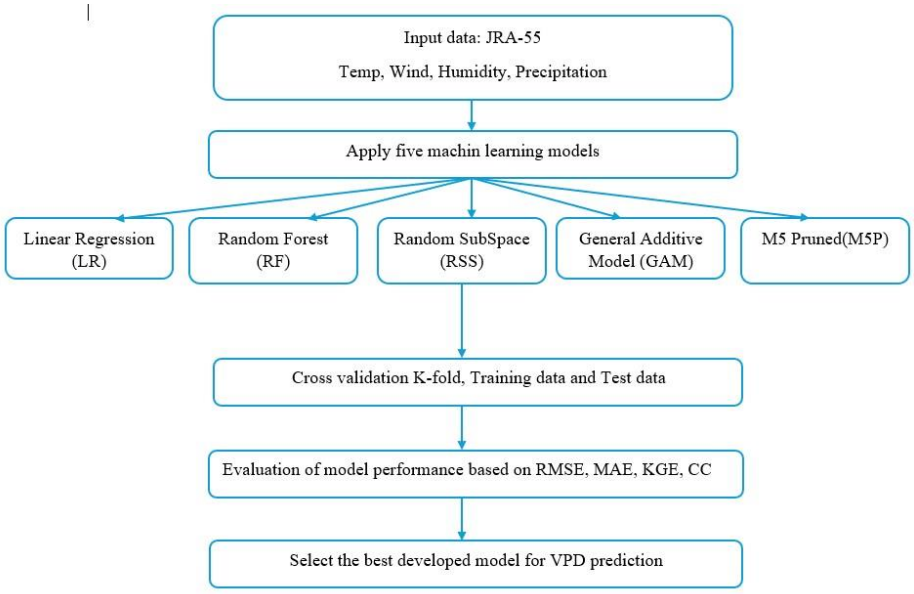


**Fig 3. Workflow diagram related to data collection, processing, and modeling of vapor pressure deficit based on machine learning algorithms**

## Evaluation of model performance

The performance of machine learning models has been evaluated using the following four performance metrics: Correlation Coefficient (CC), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Kling-Gupta Efficiency (KGE). These indices have been calculated using the following equations: where N is the total number of measurements, $X_a$ is the observed values, $Y_a$ is the estimated values, is the mean of the observed values in the X variables, and is the mean of the estimated values in the Y variables.

$$CC = \frac{\sum_a^N (X_a - \bar{X})(Y_a - \bar{Y})}{\sqrt{\sum_a^N (X_a - \bar{X})^2 \sum_a^N (Y_a - \bar{Y})^2}} \quad (7)$$

$$MAE = \frac{1}{N}\sum_a^N |X_a - Y_a| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_a^N (X_a - Y_a)^2} \quad (9)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (10)$$

## Results and Discussion

Evaluation of the statistical performance of machine learning models based on VPD prediction. The graph of vapor pressure deficit changes from 1958 to 2023 for the regions of Birjand, Qaen, Sarayan, and Tabas is shown in Figure 4.

According to the figure 4, the average increase in vapor pressure deficit is 6 Pa in Birjand, 10 Pa in Sarayan, 4 Pa in Qaen, and 5 Pa in Tabas.The performance of the LR, GAM, RSS, RF, and M5P models for monthly vapor pressure deficit prediction


**Fig4. Comparison chart of conversion in vapor pressure deficit**

using temperature, precipitation, wind speed, and humidity data was evaluated. In the process of model implementation, the Cross-Validation K-fold statistical technique was used to assess the models' performance and reduce issues caused by overfitting or underfitting. This method helps ensure that the model performs well on new data. In this method, the data is divided into K subsets (folds). In each iteration, one of these subsets is used as test data, and the remaining subsets are used as training data. Four statistical indices, CC, MAE, RMSE, and KGE, were used to evaluate the models for the regions of Birjand, Qaen, Sarayan, and Tabas during both training and testing stages. The criterion for selecting the best model was the lowest error value in the test phase. In the Birjand region, the results showed that the GAM model had the best performance for all evaluation indices, with the highest values for CC and KGE and the lowest values for MAE and RMSE. The performance of GAM, in terms of these statistics, was better than the other models. The LR model was the second best model, with the lowest RMSE and MAE values and the highest KGE and CC values. The third model, which showed good performance for predicting vapor pressure deficit, was the RF model. The performance of the RSS model was average, and the M5P model showed poor performance in predicting VPD (Table 1). In the Qaen region, by evaluating the performance of the models in the test phase, it was found that the GAM model had the best performance, with the highest value for CC and the lowest values for MAE, RMSE, and also a KGE close to 1. In this region, after the GAM model, the RF and LR models were selected as the best models, and the performance of the

