ABSTRACT
Evapotranspiration is a major component of hydrologic cycle and its accurate estimation is essential for agricultural water management. The Penman-Monteith (PM) equation is the universal accurate method for estimating reference evapotranspiration \( \text{ET}_{\text{ref}} \). Its drawback is the large climatic data required which are unavailable in many African semiarid regions such as Burkina Faso. The Hargreaves (HRG) conventional method which requires few data is still used despite of its non-universal accuracy due to the model inability to capture the effect of some important climatic parameters. Therefore, this study assesses the ability of an artificial neural network (ANN) for \( \text{ET}_{\text{ref}} \) modeling in Dédogou region, located in the Soudano-Sahelian zone of Burkina Faso. This study employs ANN and HRG models in order to evaluate their performance by comparing with the true PM. From the statistical comparison results, ANN showed a good performance than HRG which overestimated \( \text{ET}_{\text{ref}} \) for the observed condition. Furthermore, wind speed has been found as an important factor in ANN accuracy improvement. Using ANN under semiarid zone climatic condition of Africa for modeling \( \text{ET}_{\text{ref}} \) is highly superior to the other conventional method.

KEY WORDS
water management, evapotranspiration, conventional model, neural network, performance

1. Introduction
The hydrological models currently used in several studies around the world including the modeling of climate change [1], yield function [2], water budget [3], irrigation scheduling [4] and rainfall index [5] all require the evapotranspiration. [6] stated on the importance of evapotranspiration as a major component of agricultural water management due to the persistence of the water resources rarity and growing of the world population.

Water management is vital for a country such as Burkina Faso located in a dry tropical climate of Western Africa where crops are constantly under the influence of low rainfall and high temperature. Hence, efficient use of water becomes extremely a major challenge for Burkinabe farmers. Efficient water management requires an accurate reference evapotranspiration \( \text{ET}_{\text{ref}} \) which can be derived from the meteorological variables. [7] indicated that \( \text{ET}_{\text{ref}} \) is a major component in terrestrial water balance and net primary productivity models, but it is difficult to measure and predict. The most common Penman-Monteith (PM) equation has been recommended by the Food and Agriculture Organization of the United Nations as universally accurate method for estimating \( \text{ET}_{\text{ref}} \). According to [8], PM is now widely used by agronomists, irrigation engineers and other scientists in the field-practice and research. However, the large number of weather input data required by the PM equation is often difficult and expensive to obtain for practical applications in many countries of the world.

PM equation computes \( \text{ET}_{\text{ref}} \) using the minimum and maximum air temperature, relative humidity, wind velocity and sunshine hour data. The enormous data required by the equation has been indicated as constraining for irrigation information computerization in Burkina Faso [2, 9]. Hence, the conventional approach such as Hargreaves equation is still used for \( \text{ET}_{\text{ref}} \) estimation in many areas because of the advantage of its simplicity requiring only air temperature data. According to [10], although the conventional methods use few weather data, they do not have a universal suitability. [8] reported that they are often unable to capture the effect of some important climate parameters which may affect \( \text{ET}_{\text{ref}} \). According to [11], the conventional approaches miss the opportunity to incorporate some weather information.

In past decades, scientists paid considerable attention for another approach which is the artificial neural network (ANN) applied in diverse fields of hydrology engineering forecasting and modeling. ANN application in hydrology includes rainfall-runoff modeling [12]; suspended sediment forecasting [13] and evapotranspiration estimation [14]. ANN was potentially used to model \( \text{ET}_{\text{ref}} \) as a function of climatic variables. [15] and [16] in their \( \text{ET}_{\text{ref}} \) estimation simplified the neural network inputs data to air temperature, extraterrestrial solar radiation and daily light hours. Recently, [17] used similar input sets but without the daily light data for estimating successfully \( \text{ET}_{\text{ref}} \) in Iran.
Therefore, in this study, the minimum and maximum air temperature (°C), and extraterrestrial solar radiation (mm day\(^{-1}\)) were adopted as the input variables of the neural network. The present study employs the Generalized Regression Neural Network (GRNN) algorithm and Hargreaves (HRG) conventional method for ET\(_{ref}\) modeling in Dédougou region located in Burkina Faso an African semiarid country where climatic data have been collected from 1996 to 2006. The present paper’s goal is to assess the performances of the GRNN and Hargreaves (HRG) models for estimating ET\(_{ref}\) by comparison with the reference PM.

