Drought Forecasting in Alibori Department in Benin using the Standardized Precipitation Index and Machine Learning Approaches

Rodrigue B. W. VODOUNON[1], Henoc SOUDE[2], Ossénatou MAMADOU[3]
Department of Computer Science[1,2], Department of Physics[3]
Institute of Mathematics and Physics (IMSP), Porto-Novo, Benin

***Abstract*—Drought forecasting provides an early warning for the effective management of water resources to avoid or mitigate drought damage. In this study, the prediction of droughts is carried out in the department of Alibori in Benin republic using the standardized precipitation index (SPI) where two Machine Learning approaches were used to set up the drought prediction models which were Random Forest (RF) and Extreme Gradient Boosting (XGBOOST). The performance of these models was reported using metrics such as: coefficient of determination (R^2), root mean square error (RMSE), mean square error (MSE), and root mean absolute error (MAE). The results revealed that XGBOOST models gave better prediction performance for SPI 3, 6, 12 with coefficients of determination of 0.89, 0.83 and 0.99, respectively. The square root mean square error (RMSE) of the models gives 0.29, 0.40 and 0.07, respectively. This work demonstrated the potential of artificial intelligence approaches in the prediction of droughts in the Republic of Benin.**

***Keywords*—Droughts; forecasting; machine learning; SPI**

I. Introduction

Droughts come after floods and represents one of the most dangerous natural disasters affecting many countries in the world, especially the countries of West Africa which includes Benin republic [1]. Generally caused by a lack control over precipitation, droughts are very complex and difficult to identify because of their ability to occur in any climate, anywhere on earth [2]. Losses caused by droughts between 1900 and 2013 worldwide are estimated at 135 billion dollars. Moreover, in 2018, a total of 281 climatic and geophysical events were recorded with 10,733 deaths and more than 60 million people affected worldwide [3]. According to [4] and [5], climate change would be a factor in the aggravation of drought situations and their frequency in the world. Several vital sectors such as agriculture sectors, food security, hydroelectricity production, human and animal health, industrial activities [2] are seriously affected by the adverse effects of droughts.

Benin republic, like the other West African countries, is not on the sidelines and has suffered the harmful consequences of climate change. According to [3], Benin has recorded more than half a dozen droughts with ever-increasing disasters, from the 1960s to date. One of the most impacted sectors of the economy is the agricultural sector, which contributes 1/3 of the Gross domestic product (GDP) and holds up to nearly 70% of jobs in Benin republic [6]. Despite the significant impacts of drought, Benin republic, like most African countries, has no tool to either monitor or predict it's occurrence.

According to their effects and duration, droughts are classified into three categories: meteorological, agricultural and hydrological droughts. Several indices have been developed to predict and forecast the different types of drought. Gokhan et al. [7] drew up in their bibliometric analysis an exhaustive list of the various indices used in the literature as well as the conditions in which to use them. We can mention: the Standardised Precipitation Evapotranspiration Index (SPEI) dedicated for determining the onset, duration and magnitude of drought conditions; the Standardized Precipitation Index (SPI) to monitor and predict droughts on several timescales including 3, 6 and 12 months corresponding to SPI 3, SPI 6 and SPI 12, respectively. The SPI 3 is dedicated for the meteorological drought prediction, the SPI 6 is used to predict agricultural drought and the SPI is dedicated to the prediction of hydrological drought.

Abhirup et al. [8] also used artificial neural networks to understand the effect of droughts in New South Wales (NSW) using the SPEI drought index. The performance results revealed a coefficient of determination of 0.86. On the other hand, Abhirup et al. [9] used Long Short Term Memory (LSTM) to predict SPEI at 1 and 3 months time scales in New South Wales. The results of this study revealed a coefficient of determination of more than 0.99 for SPEI 1 as for SPEI 3. Poornima et al. [10] on the other hand compared the predictive capacities of SPI and SPEI in China by the LSTM and the model ARIMA at time scales of 1, 3 and 6 months. In all cases, the LSTM outperformed the capabilities of the ARIMA model. Other studies such as [11] have used Random-Forest to predict SPI at 3 and 12 month timescales in the Haive River Basin in China. These models showed good prediction accuracy. Lotfirad et al. [12] also used Random-Forest to predict SPI and SPEI at time scales of 3, 12, 48 months in Iran. The results obtained showed that in temperate climates, such as northern Iran, the correlation coefficients of SPI and SPEI were 0.94, 0.95 and 0.81 at time scales of 3, 12 and 48 months, respectively while it was 0.47, 0.35 and 0.44 in arid and hot climates.

In spite of the multitude of works, Africans are less concerned and among them we can cite the work of Mulualem et al. [13] where they used Artificial Neural Networks to predict the Normalized Precipitation and Evapotranspiration Index (SPEI) for seven stations in the upper Blue Nile (UBN) basin in Ethiopia. The results obtained found that the coefficient of determination and root mean square error of the best model ranged from 0.820 to 0.949 and 0.263 to 0.428, respectively. As in the most studies on drought forecasting, they use satellite data sources without first assessing their reliability; satellite data may not reflect the reality.