**Table 2. Performance of the models under training and testing conditions (Birjand)**

| Model | Training period (1958-2003) | | | | Testing period (2003-2023) | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | KGE | CC | RMSE | MAE | KGE | CC |
| LR | 0.291 | 0.231 | 0.914 | 0.939 | 0.312 | 0.252 | 0.928 | 0.931 |
| GAM | 0.237 | 0.183 | 0.934 | 0.960 | 0.259 | 0.206 | 0.943 | 0.952 |
| RSS | 0.453 | 0.326 | 0.729 | 0.930 | 0.462 | 0.358 | 0.727 | 0.907 |
| RF | 0.171 | 0.13 | 0.968 | 0.979 | 0.352 | 0.264 | 0.911 | 0.912 |
| M5P | 0.781 | 0.394 | 0.563 | 0.666 | 1.213 | 0.965 | -0.066 | -0.027 |

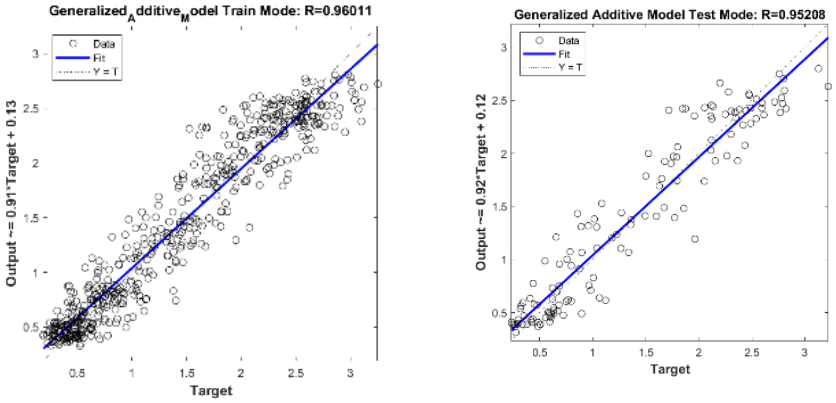


**Fig 5. Graph of the correlation results taken from GAM model for the Birjand region**.

RSS model was average, while the M5P model had a poor performance compared to the other models.

By evaluating the performance of the models in the Sarayan region, the GAM model was selected as the best model in both the test phases, with the maximum values for KGE and CC and the minimum values for MAE and RMSE. In this region, the second-best model was LR. The RF

**Table3. Performance of the models under training and testing conditions (Qaen)**

| model | Training period (1958-2003) | | | | Testing period (2003-2023) | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | KGE | CC | RMSE | MAE | KGE | CC |
| LR | 0.045 | 0.035 | 0.999 | 0.999 | 0.083 | 0.059 | 0.994 | 0.996 |
| GAM | 0.035 | 0.027 | 0.999 | 0.999 | 0.077 | 0.058 | 0.996 | 0.997 |
| RSS | 0.214 | 0.134 | 0.888 | 0.984 | 0.200 | 0.146 | 0.889 | 0.990 |
| RF | 0.192 | 0.156 | 0.968 | 0.978 | 0.189 | 0.152 | 0.960 | 0.980 |
| M5P | 0.704 | 0.360 | 0.598 | 0.754 | 0.836 | 0.648 | 0.514 | 0.683 |

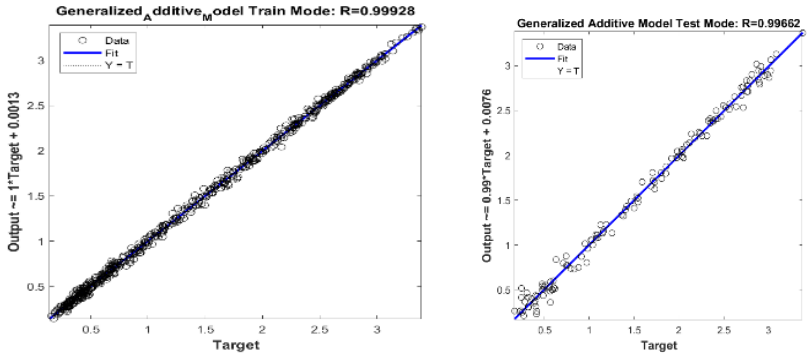


**Fig 6. Graph of the correlation results in the GAM model for the Qaen region**

model ranked third with a slight difference in performance. The RSS model showed average performance, similar to the other regions. The M5P model was also selected as a model with poor performance.

