Water Resour Manage (2012) 26:4367–4380
DOI 10.1007/s11269-012-0149-3

Regional-Specific Numerical Models of Evapotranspiration Using Gene-Expression Programming Interface in Sahel

Seydou Traore · Aytac Guven

Received: 20 April 2012 / Accepted: 17 September 2012 /
Published online: 25 September 2012

Abstract An accurate and simple Reference Evapotranspiration (ETo) numerical model eases to use for supporting irrigation planning and its effective management is highly desired in Sahelian regions. This paper investigates the performance ability of the Gene-expression Programming (GEP) for modeling ETo using decadal climatic data from a Sahelian country; Burkina Faso. For the study; important data are collected from six synoptic meteorological stations located in different regions; Gaoua, Pô, Boromo, Ouahigouya, Bogandé and Dori. The climatic data combinations are used as inputs to develop the GEP models at regional-specific data basis for estimating ETo. GEP performances are evaluated with the root mean square error (RMSE), and coefficient of correlation (R) between estimated and targeted Penman-Monteith FAO56 set as the true reference values. Obviously; from the statistical viewpoint; GEP computing technique has showed a good ability for providing numerical models on a regional data basis. The performances of GEP based on temperatures data are quite good able to substitute empirical equations at regional level to some extent. It is found that the models with wind velocity yield high accuracies by causing radical improve of the performances with R^2 (0.925-0.961) and RMSE (0.131-0.272 mmday^{-1}); while relative humidity may cause only (R^2=0.801-0.933 and RMSE=0.370-0.578 mmday^{-1}). Statistically; GEP is an effectual modeling tool for computing successfully evapotranspiration in Sahel.

Keywords Evapotranspiration modeling · Gene-expression programming · Sahel-specific data · Irrigation planning

S. Traore
African Climate Policy Center (ACPC/FSSDD), United Nation Economic Commission for Africa (UNECA), Addis-Ababa, Ethiopia

S. Traore
Ministry of Research and Scientifc Innovation, INERA/GRN-SP, Farako-bâ, Bobo Dioulasso, Burkina Faso

A. Guven (✉)
Department of Civil Engineering, University of Gaziantep, 27310 Gaziantep, Turkey
e-mail: aguven@gantep.edu.tr

1 Introduction

Irrigation is necessary for increasing the crop yields; and according to Isidoro et al. (2004) poor irrigation performance results in important social, economical and environmental problems. In Sahel, agriculture water resources shortages have pushed irrigation development and management to forefront as a key strategy to ensure crop high productivity in spite of rainwater limitation. In Sub-Saharan African regions where the rainfall is capricious, water resources effective planning and management are fundamental for enhancing crop productivity (Traore et al. 2010a). Burkina Faso lies in the Sudano-Sahelian zone, where the climate is harsh with a drastic variation characterized by high temperatures and low rainfall with a strong inter-annual and space-time variability. As a result, these climatic constraints made Burkinabe agriculture more vulnerable, as crops are essentially rainfed.

In fact, when budgeting the irrigation requirements, the evaporative demand of both soil and crop is a crucial functional variable. Although the physical-based Penman-Monteith FAO56 is the sole reliable and unanimously recommended method to compute accurate Reference Evapotranspiration (ETo) for different climate situation (Allen et al. 1998); its application in poor data situation is very limited. Indeed, Penman-Monteith equation requires full meteorological data set and its application for the manual computation of ETo is also a very hard task for irrigation technicians in the production area.

ETo has alternatively been modeled in several studies with limited inputs, but unfortunately their models accuracies generalization within a country or across regions poses problem due to the difference in regional specificity. Previous works done by Traore et al (2008) and Wang et al. (2009, 2011) have reported that to be much more appropriated to model ETo across the Sahelian regions as a way to solve the data unavailability in irrigation sites. That shows the extreme need to dispose of reliable ETo estimation methods using few data for all the respective regions of Burkina Faso. The inability to estimate ETo in each region using limited data set has even been a fundamental obstacle for irrigation development in Burkina Faso (Wang et al. 2009; Traore et al. 2008). Throughout Sub-Saharan Africa, the development of novel and simple numerical regional-specific based ETo equation is widely perceived to be the key of irrigation planning and water management. In Burkina Faso, simple and accurate ETo model is still helpfulness for technicians involved in irrigation production and extension network.