The objective of this work is to propose a prediction tool of meteorological, hydrological and agricultural droughts in the Alibori department (BENIN), using the SPI 3 SPI 6 and SPI 12 indicators respectively. The first time we focus on using satellite sources. Indeed, in the concern to have a model close to reality, we carried out an analysis of similarities between satellite data and measured data. We then proposed two models based on the algorithms random forest and XGBoost. Finally, we presented a technique to improve the performance of the model based on the XGBOOST, by combining the different variables present in our dataset.

The rest of the document is organized as follows: Section II presents study area and drought indicators. Then we described our prediction approach in Section III. In Section IV we presented and discussed our results then we concluded with Section V

II. Background

In this section, we first presented our study area followed by the description of the indices chosen for the prediction of drought in our study area. Then, we presented the algorithms used for the implementation of the different models.

A. Study area

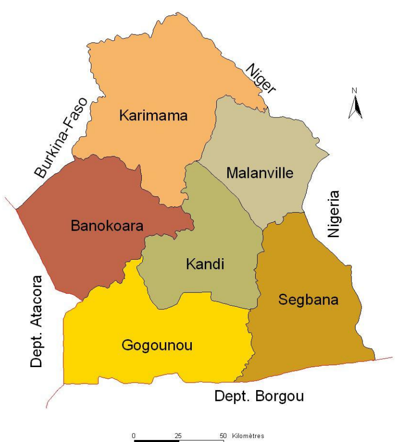

Fig. 1. Cartography of the alibori department (6)

Geographically located between $11°19'$ north latitude and $2°55'$ east longitude, Alibori is a department located in the northeast of the Republic of Benin. It is bordered to the north by the Republic of Niger, to the northwest by the Republic of Burkina Faso, to the east by the Federal Republic of Nigeria, to the west by Atacora and to the south by the department of Borgou [14]. Covering an area of $26\,242\,\mathrm{km}^2$ (nearly 23% of the national territory), Alibori is subdivided into six (06) municipalities as shown in Fig. 1 . These are: Malanville, Karimama, Segbana, Gogounou, Banikoara and Kandi, comprising 41 districts and 229 villages and city districts [14]. According to [6], the department of Alibori is the area most vulnerable and most at risk to droughts in Benin republic. [15] confirms this finding by stating that drought is more recurrent in the northern part of Benin, particularly in the municipalities of Karimama, Malanville and Segbana.

B. Standardized Precipitation Index (SPI)

The standardized precipitation index was developed in 1993 by Mc Kee, N.J. Doesken and J. Kleist of Colorado State University. Its role is not only to determine rainfall deficits for a given period [16] but also to improve the detection of drought episodes at both local and regional scales [17]. Its calculation is based solely on historical rainfall data. To calculate the SPI more efficiently, [2] suggests that it is desirable to have monthly precipitation data over a period of 20 to 60 years. The mathematical expression for calculating the standardized precipitation index is as follows:

$$\mathbf{SPI} = \frac{P_i - P_m}{\sigma} \tag{1}$$

with:

$$\begin{cases} P_i : \text{the total rain of month i;} \\ P_m : \text{the average rainfall of the series on the time scale considered;} \\ \sigma : \text{the standard deviation of the series on the time scale considered.} \end{cases}$$

The seasons can be classified according to the values of the SPI as shown in Table I.

TABLE I. Classification of Seasons based on SPI Values [18]

SPI values	Season classes
2.0 and more	Extremely wet
from 1.5 to 1.99	Very humid
from 1.0 to 1.49	Moderately humid
from -0.99 to 0.99	Near normal
from -1.0 to -1.49	Moderately dry
from -1.5 to -1.99	Very dry
-2 and less	extremely dry

C. Machine Learning Approaches

1) Random forest: Random forests represent a supervised learning method based on decision trees [19]. This method generates multiple trees for prediction. Thus, at the end of the individual predictions, these are averaged to give a final prediction of the model. Each tree being weak, the aggregation of all the weak trees compensates for this shortcoming [20] and creates a more robust model. According to [21], the mean makes a random forest better than a single decision tree, thus improving its accuracy and reducing overfitting. The functioning of Random Forests is illustrated by Fig. 2.

2) Extreme gradient boosting: Extreme Gradient Boosting, also called XGBOOST, is an optimized implementation of the gradient boosting tree algorithm [22]. It is often the winning algorithm in competitions on Kaggle [23]. Unlike the Random Forest, the XGBOOST works sequentially. In other words, the algorithm is based on previous predictions to improve the results of other estimators. Thus, we rely on the predictions of the "weak learners" to build a "strong learner" [23]. XGBoost minimizes a regularized objective function (L1 and L2) that combines a convex loss function (based on the difference between predicted and target outputs) and a penalty term for model complexity. The training proceeds iteratively, adding new trees that predict the residuals or errors of the previous trees which are then combined with the previous trees to make the final prediction [24]. Fig. 3 further illustrates how the XGBOOST algorithm works.