In the Tabas study area, the GAM model with the following error values [RMSE = 0.226, MAE = 0.180, KGE = 0.915, CC = 0.945, was selected as the best model. The second model, which showed good performance for predicting vapor pressure deficit, was the LR model, with a very slight difference from the RF model. As a result, the RF model was chosen as the third model. Additionally, the RSS model showed average performance, and the M5P model had poor performance in this region.

**Table 4. Performance of the models under training and testing conditions (Sarayan)**

| model | Training period (1958-2003) | | | | Testing period (2003-2023) | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | KGE | CC | RMSE | MAE | KGE | CC |
| LR | 0.424 | 0.341 | 0.919 | 0.942 | 0.393 | 0.312 | 0.924 | 0.952 |
| GAM | 0.284 | 0.223 | 0.958 | 0.975 | 0.378 | 0.290 | 0.953 | 0.956 |
| RSS | 0.565 | 0.406 | 0.756 | 0.952 | 0.567 | 0.435 | 0.778 | 0.945 |
| RF | 0.205 | 0.160 | 0.976 | 0.987 | 0.440 | 0.338 | 0.929 | 0.942 |
| M5P | 1.156 | 0.586 | 0.492 | 0.644 | 1.609 | 1.229 | 0.079 | 0.165 |

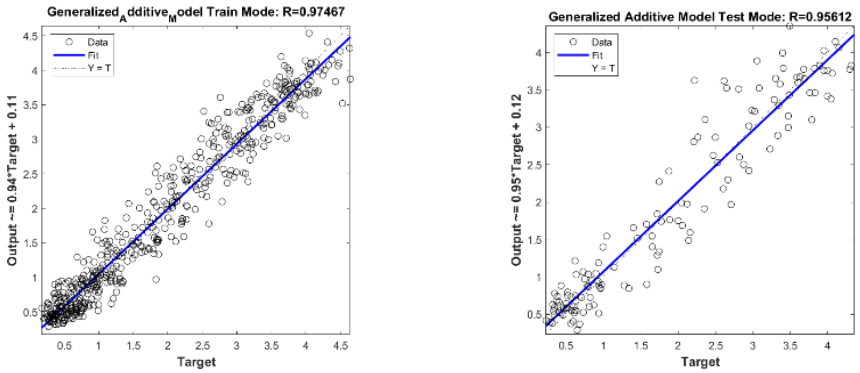


**Fig 7. Graph of the correlation results in the GAM model for the Sarayan region**

**Table 5. Performance of the models under training and testing conditions (Tabas)**

| model | Training period (1958-2003) | | | | Testing period (2003-2023) | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | KGE | CC | RMSE | MAE | KGE | CC |
| LR | 0.248 | 0.199 | 0.910 | 0.937 | 0.275 | 0.221 | 0.914 | 0.919 |
| GAM | 0.166 | 0.132 | 0.949 | 0.972 | 0.226 | 0.180 | 0.915 | 0.945 |
| RSS | 0.338 | 0.245 | 0.766 | 0.945 | 0.370 | 0.289 | 0.760 | 0.913 |
| RF | 0.135 | 0.106 | 0.974 | 0.982 | 0.278 | 0.221 | 0.912 | 0.919 |
| M5P | 0.615 | 0.304 | 0.572 | 0.686 | 0.785 | 0.586 | 0.318 | 0.390 |

Considering the GAM model as the best model for predicting vapor pressure deficit in the Birjand, Qaen, Tabas, and Sarayan regions, predictions for vapor pressure deficit for the next 10 years in these regions were made. The results of the model showed an average annual vapor pressure deficit increase of 10 Pas in Sarayan, 9 Pa in Birjand, 7 Pa in Qaen, and 5 Pa in Tabas. Figure (9). Predicted annual average vapor pressure deficit results in the GAM model for regions a- Sarayan, b- Birjand, c- Tabas, and d- Qaen.