Scientists have paid considerable attention in modeling hydrological nonlinear complex systems. The focal interest of the present research is to employ Gene-expression Programming (GEP) technique for the explicit mathematical formulation of the evapotranspiration using regional-specific data. GEP interface is a powerful tool widely applied in many engineering researches with good generalization ability. However; according to Guven and Kisi (2011), GEP is rarely applied in water engineering research. Genetic programming technique has been employed in water research with only a limited number of studies of Dorado et al. (2003), Aytek and Kişi (2008), Guven (2009), Guven and Aytek (2009), Guven and Kisi (2010), Kisi and Shiri (2011) and Azamathulla et al. (2011). GEP can express numerical functions and enhance our understandings of the physical mechanisms involved in the ETo, and has huge advantages on the artificial neural networks techniques which have been previously applied for modeling ETo in Burkina Faso by Traore et al. (2010b) and Wang et al. (2011). However, the application of GEP in developing world for solving nonlinear modeling problem is poorly referenced in literature. To our knowledge, no work has yet been done for ETo estimation in Sub-Saharan Africa by using GEP techniques excepted of our recent investigation (Traore and Guven 2011). Therefore; this study aims to model the Reference Evapotranspiration expressed in numerical functions on the regional data basis by using GEP technique under Sahel environment of Sub-Saharan Africa.

2 Material and Methods

2.1 Location and Meteorological Data Characterization

In the present study, the meteorological data are collected across the country from six synoptic stations; Gaoua (Latitude 10°33'N, Longitude 3°18'W), Pô (Latitude 11°15'N, Longitude 1°15'W), Boromo (Latitude 11°75'N, Longitude 2°93'W), Ouahigouya (Latitude 13°56'N, Longitude 2°42'W), Bogandé (Latitude 12°97'N, Longitude 0°140'W) and Dori (Latitude 14°03'N, Longitude 0°03'W). For the purpose of this study, the decadal climatic data including precipitation (mm), maximum and minimum air temperatures (°C), relative humidity (%), wind speed ($\mathrm{km\,day^{-1}}$) and sunshine duration (hours) were recorded between 1996 and 2006. According to the rainfall feature partitioning; Burkina Faso has been divided in three large climatic zones which are the Guinea, Sudan and Sahelian zones. Figure 1 presents both spatial distribution of rainfall average and location of the regions under study in the country.

2.2 Reference Evapotranspiration Models

2.2.1 FAO56 Penman-Monteith

The reference evapotranspiration model applied in the study as a true targeted value is the FAO56 Penman-Monteith (PM) which is given by Allen et al. (1998) as the following:

$$ETo = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

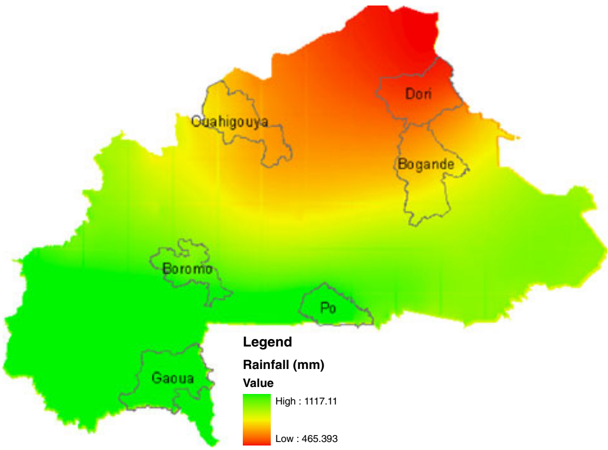

Fig. 1 Location of the regions studied herein in Burkina Faso and their annual precipitation average (1996-2006) spatially distributed

where ETo 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); Δ the slope vapor pressure curve (kPa °C^{-1}); and γ the psychrometric constant (kPa °C^{-1}).