2. Methodology

2.1. Climate Dataset

The decadal climatic data used for this study were recorded at the meteorological station of Dédougou from 1996 to 2006. Dédougou is located in the Soudano-Sahelian zone at 300 m altitude, 12°47’N latitude and 3°48’W longitude (Figure 1). The area has a semiarid climate with 809.41 mm annual average of rainfall. In the region, 80.51% of rainfall occurs between June and September with a peak in August (227.14 mm). The region has two seasons; a rainy season (short) from May to September, and a dry season (long) from October to April. The annual average air temperature are ranged from 22.16 to 38.82°C and 18.25 to 40.28°C in rainy and dry season, respectively. The relative humidity means are 33% in dry season and 69% in rainy season with an annual average of 48%. Wind velocity recorded at 2m above the ground has an annual average of 147 km day\(^{-1}\). The wind speed annual averages in rainy and dry season are 141 to 150 km day\(^{-1}\), respectively. The data collected for this study were comprised of maximum and minimum air temperature (°C), precipitation (mm), relative humidity (%), wind velocity (km day\(^{-1}\)) and sunshine duration (hours).

2.2. Reference Evapotranspiration (ET\(_{ref}\)) Estimation Methods

-Penman-Monteith (PM) equation used in this study is given as the following:

\[
ET_{ref} = \frac{0.408 \Delta (R_n - G) + \frac{900}{\Delta + 273} u_2 (e_s - e_a)}{\Delta + y (l_0 + 0.34u_2)}
\]

Where ET\(_{ref}\) is the reference evapotranspiration (mm day\(^{-1}\)); \(R_n\) the net radiation at the crop surface (MJ m\(^{-2}\) day\(^{-1}\)); \(G\) the soil heat flux density (MJ m\(^{-2}\) day\(^{-1}\)); \(T\) the mean daily air temperature at 2 m height (°C); \(u_2\) the wind speed at 2 m height (m s\(^{-1}\)); \(e_s\) the saturation vapor pressure (kPa); \(e_a\) the actual vapor pressure (kPa); \(e_s - e_a\) the saturation vapor pressure deficit (kPa); \(\Delta\) the slope vapor pressure curve (kPa °C\(^{-1}\)); and \(\gamma\) the psychrometric constant (kPa °C\(^{-1}\)).

-Hargreaves (HRG) equation is used for ET\(_{ref}\) estimation when the solar radiation, relative humidity and wind speed data are missing. This method estimates ET\(_{ref}\) using only the maximum and minimum air temperature with the following equation:

\[
ET_{ref} = C_o (T_{max} - T_{min})^{0.5} (T_{mean} + 17.8) R_a
\]

Where ET\(_{ref}\) is the reference evapotranspiration (mm day\(^{-1}\)); \(T_{max}\) and \(T_{min}\) are the maximum and minimum temperature (°C); \(T_{mean}\) is the mean temperature (°C); \(R_a\) is the extraterrestrial radiation (mm day\(^{-1}\)); and \(C_o\) is the conversion coefficient (°C) (\(C_o = 0.0023\)).

2.3. Artificial Neural Network

The artificial neural network (ANN) is a mathematical model so called black-box in which the network takes only the input and output data for learning the complex relationship, and then produces approximately a new output from the input data. The present study employs the generalized regression neural networks (GRNN) algorithm for ET\(_{ref}\) modeling. GRNN is preferred instead of the multilayer networks due to its performances [18], it does not also require an iterative training procedure as the multilayer perceptron neural networks model, and then the local minimum problem is not faced in the GRNN modeling. Figure 2 shows a schematic diagram of generalized regression neural network architecture. GRNN consists four layers: input layer, pattern layer, summation layer and output layer. The number of input units in the first layer is equal to the total number of parameters. The first layer is fully connected to the second pattern layer, where each unit represents a training pattern and its output is a measure of the distance of the input from the stored patterns. Each pattern layer unit is connected to the two neurons in the summation layer. The data for this study collected from 1996 to 2006 were divided in three sets for the purpose of training, cross-validation and testing.
The GRNN can be treated as a normalized radial basis function network in which the hidden unit is centered at every training case. These radial basis function units are usually probability of density functions such as the Gaussian. GRNN is a method for estimating the joint probability density function of input and output, given only a training set. Since the probability density function is derived from the data with no preconceptions about its form, the system is perfectly general. By definition, the regression of a dependent variable (output) on an independent (input) estimates the most probable value for output, given input and a training set. This study considers independent (input) and also according to the smoothing parameter.