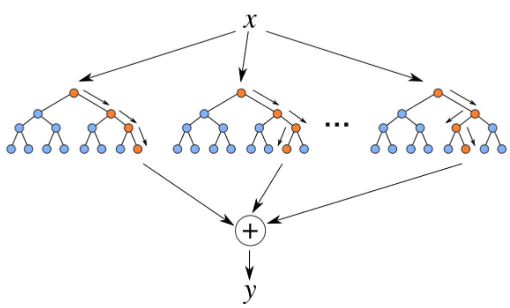

Fig. 2. Structure of a random forest (Random forest) [25]

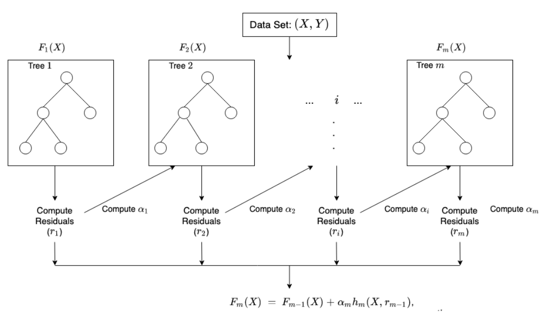

Fig. 3. How XGBOOST works [24]

3) Performance assessment: The mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and coefficient of determination (R^2) are five typical performance indicators used to assess the models performance. The following mathematical expressions illustrate how to calculate these metrics.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Actual - Predicted| \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(Actual - Predicted\right)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(Actual - Predicted\right)^2} \tag{4}$$

$$R^2 = 1 - \frac{Squared\ sum\ error\ of\ regression\ line}{Squared\ sum\ error\ of\ mean\ line} \tag{5}$$

with n : The number of samples.

III. Prediction Approach

A. Methodology

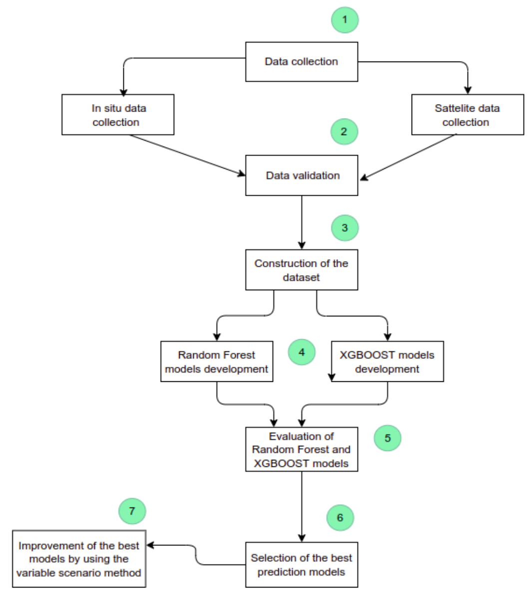

Fig. 4. Methodology of the study conducted

Our solution consists in predicting droughts in the department of Alibori using the drought index SPI$\{3, 6, 12\}$. The prediction is performed using machine learning algorithms, as shown in Fig. 4. First, we collected satellite data covering our entire study area over a period of 30years. Then, we carried out the validation of these data using observation data from a station of the AMMA-CATCH program (Multidisciplinary Analysis of the African Monsoon-Coupling of the Tropical Atmosphere and the Hydrological Cycle) [26]. After that, we built our dataset and our models based on the Random Forest and XGBOOST algorithms. Finally we selected the best of the models and tried to improve them using a variable scenario method described in [27]. It consists of identifying the most useful variables when predicting indicators. First, we determined the correlation between the variables using the Pearson correlation matrix, then we defined the different combinations of variables, starting with the most correlated variables. After construction of the variable scenarios, the models were retrained and re-evaluated on the basis of each variable scenario.

B. Data

The use of Machine Learning in the prediction of drought indicators requires data over a period of at least 30 years. Unfortunately, we do not have any data for such a period. So as an alternative, we therefore opted for the use of satellite data from NASA's Power Data Access Viewer platform(https://power.larc.nasa.gov/) which has data over a period of 36 years (1985-2021), covering the entirety of our study area and containing all the variables allowing the calculation of the SPI: Temperature, total precipitation, wind speed, wind direction, surface pressure, relative humidity, cloud amount, soil moisture, Top-Of-Atmosphere Shortwave Downward Irradiance and all sky and All Sky Surface Shortwave Downward Irradiance. In addition to this, its use is reported in several works [28], [29], [30] to quote only those.

1) Validation: It consists in evaluating the reliability of satellite data before using them. To this end, we considered the Bellefoungou station (Benin) [31] for which we already had in situ data over (08) years and we collected satellite data corresponding to this region and the same period. We then used the Taylor diagram and the Pearson correlation matrix to study the similarity between in situ and satellite data.The results of this validation phase are presented in Fig. 5 and 6. As shown in Fig. 5, there is a strong correlation (0.91) between observation data and satellite data. Fig. 6 also reveals that there is a strong correlation (0.91) between observation data (represented by the white circle on the x-axis) and satellite data (represented by the blue dot). This figure also shows that the difference in the standard deviation between the observation data and the satellite data is not very high (approximately equal to 0.3) and the error between these data is above 0.5. From all the above, it can be seen that there is a strong similarity between observational data and satellite data.