2.2.2 GEP ETo Model Formulation

GEP is a research technique that allows the solution of problems by automatically generating algorithms and expressions. The flowchart from the Fig. 2 describes the GEP algorithm employed herein. It has five major steps to solve the present problem in this research paper. The first step in preparing to employ the GEP paradigm is to identify the set of terminals to be used in the individual computer programs. The major types of terminal sets contain the independent variables of the problem, the state variables of the system and the functions with no arguments. The second major step is to determine the set of functions (e^x, x^a, sin(x), cos(x), ln(x), log(x), 10^x, etc.) and arithmetic operations (+,-,/,*). In this study, various function sets were tried and (√, x^2,ln(x), e^x, a^x, x^a) was found to give best results. The third major step is fitness measure, which identifies the way of evaluating how good a given program solves a particular problem. In the present study, the root mean square error (RMSE) of the training set is taken as fitness function. The fourth major step is the selection of certain parameters to control the runs. The control parameters contain the size of the population, the rate of crossover, etc. The genetic operators used in this study were the Chromosomes (20-30), Head size (6-8), Number of genes (2.3), Mutation rate (0.044), Inversion rate (0.1), One-point and two-point recombination rate (0.3), Gene recombination and transposition rates (0.1). The last step is the determination of the criteria to terminate the run. Once the terminal and non-terminal operators are specified, the automatic program generation is carried out by means of a process derived from Darwin's evolution theory, in which, after subsequent generations, new trees (individuals) are produced from former ones via crossover, copying, and mutation (Guven et al. 2008). Based on natural selection, the best trees will have more chances of being chosen to become part of the next generation. Thus, a stochastic process is established where, after successive generations, a well-adapted tree is obtained.

The Gene-expression programming models developed in the present study are coded as $GEP_{code1\text{-}6}$ according to their input structure. Hence, GEP_{code1} is designed as the temperature-based model using only minimum temperature (Tmin), maximum temperature (Tmax) and extraterrestrial radiation (Ra) data. GEP_{code2} and GEP_{code3} are formed by inserting wind velocity (u2) and relative humidity (Rh) into GEP_{code1} combination, respectively. The model of GEP_{code1} integrating both Rh and u2 are presented by GEP_{code4}. Then, the model of GEP_{code5} is the combination of GEP_{code4} inputs plus sunshine (sun) variable. Finally; GEP_{code6} inputs are illustrated by mean temperature (Tmean), Ra, Rh and u2.

2.3 Models Performances Evaluation

The statistical indicators employed to evaluate the models performances are the goodness-of-fit measures of root mean square error (RMSE) and coefficient of determination (R^2)

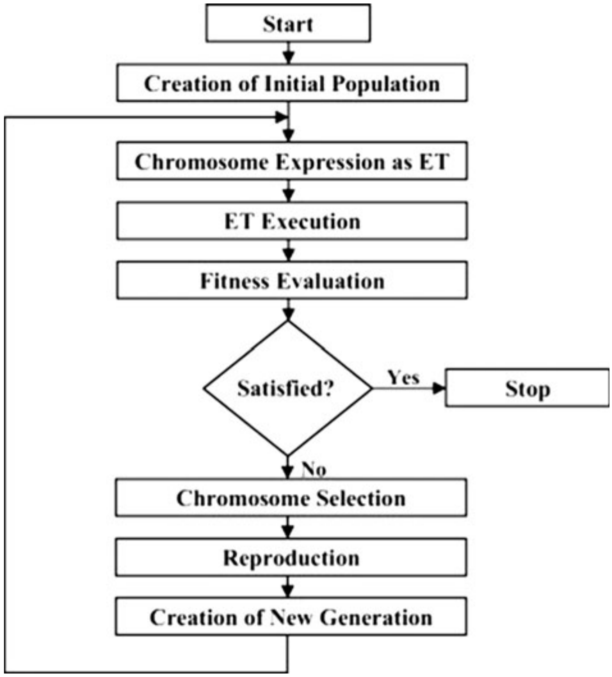

Fig. 2 Flowchart of the employed GEP algorithm

which are expressed as the following:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(y_i - y_i'\right)^2}{N}} \tag{2}$$

$$R^2 = \frac{\left(\sum_{i=1}^{N} (y_i - \overline{y})(y'_i - \overline{y}')\right)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2 \sum_{i=1}^{N} (y_i - \overline{y}')^2} \tag{3}$$

where y_i represents the PM targeted ETo, y_i' is the GEP estimated ETo for the ith values; $\overline{y}$ and $\overline{y}'$ represent the averages values of the corresponding variable; and N represents the number of data considered. Additionally, a linear regression $y = \alpha_1 x + \alpha_0$ is applied for evaluating the models' performance statistically, where y is the dependent variable (GEP methods); x the independent variable (PM); α_1 the slope and α_0 the intercept.