The density function can be estimated from the training set using the Parzen’s nonparametric estimator [19]:

\[
f(x, y) = \frac{1}{np} \sum_{i=1}^{n} e^{-d(x, x_i)} e^{-d(y, y_i)}
\]

(4)

Where \(d(x, x_i) = \sum_{j=1}^{p} [(x_j - x_{i,j})/(\sigma_j)]^2\) and \(d(y, y_i) = [(y - y_i)/(\sigma)]^2\) the number of training patterns and the number of independent variables are denoted \(n\) and \(p\), respectively. The density function \(f(x, y)\) is therefore estimated by a weighted sum of the “Kernel function” [12]. The parameter \(\sigma\) represents the smoothing parameter.

The estimator \(f(x, y)\) is asymptotically unbiased and consistent [20]. An interpretation of the probability estimate \(f(x, y)\) is that it assigns sample probability of width \(\sigma\) for each \(i\) th value of \(x\) and \(y\). The indicated integration yields as the following:

\[
\hat{y}(x) = \frac{\sum_{i=1}^{n} y_i e^{-d(x, x_i)}}{\sum_{i=1}^{n} e^{-d(x, x_i)}}
\]

(5)

The predictor (5) is a weighted sum over all training patterns. It is directly applicable to problems involving numerical data. Each training pattern is weighted exponentially according to its Euclidean distance to the unknown pattern \(x\) and also according to the smoothing factors. This predictor was mapped into a neural network.

2.4. Data Preparation

The data used in the neural network for the ET_ref modeling were normalized in order to overcome the problem associated with extreme values. Hence, the input and output data sets were scaled in the range of [0 1] using the following equation [21]:

\[
y_{\text{norm}} = \frac{y_i - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}
\]

(6)

Where \(y_{\text{norm}}\) is the normalized dimensionless variable; \(y_i\) is the observed value of variable; then \(y_{\text{min}}\) and \(y_{\text{max}}\) are the minimum and the maximum values of the observed variable.

2.5. Statistical Analysis

Three statistical indicators were used for comparing the models performances, namely, root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (r), expressed as a percentage of the arithmetic mean of observed values:

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - y'_i)^2}{N}}
\]

(7)

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| y_i - y'_i \right|
\]

(8)

\[
r = \frac{\sum_{i=1}^{N} (y_i - \bar{y}_i)(y'_i - \bar{y}'_i)}{\sqrt{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2 \sum_{i=1}^{N} (y'_i - \bar{y}'_i)^2}}
\]

(9)

Where \(y_i\) represents the PM observed ET_ref, \(y'_i\) is the alternative methods estimated ET_ref for the \(i\)th values; \(\bar{y}_i\) and \(\bar{y}'_i\) represent the average values of the corresponding variable; and \(N\) represents the number of data considered. Additionally, a linear regression \(y = a_0 + a_1 x\) is applied for evaluating the models'
performance statistically, where $y$ is the dependent variable (PM); $x$ the independent variable (alternative methods); $a_0$, the intercept; and $a_1$, the slope.

3. Results and Discussion

3.1. ET_ref Estimation

The data collected between 1996 and 2006 in Dédougou region had a total of 396 patterns divided in three parts for the purpose of training from January 1996 to December 2003, cross-validation from January 2004 to December 2005, and testing from January 2006 to December 2006. The training data are used to train the network by minimizing the error data, the cross-validation used to find the network performance and guard against overtraining, and then the testing data used for checking the overall performance of the trained network. The testing data used were not given during the training phase previously. The present study used the decade time step data since it has been well documented by [22] and [23] as a suitable step for ET_ref computation when using the temperature-based models. The Hargreaves conventional method behaves best for decades predictions although some accurate ET_ref daily estimations have been reported in literature [24]. According to [25], the temperature-based models are not accurate as the PM for period less than 5 days.

In this study, ET_ref was estimated with the Generalized Regression Neural Networks (GRNN) and Hargreaves (HRG). In the GRNN modeling, very few user decisions are required. Among the decisions that are required is the selection of the appropriate smoothing factors ($\alpha$) to be applied to each of the model inputs. The success of the GRNN depends heavily on this smoothing factor [26]. The smoothing factor ($\alpha$) producing the best estimation results in this study has been found at 0.01. The statistical evaluation results obtained with HRG are 0.850, 0.429 mm day$^{-1}$ and 0.363 mm day$^{-1}$ for $r^2$, RMSE and MAE, respectively. While with GRNN, the results are 0.917, 0.272 mm day$^{-1}$, and 0.193 mm day$^{-1}$ for $r^2$, RMSE and MAE, respectively. GRNN produces the best result based on the $r^2$, RMSE and MAE. The good performances of GRNN vis à vis to PM can be seen from the scatter and plot representation in Figures 3a and b. More recently, [27] also obtained good results by using similar input data sets with the neural network.