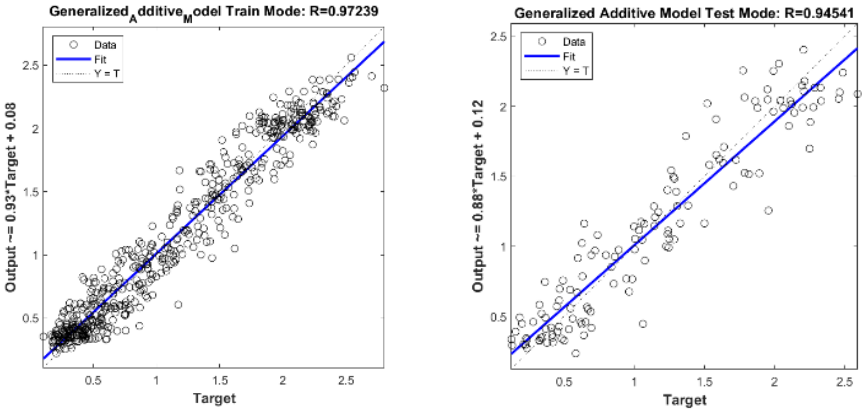


**Fig 8. Graph of the correlation results in the GAM model for the Tabas region**

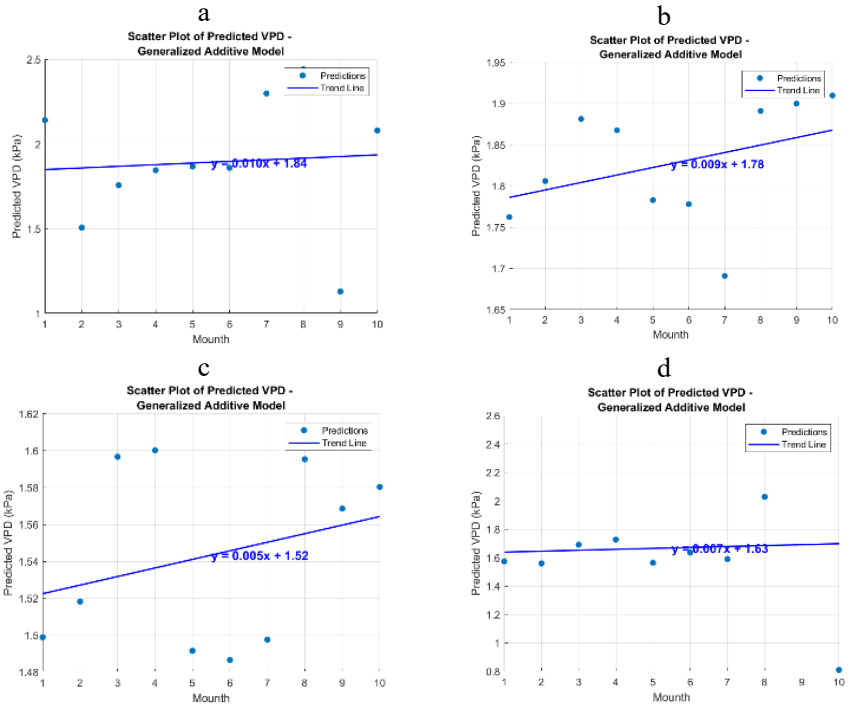


**Fig 9. The results of the annual average vapor pressure deficit prediction in the GAM model**.

## Conclusion

In this study, data extraction process from the JRA-55 database in the regions of Birjand, Qaen, Sarayan, and Tabas were discussed. The data analysis indicated an annual increase in vapor pressure deficit over 65 years in these regions. To predict vapor pressure deficit in these areas, machine learning techniques, including LR, RSS, GAM, M5P, and RF, were used. The results showed that the geographic region did not affect the performance of the machine learning models, as GAM was the best model for estimating vapor pressure deficit across all areas. The GAM model demonstrated high ability to uncover complex, nonlinear relationships between independent and dependent variables,

suggesting that most of the data used in this study had nonlinear relationships. The LR model, on the other hand, is more suitable for simple and linear modeling without deep complexity, which could explain its error weaknesses due to the nonlinear relationships of climatic parameters. The RF model, based on decision trees, uses the combination of multiple trees to reduce variance. Its weakness was its sensitivity to small changes in data (vapor pressure deficit values). The RSS model is useful in analyzing spatial and nonlinear data, but the deep complexity and multiple interactions in the data used in this study posed a limitation for this model. The M5P model, which combines decision tree models and linear regression, is highly effective in uncovering both linear and nonlinear relationships. However, it tends to make errors with highly complex data, leading to the highest error among the models in this study. The GAM model's predictions showed a more significant increase in vapor pressure deficit than the long-term statistics in the studied regions, signaling a potential increase in water consumption in the region's warm and dry climatic conditions. Finally, this study recommends using the GAM model for future research, particularly in smart irrigation systems.