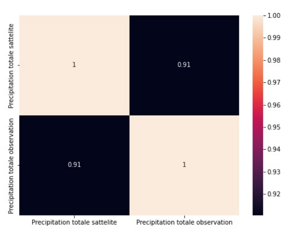

Fig. 5. Correlation matrix of the total precipitation variable

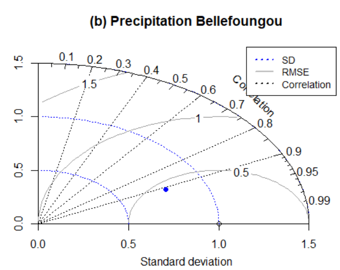

Fig. 6. Taylor diagram of the total precipitation variable

2) Construction of the dataset: By exploiting the precipitation data, the calculation of the SPIs at the time scales of 3, 6 and 12 months corresponding respectively to SPI 3, SPI 6 and SPI 12 has been implemented according to the formula 1 presented in the Section II.

After calculating the different SPIs, observations with missing data were removed. Once the datasets were purged of missing values, the data from the (06) municipalities were merged together in order to have a single dataset (Table II).

C. Model Building

We used the approaches described in the subsection II-C to implement models for the prediction of SPI 3 (meteorological drought), SPI 6 (agricultural drought), SPI 12 (hydrological drought). The different models have been implemented using the python programming language as well as different data science libraries such as: NumPy, Pandas, Scikit-learn, Matplotlib and Seaborn.

TABLE II. DETAILS OF THE SATELLITE DATASET

VARIABLES	UNITS	TIME SCALE
Surface Pressure	KPa	Monthly
Total Precipitation	mm	Monthly
Air Temperature at 2 meters	C	Monthly
Relative Humidity at 2 meters	%	Monthly
Wind Speed at 2 meters	m/s	Monthly
Wind Direction at 2 meters	Degree (°)	Monthly
Soil Moisture	-	Monthly
Cloud Amount	%	Monthly
Top-of-Atmosphere Shortwave Downward Irradiance	$MJ/m^2/day$	Monthly
All Sky Surface Shortwave Downward Irradiance	$MJ/m^2/day$	Monthly

In order to effectively train and validate the developed models, the dataset was divided into two parts: the training data (77%) and the test data (23%). We also used GridsearchCV to obtain the best hyperparameters for the models (see Tables III and IV).

TABLE III. RANDOM FOREST MODEL HYPERPARAMETER VALUES

Index	Hyperparameters	Values
SPI 3	n_estimators	70
	random_state	15
SPI 6	n_estimators	100
	random_state	9
SPI 12	n_estimators	100
	random_state	9

TABLE IV. XGBOOST MODEL HYPERPARAMETER VALUES

Hyperparameters	Values
Subsample	0,7
colsample_bytree	1
eta	0,1
max_depth	10
n_estimators	100
booster	gbtree
verbosity	0

IV. RESULTS AND DISCUSSION

We first evaluated the different models based respectively on Random-Forest and XGBOOST to identify the best models and then we improved them using variable scenario.

A. Random Forest Evaluation

Considering all the variables whose data were collected, prediction models for SPI 3, 6 and 12 were implemented and evaluated using the performance metrics presented in Section II. As it can be seen in the Table V, the models gave SPI 3, SPI 6 and SPI 12, respectively for performance R^2 of 0.832 (83%), 0.742 (74%) and 0.992 (99%).

After evaluating the models, it can be seen that the SPI 3 prediction model gives errors RMSE = 0.379, MAE = 0.276 and MSE = 0.143 respectively. As for the prediction model of SPI 6, we had the following errors: RMSE=0.479, MAE=0.344 and MSE=0.230. The SPI 12 prediction model gave the lowest errors with RMSE= 0.082, MAE = 0.041

and MSE = 0.006 values, respectively. This would mean that the variables considered are more relevant for predicting hydrological drought (SPI 12).

B. XGBOOST Evaluation

The Table VI presents the evaluation results of the prediction models of SPI 3, 6 and 12 based on the XGBOOST algorithm. This table shows good performance for all models with coefficient of determination R^2 respectively equal to: 88% for the SPI 3 prediction model, 81% for the SPI 6 prediction model and 99% for that of the SPI 12.

As for the errors made by the models, the results of the evaluation revealed that the prediction model of the SPI 3 gave the values RMSE = 0.317, MAE = 0.221 and MSE = 0.100, respectively. The SPI 6 prediction model gave the following results: RMSE = 0.423, MAE = 0.301 and MSE = 0.179. The SPI 12 prediction model provides the best performance and minimizes all errors the most with errors RMSE = 0.076, MAE = 0.041 and MSE = 0.005.

By comparing the performances of the Random Forest models and those of XGBOOST, we find that the XGBOOST algorithm is much more efficient in the case of the present study.