3 Results and Discussion

In the study, the meteorological data used as inputs with GEP to estimate ETo are composed of temperatures, sunshine, wind velocity (u2) and relative humidity (Rh) data. For the models inputs data structures, six combinations have been composed for the reason that in the past, Traore et al. (2010b) have apprehended both u2 and Rh as critical parameters in ETo modeling to some extent under this African semiarid environment. Therefore; the study does not consider the full range of the sixteen possible combinations from the recorded inputs sets. The Gene-expression Programming numerical models corresponding to each region under study are developed by using six different meteorological combinations sets. These models are the regions-specific equations developments that are function of the data availability opportunity in this Sahelian environment. Note that simpler equations have a limited range of applicability (Kisi 2006). According to Trajkovic (2005), people should adapt all ETo calculation to their local conditions. Khoob (2008) has considered a regional level study in order to give to the developed model, a higher regional capacity that could be applied to estimate ETo. Lopez-Urrea et al. (2006) have even suggested the diffusion of their results in local ambit. So, in Burkina Faso, there is difficulty to obtain the data, and also the data availability can differ from one region to another, in such circumstances GEP derived regional models that can provide available options for each region are highly desired. Table 1 shows the GEP models performances results at the testing period for the region of Gaoua, Pô, Boromo, Ouahigouya, Bogandé and Dori. In the Table 1, noted that the statistical evaluation criteria of the RMSE and R^2 inform about the models predictive capabilities and measure the degree to which estimated and targeted variables are linearly related, respectively. The GEP_{code1} generated as temperature-based model with the inputs structure combining only temperatures data i.e., Tmin, Tmax and Ra produced the lowest performances. From the results of the study, the performances of the GEP_{code1} are ranged from 0.525 to 0.856 for the R^2 and 0.269 to 0.644 mm day^{-1} for the RMSE. In the country, it is common to see technicians employ alternative models such as Hargreaves (HRG), Blaney-Criddle (BCR) and Reference Model for Burkina Faso (RMBF) in the ETo estimation despite their low performances. Based on the results from this present study, GEP_{code1} show better performances than those of the above alternative methods which poor performances ($R^2\sim$ 0.367, RMSE~1.845 mm day^{-1}) in Burkina Faso were previously reported in Traore et al. (2008) and Wang et al. (2011) studies'. In spite of the quite performances of the GEP_{code1}, and in comparison with our past studies, GEP_{code1} can be an alternative to the empirical equations to some extent. The formulae for GEP_{code1} comprised of three climatic variables obtained for Gaoua, Pô, Boromo, Ouahigouya, Bogandé and Dori are given below in Eqs. 4, 5, 6, 7, 8 and 9, respectively as:

$$ETo = \frac{T_{\max}(T_{\min}+4.73)^2}{171.27T_{\min}} - \frac{2.45T_{\max}+6.13}{R_a+8.27} + \frac{19.25}{T_{\min}} + R_a + 3.37 \tag{4}$$

$$ETo = \frac{(0.11R_a-0.74)^{0.15T_{\min}}}{T_{\max}-R_a} + \frac{R_a^2+6.54}{16.21R_a\left(\sqrt{T_{\min}}-3.21\right)} + \frac{T_{\max}}{13.04-0.37T_{\min}} \tag{5}$$

$$ETo = (T_{\max}-R_a)^{2\sqrt{R_a}-5.1} + 9.09^{1.5\sqrt{T_{\min}}-7.89} + \frac{T_{\min}^3-6.05}{T_{\max}^3} + \frac{T_{\min}+1.92}{T_{\max}+2.14} + \sqrt{2T_{\max}} - 0.79 \tag{6}$$

Table 1 Statistical performances of the GEP models in ETo estimation at the testing period in the study areas