**References**

Ahmar, S., Gill, R.A., Jung, K.H., Faheem, A., Qasim, M.U., Mubeen, M. & Zhou, W. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. *International Journal of Molecular Sciences*. 21(7), 2590. https://doi.org/10.3390/ijms21072590.

Azarm, k., mohebolhojeh, A. & Mirzaie, M. (2022). Climatological Study of Wintertime Blocking Events in the Northern Hemisphere. 20th Iranian Geophysical Conference, Iran.

Blaifi, S.A., Moulahoum, S., Benkercha, R., Taghezouit, B.& Saim, A. (2018). M5P model tree-based fast fuzzy maximum power point tracker. *Solar Energy* 163, 405–424. https://doi.org/10.1016/j.solener.2018.01.071

Bolton, D. (1980). The computation of equivalent potential temperature. *Monthly Weather Review*. 108, 1046–1053. https://doi.org/10.1175/1520-0493(1980)108

Carnicer, J., Barbeta, A., Sperlich, D., Coll, M. & Pe˜nuelas, J. (2013). Contrasting trait syndromes in, angiosperms and conifers are associated with different responses of tree growth to, temperature on a large scale. *Frontiers in Plant Science*. 4, 409. http://doi.org/10.3389/fpls.2013.00409

Dai, G., 2006. Recent climatology, variability, and trends in global surface humidity. *Journal of Climate*. 19(15), 3589–3606.

Dai, A., Zhao, T. & Chen, J. (2018). Climate change and drought: a precipitation and evaporation perspective. *Current Climate Change Reports*, 4 (3), 301–312. http://doi.org/10.1007/s40641-018-0101-6

Devi, M.J., Reddy, V.R. (2018). Effect of temperature under different evaporative demand conditions on maize leaf expansion. *Environmental and Experimental Botany*. 155, 509–517. https://doi.org/10.1016/j.envexpbot.2018.07.024

Ding, J., Yang, T., Zhao, Y., Liu, D., Wang, X., Yao, Y., Peng, S., Wang, T. & Piao, S. (2018). Increasingly important role of atmospheric

aridity on Tibetan alpine grasslands. *Geophysical Research Letters*, 45(6), 2852–2859. https://doi.org/10.1002/2017GL076803

Emami, M., Ahmadi, A., Daccache, A., Nazif, S., Mousavi, S.F. & Karami, H. (2022). County-level irrigation water demand estimation using machine learning: case study of California. *Water*, 14 (12), 1937. https://doi.org/10.3390/w14121937.

Feng, Y., Cui, N., Hao, W., Gao, L.& Gong, D. (2019). Estimation of soil temperature from meteorological data using different machine learning models. *Geoderma* 338, 67–77. https://doi.org/10.1016/j.geoderma.2018.11.044

Ficklin, D.L. & Novick, K.A. (2017). Historical and projected changes in vapor pressure deficit suggest a continental-scale drying of the United States atmosphere. *Journal of Geophysical Research Atmospheres*, 122 (4), 2061–2079. https://doi.org/10.1002/2016JD025855

Fletcher, L., Sinclair, T.R. & Allen Jr., L.H. (2007). Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. *Environmental and Experimental Botany*, 61(2), 145–151. https://doi.org/10.1016/j.envexpbot.2007.05.004

Gong, X.W., Qiu, R.J., Sun, J.S., Ge, J.K., Li, Y.B. & Wang, S.S. (2020). Evapotranspiration and crop coefficient of tomato grown in a solar greenhouse under full and deficit, irrigation. *Agricultural Water Management* 235, 106154. https://doi.org/10.1016/j.agwat.2020.106154

Grossiord, C., Buckley, T.N., Cernusak, L.A., Novick, K.A., Poulter, B., Siegwolf, R.T.W., Sperry, J.S. & McDowell, N.G. (2020). Plant responses to rising vapor pressure deficit. *Wiley New Phytologist*. 226 (6), 1550–1566. https://doi.org/10.1111/nph.16485.