TABLE V. Random Forest Model Evaluation Result

Index	R^2	RMSE	MAE	MSE
SPI 3	0.832	0.379	0.276	0.143
SPI 6	0.742	0.479	0.344	0.230
SPI 12	0.992	0.082	0.041	0.006

TABLE VI. XGBOOST model Evaluation Result

Index	R^2	RMSE	MAE	MSE
SPI 3	0.882	0.317	0.221	0.100
SPI 6	0.817	0.423	0.301	0.179
SPI 12	0.993	0.076	0.041	0.005

C. Variable Scénario

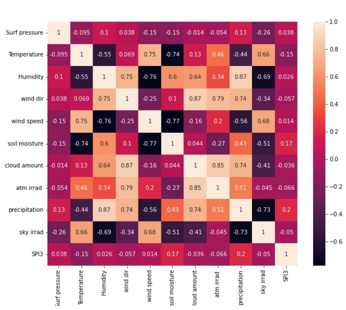

Fig. 7. Correlation matrix between input variables and SPI 3

According to the Fig. 7, the highest correlation is equal to 0.2 and this corresponds to the correlation between the total precipitation and the SPI 3. We also note a correlation of 0.17 between the humidity of the ground and SPI 3. Then follow weaker positive correlations respectively equal to 0.038, 0.026 and 0.014 corresponding respectively to the correlations between SPI 3 and surface pressure, relative humidity and wind speed. In addition to the positive correlations, some variables have a negative correlation with the SPI 3. These are the amount of cloud (-0.036), All Sky Surface Shortwave Downward Irradiance (-0.05), wind direction (-0.057), Top-Of-Atmosphere Shortwave Downward Irradiance (-0.066), Air temperature (-0.15). This correlation matrix made it possible to identify the variables that have a strong correlation with the SPI 3. This is the variable: total precipitation. It can be deduced that the total precipitation represents a relevant variable for the prediction of SPI 3 (meteorological drought). The same approach was adopted for the cases of SPI 6 (agricultural drought) and 12 (hydrological drought).

The Tables VII, VIII and IX respectively present the different variable scenarios according to the type of drought. Indeed, these scenarios are only combinations of the variables obtained from the Pearson correlation matrices. For the experimentation phase, they were successively used as input parameters to the XGBOOST models. These matrices made it possible to construct 11 scenarios per type of drought. The presence of a (✓) symbol in the cell of a column means that the variable concerned is part of the scenario corresponding to this column. For illustrative purposes, scenario 1 of the Table VII consists only of the Total precipitation variable. Scenario 3 consists of the variables: soil moisture, total precipitation and surface pressure.

TABLE VII. Scenarios of Variables for the Prediction of SPI 3 (Meteorological Drought)

Variables \ Scenarios	1	2	3	4	5	6	7	8	9	10	11
Soil moisture		✓	✓	✓	✓	✓	✓	✓	✓		
Total recipitation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Relative humidity				✓	✓	✓	✓	✓	✓	✓	✓
Wind direction								✓	✓	✓	✓
Surface pressure			✓	✓	✓	✓	✓	✓	✓	✓	✓
Cloud amount						✓	✓	✓	✓	✓	✓
Top-Of-Atmosphere Shortwave Downward Irradiance									✓	✓	✓
Wind speed					✓	✓	✓	✓	✓	✓	
All Sky Surface Shortwave Downward Irradiance							✓	✓	✓	✓	✓
Air temperature										✓	

TABLE VIII. Scenarios of Variables for the Prediction of SPI 6 (Agricultural Drought)

Variables \ Scenarios	1	2	3	4	5	6	7	8	9	10	11
Soil moisture	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Total precipitation		✓	✓	✓	✓	✓	✓	✓	✓	✓	
Relative humidity			✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind direction				✓	✓	✓	✓	✓	✓	✓	✓
Surface pressure					✓	✓	✓	✓	✓	✓	✓
Cloud amount						✓	✓	✓	✓	✓	✓
Top-Of-Atmosphere Shortwave Downward Irradiance							✓	✓	✓	✓	✓
Wind speed								✓	✓		
All Sky Surface Shortwave Downward Irradiance									✓	✓	
Air Temperature									✓		

Once the variable scenarios were established, the XGBOOST models were trained and validated by considering each variable scenario. Then, these models were evaluated using previously used performance metrics to identify not only the most accurate model but also the scenarios that provide better prediction performance. The performance results of the models according to each scenario of variables for each SPI

TABLE IX. SCENARIOS OF VARIABLES FOR THE PREDICTION OF SPI 12 (HYDROLOGICAL DROUGHT)

Variables \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
Soil moisture			✓	✓	✓	✓	✓	✓	✓	✓	✓
Total precipitation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Relative humidity		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind direction				✓	✓	✓	✓	✓	✓	✓	✓
Surface pressure						✓	✓	✓	✓	✓	✓
Cloud amount					✓	✓	✓	✓	✓	✓	✓
Top-Of-Atmosphere Shortwave Downward Irradiance							✓	✓	✓	✓	✓
Wind speed									✓	✓	✓
All Sky Surface Shortwave Downward Irradiance								✓	✓		
Air temperature										✓	