Region	Model code	Input set	R^2	RMSE ($mm\,day^{-1}$)
Gaoua	GEP_{code1}	Tmin,Tmax,Ra	0.856	0.269
Pô		Tmin,Tmax,Ra	0.737	0.366
Boromo		Tmin,Tmax,Ra	0.816	0.287
Ouahigouya		Tmin,Tmax,Ra	0.646	0.592
Bogandé		Tmin,Tmax,Ra	0.525	0.644
Dori		Tmin,Tmax,Ra	0.719	0.429
Gaoua	GEP_{code2}	Tmin,Tmax,Ra,u2	0.952	0.157
Pô		Tmin,Tmax,Ra,u2	0.942	0.175
Boromo		Tmin,Tmax,Ra,u2	0.961	0.131
Ouahigouya		Tmin,Tmax,Ra,u2	0.925	0.272
Bogandé		Tmin,Tmax,Ra,u2	0.934	0.246
Dori		Tmin,Tmax,Ra,u2	0.956	0.169
Gaoua	GEP_{code3}	Tmin,Tmax,Ra,Rh	0.933	0.370
Pô		Tmin,Tmax,Ra,Rh	0.849	0.456
Boromo		Tmin,Tmax,Ra,Rh	0.898	0.403
Ouahigouya		Tmin,Tmax,Ra,Rh	0.801	0.578
Bogandé		Tmin,Tmax,Ra,Rh	0.804	0.546
Dori		Tmin,Tmax,Ra,Rh	0.818	0.534
Gaoua	GEP_{code4}	Tmin,Tmax,Ra,Rh,u2	0.965	0.131
Pô		Tmin,Tmax,Ra,Rh,u2	0.949	0.161
Boromo		Tmin,Tmax,Ra,Rh,u2	0.966	0.125
Ouahigouya		Tmin,Tmax,Ra,Rh,u2	0.964	0.184
Bogandé		Tmin,Tmax,Ra,Rh,u2	0.988	0.102
Dori		Tmin,Tmax,Ra,Rh,u2	0.973	0.141
Gaoua	GEP_{code5}	Tmin,Tmax,Ra,Rh,u2,Sun	0.972	0.126
Pô		Tmin,Tmax,Ra,Rh,u2,Sun	0.952	0.163
Boromo		Tmin,Tmax,Ra,Rh,u2,Sun	0.957	0.148
Ouahigouya		Tmin,Tmax,Ra,Rh,u2,Sun	0.982	0.134
Bogandé		Tmin,Tmax,Ra,Rh,u2,Sun	0.960	0.188
Dori		Tmin,Tmax,Ra,Rh,u2,Sun	0.980	0.119
Gaoua	GEP_{code6}	Tmean,Ra,Rh,u2	0.964	0.135
Pô		Tmean,Ra,Rh,u2	0.968	0.132
Boromo		Tmean,Ra,Rh,u2	0.937	0.167
Ouahigouya		Tmean,Ra,Rh,u2	0.971	0.168
Bogandé		Tmean,Ra,Rh,u2	0.943	0.223
Dori		Tmean,Ra,Rh,u2	0.964	0.157

$$ETo = \sqrt{T_{\max}\sqrt{T_{\min}} + T_{\min} + 5.19R_a} + (4.94 - 0.14T_{\min})^{-3.08} - \frac{9.36T_{\min} - 22.53}{22.53 + T_{\min}} - 8.02 \tag{7}$$

$$ETo = \sqrt{R_a + T_{\min}} + T_{\min}^{-1.25} + \sqrt{T_{\max} + 0.027T_{\min}R_a - \sqrt{R_a} - 0.5} + 10T_{\max}\left(\frac{R_a}{T_{\max}}\right)^{T_{\min}} - 8.59 \quad (8)$$

$$ETo = \left(\frac{R_a}{T_{\min} - 44.43}\right)^2 + \frac{T_{\max}^2}{T_{\min}^{3.3}} + \frac{T_{\min} - R_a + 3.76}{R_a - T_{\max} + 3.44} + \sqrt{T_{\min}} + 3.73 \quad (9)$$

The GEP_{code2} model whose input combinations are composed of the temperatures and wind velocity data (Tmin, Tmax, Ra, u2) increases drastically in performances with the R^2 and RMSE ranged between 0.925-0.961 and 0.131-0.272 $mm day^{-1}$, respectively. By referring to GEP_{code1}, it is observed that the insertion of wind alone into the model represented by GEP_{code2} causes drastic increasing of the R^2 for about 11.21 (Gaoua), 17.77 (Boromo), 27.82 (Pô), 32.96 (Dori), 43.19 (Ouahigouya) and 77.90 % (Bogandé). Strong wind velocity was recorded in all sites; this may explain its significant influence on ETo in Sahel. The soil evaporative demand and crop transpiration increase in dry windy area. GEP provided an opportunity to capture the critical variables involving in the ETo Process. The effect of wind on ETo is particularly more heightened for both Bogandé (310.0 $km day^{-1}$) and Ouahigouya (317.0 $km day^{-1}$) where the highest values were found. The factor such as wind velocity is ultimately relevant in evapotranspiration estimation especially for semiarid regions, and wind has been found by Traore et al. (2010b, 2011) to be the most effective variable in modeling the ETo physical-based nonlinear complex process in the Sudano-Sahelian zone of Africa. According to Popova et al. (2006), the impact of wind speed on the ETo results is relatively smaller except for arid windy areas. The behavior of the ET temperature-based model at the windy and non-windy regions could be explained at some extent by the fact that wind in the atmosphere decreases the temperature during the daytime and increases it during the nighttime (Temesgen et al. 1999). It has also been documented by Cob and Juste (2004) that wind occurred an important variation of ETo computed from limited model input such as temperature-based methods. So, in such particular weather condition, previous investigations have shown that temperature data alone may not stand significantly when expressing ETo formulation with the highest accuracy. This is clearly evidenced from the results of this present study when comparing GEP_{code1} and GEP_{code2}. The latter model which is wind incorporated yields high accuracy. In fact, wind speed is a much serious source of errors, and Allen et al. (1998) stated that the climatic parameters such as wind velocity deteriorates the ETo estimated.