Huntington, T., Cui, X., Mishra, U. & Scown, C.D. (2020). Machine learning to predict biomass sorghum yields under future climate scenarios. *Biofuels Bioprod. Bioref.* 14 (3), 566–577. https://doi.org/10.1002/bbb.2087

Iribarne, J.V., and W.L. Godson. 1981. Atmospheric Thermodynamics. D. Reidel, p. 65. Islam, A.R.M., Talukdar, S., Akhter, S., Eibek, K.U., Rahman, M., Pal, S. & Mosavi, A. (2022). Assessing the impact of the farakka barrage on hydrological alteration in the Padma River with future insight. *Sustainability* 14 (9), 5233. https://doi.org/10.3390/su14095233

Khosravi, K., Daggupati, P., Alami, M.T., Awadh, S.M., Ghareb, M.I. & Panahi, M. (2019). Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: a case study in Iraq. *Computers and Electronics in Agriculture 167*. https://doi.org/10.1016/j.compag.2019.105041

Khosravi, A., Zucchini, M., Giorgi, V., Mancini, A. & Neri, D. (2021). Continuous monitoring of olive fruit growth by automatic extensimeter in response to vapor pressure deficit from pit hardening to harvest. *Horticulturae* 7 (10), 349. https://doi.org/10.3390/horticulturae7100349

Kimball, J.S., Running, S.W. & Nemani, R. (1997). An improved method for estimating surface humidity from daily minimum temperature. *Agricultural and Forest Meteorology*. 85 (1–2), 87–98. http://doi.org/10.1016/S0168-1923(96)02366-0

Konings, A.G., Williams, A.P. & Gentine, P. (2017). Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation. *Nature Geoscience*, 10 (4), 284–288. https://doi.org/10.1038/ngeo2903

Kobayashi, S., Ota, Y., Harda, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. & Gentine, P. (2015). The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan Ser II*.5-48. https://doi.org/10.2151/jmsj.2015-001

Li, Y., Wang, W., Wang, G. & Tan, Q. (2022).

Actual evapotranspiration estimation over the Tuojiang River Basin based on a hybrid CNN-RF model. *Journal of Hydrology*, 610, 127788. https://doi.org/10.1016/j.jhydrol.2022.127788

Liu, W. & Sun, F. (2017). Projecting and attributing future changes of evaporative demand over China in CMIP5 climate models. *Journal of Hydrometeorology*. 18 (4), 977–991. https://doi.org/10.1175/JHM-D-16-0204.1

Maulud, D. & Abdulazeez, A.M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends* 1 (4), 140–147. https://doi.org/10.38094/jastt1457

Mokhtar, A., Elbeltagi, A., Maroufpoor, S., Azad, N., He, H. & Alsafadi, K. (2021). Estimation of the rice water footprint based on machine learning algorithms. *Computers and Electronics in Agriculture*. 191, 106501. https://doi.org/10.1016/j.compag.2021.106501

Mollasharifi, A., Mohebalhojeh, A. R. & Ahmadi-Givi, F. (2019). A study of the impacts of the NAO on the relation between the North Atlantic and Mediterranean storm tracks using the NCEP/NCAR and JRA-55 reanalysis data. *Journal of the Earth and Space Physics*. 45(2),423-440. https://doi.org/ 10.22059/JESPHYS.2019.267521.1007050

Otieno, D., Lindner, S., Muhr, J. & Borken, W. (2012). Sensitivity of peat land herbaceous vegetation to vapor pressure deficit influences net ecosystem CO2 exchange. *Wetlands,* 32 (5), 895–905. https://doi.org/10.1007/s13157-012-0322-8.

Pierce, W., Westerling, A.L. & Oyler, J. (2013). Future humidity trends over the western United States in the CMIP5 global climate models and variable infiltration capacity hydrological modeling system. *Hydrology and Earth System Sciences Discussions*, 9(12), 13651-13691. https://doi.org/10.5194/hessd-17-18332013

Qiu, R.J., Liu, C.W., Cui, N.B., Wu, Y.J., Wang, Z.C. & Li, G. (2019). Evapotranspiration estimation using a modified Priestley-Taylor model in a rice-wheat rotation system. *Agricultural Water Management* 224, 105755. http://doi.org/10.1016/j.agwat.2019.105755.