(SPI 3, SPI 6 and SPI 12) are presented in the Tables X, XI, XII. The Table X, reveals that scenario 9 is the one that provides the best results with a determination coefficient R^2 equal to 0.898 (89%), an RMSE of 0.29, an MAE of 0.211 and an MSE of 0.087. This table also reveals two important elements. First, it can be noticed that the more one adds a variable to a scenario, the more the precision of the model increases (even the variables having a weak correlation with the SPI 3) with the exception of the variables soil humidity and air temperature . Secondly, when we use as input parameters, scenario 2 which contains only the variables most correlated to SPI 3, the performance of the model deteriorates completely with a negative R^2 (-0.241). This actually means that the fact that the variables are the most correlated to the SPI 3, does not necessarily mean that they are the most essential for the prediction of the SPI 3. Moreover, a variable having a negative correlation (quantity of cloud for example) with SPI 3 can increase the predictive performance of the model. As for the Table XI, it shows that scenario 9 is still the one that provides the best results for the prediction of SPI 6 with a coefficient of determination R^2 equals to 0.830 (83%), an RMSE of 0.408, an MAE of 0.289 and an MSE of 0.167. This table also shows two other key pieces of information. We can also notice that the more we add a variable to a scenario, the precision of the model increases (even the variables having a negative correlation with the SPI 6) with the exception this time of the temperature variable of the air whose addition lowered accuracy and increased errors. Although the total precipitation and soil moisture variables constituting scenario 2 are the most correlated with SPI 6, the performance of the model was only 0.017 for this scenario. This actually shows that the fact that the variables are the most correlated to the SPI 6, does not necessarily mean that they are the most essential for the prediction of the SPI 6. Moreover, a variable having a negative correlation (wind speed per example) with SPI 6 can increase the predictive performance of the model.

The performance results of the XGBOOST model for the prediction of SPI 12 are presented in the Table XII. Unlike SPI 3 and SPI 6, this time it is scenario 8 that gives the best results with a determination coefficient R^2 equal to 0.994 (99%), an RMSE of 0.073, an MAE of 0.039 and an MSE of 0.005. Several other relevant information can be extracted from this table. Firstly, as in the case of SPI 3 and SPI 6, it can also be seen that the more a variable is added to a scenario, the more the precision of the model increases (even the variables having a negative correlation with the SPI 12) with the exception this time variables: Top-Of-Atmosphere Shortwave Downward Irradiance, wind speed, air temperature, the addition of which decreased accuracy and increased errors. Secondly, considering the three variables most correlated to SPI 12 which are: total precipitation, soil moisture and relative humidity forming scenario 3, the performances are already starting to be good (R^2 = 86%) . This is not the case for the prediction of SPI 3 and 6. This actually shows that the total precipitation, soil moisture and relative humidity have a strong influence on the prediction of SPI 12.Having a negative correlation with the SPI 12, only the addition of the variable All Sky Surface Shortwave Downward Irradiance slightly increased the performance of the model 0.994 (99%).

The variation of predicted SPI values and calculated SPI values is illustrated in Fig. 8, 9 and 10. As it can be seen, the predicted SPI (in blue) have a similar variation as the calculated SPIs 3 (in red). By considering these three figures, it can be noticed that the predicted SPI 12 is very close to the calculated SPI 12. This justifies the 99% accuracy of the SPI 12 prediction model. It can therefore be said that the XGBOOST is better suited to the prediction of hydrological drought in the department of Alibori.

TABLE X. PERFORMANCE RESULTS OF THE SPI 3 PREDICTION MODEL BASED ON THE XGBOOST ALGORITHM

Métrique \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
R^2	-0.0005	-0.241	0.200	0.449	0.526	0.636	0.646	0.665	0.898	0.884	0.874
RMSE	0.926	1.032	0.828	0.687	0.637	0.558	0,551	0.536	0.295	0.314	0.328
MAE	0.712	0.722	0.567	0.463	0.441	0.393	0.391	0.374	0.211	0.226	0.225
MSE	0.859	1.065	0.686	0.472	0.406	0.311	0.303	0.287	0.087	0.176	0.107

TABLE XI. PERFORMANCE RESULTS OF THE SPI 6 PREDICTION MODEL BASED ON THE XGBOOST ALGORITHM

Métrique \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
R^2	0.015	0.017	0.317	0.458	0.519	0.572	0.784	0.813	0.830	0.810	0.327
RMSE	0.985	0.984	0.189	0.730	0.688	0.649	0.460	0.428	0.408	0.432	0.814
MAE	0.792	0.722	0.136	0.532	0.480	0.452	0.315	0.294	0.289	0.304	0.656
MSE	0.971	0.969	0.035	0.534	0.474	0.421	0.212	0.183	0.167	0.187	0.663