Statistically, the GEP_{code2} formulation produced a good estimation of ETo in all regions, and accordingly this attest the powerfulness of wind variable which cannot be overlooked in Sahelian environment. Obviously, it is observed that, Gene-expression Programming technique is capable to catch up the ETo trend in Sahel. Figure 3 shows the comparison plots of the models of GEP_{code1} and GEP_{code2} between their ETo estimated and targeted values for the regions under study. Obviously, it is observed that, Gene-expression Programming technique is capable to catch up the ETo trend in Sahel. In Fig. 4, the scatter plots between ETo targeted and estimated values for the GEP_{code2} models that yield good performances are represented for the different regions studied. From these results given in Fig. 4, GEP_{code2} ETo estimates are more closer to the targeted values showing the fit lines closer to the exact

(45°) line with the slopes closer to 1, the intercept almost 0 with higher R^2(0.92-.94) than those of $GEP_{code1.}$ Since the developed GEP_{code2} model produced more accurate results than the GEP_{code1} temperature-based model, it can be suggested for computing ETo in each region when the others variables are missing excepted wind and temperatures. In methodology development, this study associated temperatures information to a key parameter (wind) that is influencing ETo in Sahel translated into mathematical expression given in $GEPcode_2$. GEP converts the nonlinearity connection between parameters taking as independent variables to estimate ETo which is considered as the dependent variable. The formulae for GEP_{code2} obtained for Gaoua, Pô, Boromo, Ouahigouya, Bogandé and Dori are given below in Eqs. 10, 11, 12, 13, 14 and 15, respectively as:

$$ETo = \frac{85.26 - u_2}{T_{\max} - 7.2T_{\min}} + \sqrt{R_a + \frac{74.55}{R_a - T_{\max}}} + \left(\sqrt{T_{\max} + T_{\min} - \frac{7.99}{R_a - 6.19}} - 6.29\right) \tag{10}$$

$$ETo = \left(\frac{10.90}{6.76 - T_{\min}} - 1.38\right) + \sqrt{\frac{(u_2 - T_{\max})^2}{100T_{\min}} + R_a + 145.56\sqrt{T_{\max}T_{\min}(u_2 + T_{\max})}} \tag{11}$$

$$\begin{aligned} ETo = &\left(\frac{\left(\sqrt{T_{\min} + u_2} - 6.90\right)u_2}{T_{\min}(R_a + 5.03)}\right) + \left(\frac{T_{\min}/u_2 - 24.62}{T_{\max} - 11.97}\right) \\ &+ \left(\sqrt{T_{\min} + 0.057(R_a - 6.0)^2} - 7.13\right) \end{aligned} \tag{12}$$

$$\begin{aligned} ETo = &\left(T_{\max} + R_a - \sqrt{R_a - 12.85}\right) + \frac{T_{\max}^2 + 2u_2}{4.66T_{\max} + 4.33} \\ &- (T_{\max} + 0.72R_a + 7.52) \end{aligned} \tag{13}$$

$$\begin{aligned} ETo = &\left((0.04R_a)^{T_{\max} - R_a} + 5.83\right) + \sqrt{1.23T_{\min} + 0.2u_2} \\ &+ \sqrt{T_{\max} - \frac{12.54R_a + 8.2T_{\min}}{R_a + 0.08}} \end{aligned} \tag{14}$$

$$ETo = \left(0.28\sqrt{u_2 - 0.2} - 2.76\right) + \sqrt{1.56^{\frac{T_{\max} - u_2}{2}} + T_{\min} + \frac{(0.11R_a)^{7.61} - 9.18T_{\min}}{84.05(R_a - 6.92)}} \tag{15}$$