HRG model shows clearly poor performance results in comparison with GRNN. The ET_ref comparison results between Hargreaves and PM taken as reference values showed an overestimation from the nineteen decade (July) of the year. This overestimation is most pronounced during the rainy season as shown in Figure 3b. In general, the conventional methods due to their models simplicity are unable to capture the effect of some important climatic parameters which affect ET_ref. The behavior of the HRG could be explained by the influence of other parameters such as wind which is not considered into the model. The weather conditions of the study area characterized by a low rainfall, high temperature variation and high wind speed, might affect the accuracy of the temperature-based estimates ET_ref. [8] found also an overestimation with HRG up to 28%. [28] observed an overestimation with HRG in the arid region of India. The performance of the HRG conventional method may strongly dependent of climatic condition. There is therefore a consensus that, the performances of most alternative methods have been found to vary from one climate to another [29]. Since the GRNN showed high performance than the HRG temperature-based method, this algorithm can be considered as a potential alternative approach for estimating the ET_ref in the semiarid zone of Africa. The accuracy of the GRNN might be improved by considering other parameters such as wind velocity which has affected significantly the temperature-based ET_ref in many other arid areas. In order to understand the influence of wind on ET_ref in this semiarid environment, the sensitivity analysis was carried out by considering wind speed as an additional input variable of the neural network.

![Figure 3. Models scatter (a) and plot (b) comparison during the testing period.](image)

3.2. ET_ref Sensitivity Analysis

Under the consideration of wind into the neural network input data sets, the ET_ref has been estimated and the statistical results summarized in Table 1. The scatter and
plot of ET_ref estimated when wind velocity is incorporated into the network are given in Figures 4a and b, respectively. For this windy region, GRNN accuracy improves significantly with wind velocity ($r^2=0.970$, RMSE=0.025 mm day$^{-1}$, MAE=0.124 mm day$^{-1}$). These good results show the evidence of the high sensitivity of ET_ref for the wind speed under the Dédougou weather condition. ET_ref is sensitive to wind [7] and its performance may be also influenced [30]. [31] observed a positive correlation between ET_ref and wind speed in the East Arid Zone of Nigeria in Africa. It has been documented by [32] that, the climatic parameters such as wind velocity simultaneously results by deteriorating ET_ref from temperature-based methods. [26] by using the ANN found that the wind speed was more effective for estimating ET_ref with the temperature-based method. According to [33] the impact of wind speed on the ET_ref results is relatively smaller except for arid windy areas. Wind is extremely required to be in the model for the neural network accuracy improvement in this African semiarid zone. When wind data is not available, ET_ref can still be better estimated with the GRNN than HRG using air temperature and extraterrestrial radiation data. It could be concluded that, the wind velocity is an important source for improving the network accuracy in ET_ref estimation for the semiarid zone of Africa.

**Table 1.** Summary of models statistical performances during the testing period estimation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Neural Network Input</th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>$r^2$</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRG</td>
<td>Tmax, Tmin, Tmean, Ra</td>
<td>-1.104</td>
<td>1.153</td>
<td>0.850</td>
<td>0.429</td>
<td>0.363</td>
</tr>
<tr>
<td>GRNN</td>
<td>Tmax, Tmin, Ra</td>
<td>0.293</td>
<td>0.930</td>
<td>0.917</td>
<td>0.272</td>
<td>0.193</td>
</tr>
<tr>
<td>GRNN_Wind</td>
<td>Tmax, Tmin, Ra, Wind</td>
<td>0.181</td>
<td>0.972</td>
<td>0.970</td>
<td>0.025</td>
<td>0.124</td>
</tr>
</tbody>
</table>

**4. Conclusion and Recommendation**

The accurate estimation of evapotranspiration is crucial for an efficient agriculture water management in the area where there is water resources rarity problem. Since it is well known that the large numbers of meteorological data required for PM are not always available in Africa, this study adopted an approach using few input variables. It was clearly showed from the ET_ref modeling results that, HRG performs less than GRNN. Beside, HRG overestimated under the weather condition of this semiarid zone studied. GRNN produced the best estimation values close to the PM ET_ref. Using ANN with only temperature data under semiarid zone climatic condition of Africa for estimating ET_ref is highly superior to the others conventional methods. It has been observed that the accuracy of the GRNN improves significantly when wind velocity is considered into the neural network model. Wind is an important variable and extremely recommended to be into the model for this semiarid zone of Africa.

**References**