TABLE XII. PERFORMANCE RESULTS OF THE SPI 12 PREDICTION MODEL BASED ON THE XGBOOST ALGORITHM

Métrique \ Scénarios	1	2	3	4	5	6	7	8	9	10	11
R^2	0.663	0.780	0.860	0.896	0.913	0.994	0.993	0.994	0.993	0.993	0.993
RMSE	0.559	0.451	0.360	0.310	0.283	0.07	0.075	0.073	0.077	0.075	0.076
MAE	0.398	0.300	0.197	0.173	0.157	0.044	0.040	0.039	0.042	0.041	0.041
MSE	0.313	0.204	0.129	0.096	0.080	0.005	0.005	0.005	0.005	0.005	0.005

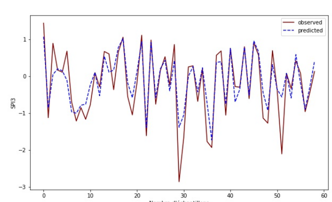

Fig. 8. Comparison of predicted SPI 3 and target SPI 3

V. CONCLUSION

This study consisted in designing prediction models of SPI 3 (meteorological drought), SPI 6 (agricultural drought) and SPI 12 (hydrological drought) in the department of Alibori based respectively on the Random forest and Extreme Gradient Boosting (XGBOOST) algorithms. Performance results

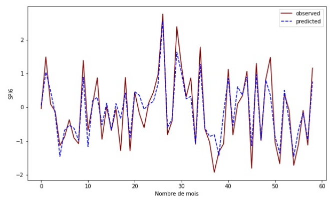

Fig. 9. Comparison of predicted SPI 6 and target SPI 6

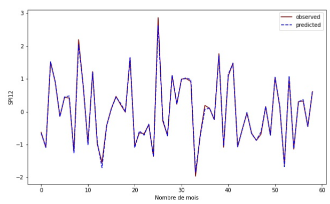

Fig. 10. Comparison of predicted SPI 12 and target SPI 12

indicated that XGBOOST performs better than Random forest in predicting SPI 3, SPI 6 and SPI 12 with respectively the coefficients of determination R^2 équal to 88%, 81% and 99%. To further improve the performance of XGBOOST models, 11 variable scenarios were designed. These scenarios were then used respectively to train, validate and evaluate the models in order to identify the scenarios that allow good prediction performance. At the end of this improvement step, the best performances obtained are respectively 89% for SPI 3, 83% for SPI 6 and 99% for SPI 12. The use of this technique based on the scenarios of variables allowed to deduce that the air temperature was not relevant for the prediction of the SPI 3. Similarly the variables air temperature and wind speed were not essential for the prediction of SPI 12. Future prospects consist of using other drought indices like SPEI in order to compare their results with those obtained in the present study in order to see the drought index that lends itself to drought prediction in Benin in general and in the Alibori in particular).

VI. Perspectives

In the future works, historical data will be used to predict future droughts, which would be useful to meteorological decision makers in developing drought mitigation measures and actions.

References

[1] D. Tsegai, M. Medel, P. Augenstein, and Z. Huang, "La sécheresse en chiffres," Convention des Nations Unies sur la Lutte contre la Désertification (CNULCD), Côte d'Ivoire, Abidjan, Rapport de la 15ème Conférence des Parties (COP 15), Mai 2022.

[2] O. météorologique mondiale (OMM) et Partenariat mondial pour l'eau (GWP), *Manuel des indicateurs et indices de sécheresse*, ser. Programme de gestion intégrée des sécheresses, Integrated Drought Management Programme Tools and Guidelines Series 2, Genève, 2016.

[3] A. N. de Protection Civile (ANPC), "Plan national secheresse 2019-2024," Bénin, Tech. Rep., Janvier 2019.

[4] réseau action climat france. (2019) Le réchauffement va aggraver le risque de sécheresses dans certaines régions. [Online]. Available: https://reseauactionclimat.org/rechauffement-climatique-aggrave-secheresse/

[5] M. France. (2020) Sécheresses et changement climatique. [Online]. Available: https://reseauactionclimat.org/rechauffement-climatique-aggrave-secheresse/

[6] F. de recherche CIMA, C. W. C. pour l'OMM, and A. N. au Bénin, "Projet intégrer la gestion des inondations et des sécheresses et de l'alerte précoce pour l'adaptation au changement climatique dans le bassin de la volta," World Meteorological Organization, Volta Basin Authority, Global Water Patnership West Africa, Rapport de consultation national au Bénin, Janvier 2021.

[7] G. Yildirim, A. Rahman, and V. P. Singh, "A bibliometric analysis of drought indices, risk, and forecast as components of drought early warning systems," *Water*, vol. 14, no. 2, p. 253, 2022.

[8] A. Dikshit, B. Pradhan, and A. M. Alamri, "Temporal hydrological drought index forecasting for new south wales, australia using machine learning approaches," *Atmosphere*, vol. 11, no. 6, p. 585, 2020.

[9] A. Dikshit, B. Pradhan, and A. Huete, "An improved spei drought forecasting approach using the long short-term memory neural network," *Journal of environmental management*, vol. 283, p. 111979, 2021.