The exclusion of wind and inclusion of Rh represented by the GEP_{code3} model whose inputs structures are composed of Tmin, Tmax, Ra and Rh decreased the models accuracies in respect of the reduction in the performances statistics (R^2=0.801-0.933; RMSE=0.370-0.578 mm day^{-1}). While the insertion of wind into GEP_{code4} that also has relative humidity variable has given better accuracy than its exclusion. GEP_{code4} model performances are ranged from 0.949 to 0.988 and 0.102 to 0.184 mm day^{-1} for the R^2 and RMSE, respectively.

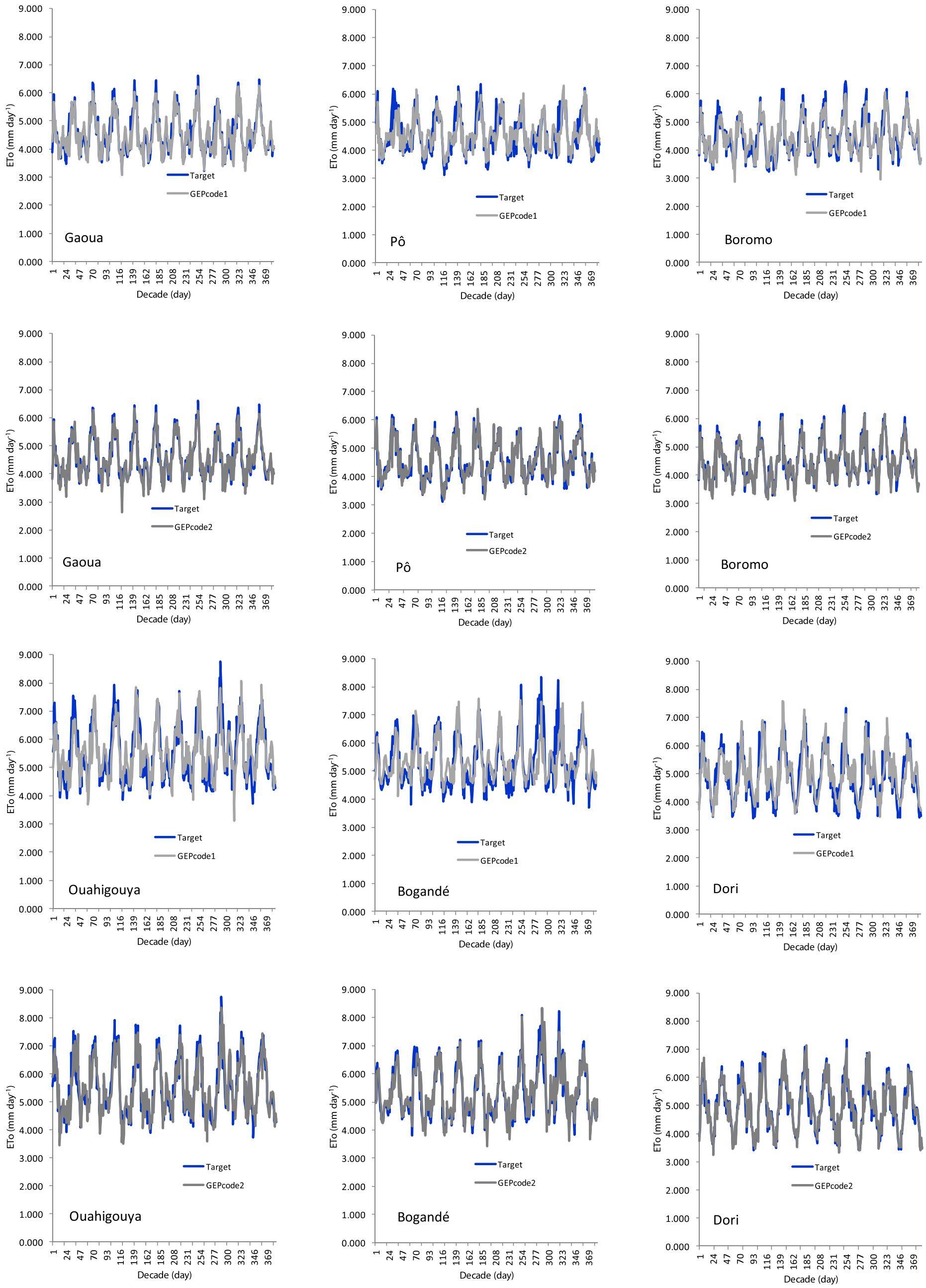

Fig. 3 Comparison plots between estimated and targeted ETo values for the models of GEPcode1 and GEPcode2