Qiu, R.J., Katul, G.G. (2020). Maximizing leaf carbon gain in varying saline conditions: an optimization model with dynamic mesophyll conductance. *The Plant Journal*. 101 (3), 543–554. https://doi.org/10.1111/tpj.14553

Rawson, H.M., Begg, J.E., Woodward, R.G. (1977). The effect of atmospheric humidity on photosynthesis, transpiration and water use efficiency of leaves of several plant species. *Planta* 134, 5–10. https://doi.org/10.1007/BF00390086.

Restaino, M., Peterson, D.L. & Littell, J. (2016). Increased water deficit decreases Douglas firgrowth throughout western US forests. *Proceedings of the National Academy of Sciences*. 113(34), 9557–9562.https://doi.org/10.1073/pnas.1602384113

Sangin´es de C´arcer, P., Vitasse, Y., Pe˜nuelas, J., Jassey, V.E.J., Buttler, A. & Signarbieux, C. (2018). Vapor–pressure deficit and extremeclimatic variables limit tree growth. *Global Change Biology*. 24 (3), 1108–1122. https://doi.org/10.1111/gcb.13973.

Sellin, Taneda, H. & Alber, M. (2019). Leaf structural and hydraulic adjustment with respect to air humidity and canopy position in silver birch (Betula pendula). *Journal of Plant Research*, 132, 369–381. https://doi.org/10.1007/s10265-019-01106-w.

Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S. & Chau, K.W.(2020). Predicting standardized stream flow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*. 14 (1), 339–350. https://doi.org/10.1080/19942060.2020.17158

44
Simmons, J., Willett, K.M., Jones, P.D., Thorne, P.W. & Dee, D.P. (2010). Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences fromreanalyses and monthly gridded observational data sets. *Journal of Geophysical Research Atmospheres*. 115(1), https://doi.org/10.1029/2009JD012442
Skurichina, M. & Duin, R.P. (2002). Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis & Applications*. 5 (2), 121–135. https://doi.org/10.1007/s100440200011
Smidte, S.J., Haacker, E.M.K., Kendall, A.D., Deines, J.M., Pei, L., Cotterman, K.A., Li, H., Liu, X., Basso, B. & Hyndman, D.W. (2016). Complex water management in modern agriculture: Trends in the water–energy–food nexus over the High Plains aquifer. *Science of the Total Environment*. 566–567, 988–1001. https://doi.org/10.1016/j.scitotenv.2016.05.127
Sun, X., Lai, P., Wang, S., Song, L., Ma, M. & Han, X. (2022). Monitoring of extreme agricultural drought of the past 20 years in southwest china using gldas soil moisture. remote sensing. 14, 13-23. https://doi.org/10.3390/rs14061323.
Wada, Y., Bierkens, M.F.P. (2014). Sustainability of global water use: past reconstruction and future projections. *Environmental Research Letters*, 9(10), 104003. https://doi.org/10.1088/1748-9326/9/10/104003.
Wang, K., Dickinson, R.E. & Liang, S. (2012). Global atmospheric evaporative demand over land from 1973 to 2008. *Journal of Climate*. 25 (23), 8353–8361. https://doi.org/10.1175/JCLI-D-11-00492.1
Willett, K.M., Jones, P.D., Gillett, N.P. & Thorne, P.W. (2008). Recent changes in surface humidity: development of the HadCRUH dataset. *Journal of Climate*, 21(20), 5364–5383. https://doi.org/10.1175/2008jcli2274.1.
Willett, K.M., Dunn, R.J.H., Thorne, P.W., Bell, S., de Podesta, M., Parker, D.E., Jones, P. D. & Williams Jr., C.N. (2014). HadISDH land surface multi-variable humidity and temperature record for climate monitoring. *Climate of the Past,* 10(6), 1983–2006. https://doi.org/10.5194/cp-10-1983-2014.
Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y. & Yang, S. (2019). Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Science Advances*, *5*(8), eaax1396. https://doi.org/10.1126/sciadv.aax1396.
Zhang, D., Liu, Y., Li, Y., Qin, L., Li, J. & Xu, F. (2018). Reducing the excessive evaporative demand improved the water-use efficiency of greenhouse cucumber by regulating the trade-off between irrigation demand and plant productivity. *HortScience*, 53(12), 1784–1790. https://doi.org/10.21273/HORTSCI13129-18.