[10] S. Poornima and M. Pushpalatha, "Drought prediction based on spi and spei with varying timescales using lstm recurrent neural network," *Soft Computing*, vol. 23, no. 18, pp. 8399–8412, 2019.

[11] J. Chen, M. Li, and W. Wang, "Statistical uncertainty estimation using random forests and its application to drought forecast," *Mathematical Problems in Engineering*, vol. 2012, 2012.

[12] M. Lotfirad, H. Esmaeili-Gisavandani, and A. Adib, "Drought monitoring and prediction using spi, spei, and random forest model in various climates of iran," *Journal of Water and Climate Change*, vol. 13, no. 2, pp. 383–406, 2022.

[13] G. M. Mulualem and Y.-A. Liou, "Application of artificial neural networks in forecasting a standardized precipitation evapotranspiration index for the upper blue nile basin," *Water*, vol. 12, no. 3, p. 643, 2020.

[14] B. Alexandre, S. and A. Djabar, Dine, "Cahier des villages et quartiers de ville du departement de l'alibori," INSTITUT NATIONAL DE LA STATISTIQUE ET DE L'ANALYSE ECONOMIQUE (INSAE), Benin, RECENSEMENT GENERAL DE LA POPULATION ET DE L'HABITATION (RGPH-4), Août 2016.

[15] "Catastrophes naturelles et objectifs du millÉnaire pour le dÉveloppement au bÉnin: Les propositions alternatives de social watch bÉnin," SOCIAL WATCH BÉNIN, Benin, Rapport Alternatifs annuel des Organisations de société civile sur les OMD, Décembre 2010.

[16] R. Mehdaoui, E. MILI, and A. SEGHIR, "Caractérisation à l'aide du spi de la sécheresse climatique dans le bassin versant de ziz (sud-est, maroc)," *Eur Sci J*, 2018.

[17] A. Mokhtar, M. Jalali, H. He, N. Al-Ansari, A. Elbeltagi, K. Alsafadi, H. G. Abdo, S. S. Sammen, Y. Gyasi-Agyei, and J. Rodrigo-Comino, "Estimation of spei meteorological drought using machine learning algorithms," *IEEE Access*, vol. 9, pp. 65 503–65 523, 2021.

[18] S. M., H. M., and W. D., *Guide d'utilisation de l'indice de précipitations normalisé*, Genève, 2012, no. OMM-N° 1090.

[19] E. S. R. I. (Esri). (2020) Fonctionnement de l'outil régression et classification de forêt aléatoire. [Online]. Available: https://pro.arcgis.com/fr/pro-app/latest/tool-reference/geoai/how-random-forest-classification-and-regression-works.htm

[20] M. FERECATU. Apprentissage, réseaux de neurones et modèles graphiques. Paris, France. [Online]. Available: http://cedric.cnam.fr/vertigo/cours/ml2/coursForetsAleatoires.html

[21] D. Mwiti. Random forest regression: When does it fail and why? [Online]. Available: https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why

[22] B. Cayla. La star des algorithmes de ml : Xgboost. [Online]. Available: https://datacorner.fr/xgboost/

[23] Xgboost, le grand gagnant des compétitions. [Online]. Available: https://datascientest.com/xgboost-grand-gagnant-des-competitions-machine-learning-algorithme

[24] How xgboost works. [Online]. Available: https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html

[25] Random forest regression. [Online]. Available: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

[26] S. Galle, M. Grippa, C. Peugeot, I. B. Moussa, B. Cappelaere, J. Demarty, E. Mougin, G. Panthou, P. Adjomayi, E. Agbossou *et al.*, “Amma-catch, a critical zone observatory in west africa monitoring a region in transition,” *Vadose Zone Journal*, vol. 17, no. 1, pp. 1–24, 2018.

[27] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, “A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index,” *Journal of climate*, vol. 23, no. 7, pp. 1696–1718, 2010.

[28] A. M. Abdel-Hameed, M. E.-S. Abuarab, N. Al-Ansari, H. S. Mehawed, M. A. Kassem, H. He, Y. Gyasi-Agyei, and A. Mokhtar, “Winter potato water footprint response to climate change in egypt,” *Atmosphere*, vol. 13, no. 7, p. 1052, 2022.

[29] O. A. Marzouk, “Assessment of global warming in al buraimi, sultanate of oman based on statistical analysis of nasa power data over 39 years, and testing the reliability of nasa power against meteorological measurements,” *Heliyon*, vol. 7, no. 3, p. e06625, 2021.

[30] G. C. Rodrigues and R. P. Braga, “Evaluation of nasa power reanalysis products to estimate daily weather variables in a hot summer mediterranean climate,” *Agronomy*, vol. 11, no. 6, p. 1207, 2021.

[31] M. Hounsinou, O. Mamadou, M. Wudba, B. Kounouhewa, and J.-M. Cohard, “Integral turbulence characteristics over a clear woodland forest in northern benin (west africa),” *Atmospheric Research*, vol. 268, p. 105985, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016980952100541X