The regression of $\mathrm{GEP}_{\mathrm{code3}}$ estimated ETo values versus to the targeted values in the regions under study indicated that $\mathrm{GEP}_{\mathrm{code3}}$ models formulated are not good as $\mathrm{GEP}_{\mathrm{code2}}$ and $\mathrm{GEP}_{\mathrm{code4}}$. The scatter plots from Fig. 4 that represent both $\mathrm{GEP}_{\mathrm{code2}}$ and $\mathrm{GEP}_{\mathrm{code4}}$ estimates

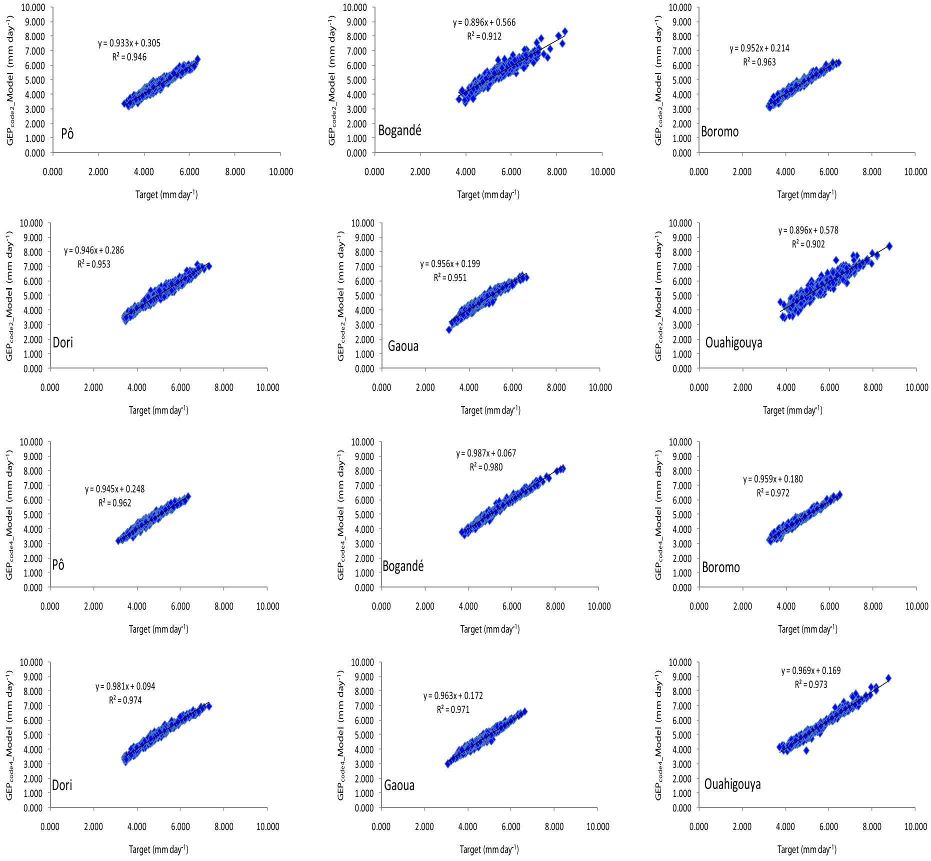

Fig. 4 Comparative scatter plots between estimated and targeted ETo values for the models of GEP_{code3} and GEP_{code4}

values regressed on their respective targeted ETo values confirm the statistics given in Table 1. Referring to GEP_{code1}, GEP_{code3} increases only the R^2 for about 9.00 (Gaoua), 10.05 (Boromo), 13.77 (Dori), 15.20 (Pô), 23.99 (Ouahigouya) and 53.14 % (Bogandé). So, by comparing with GEP_{code2} wind model, the increasing of the GEP_{code3} performance due to Rh is not as good as wind.

Furthermore, GEP_{code4} models whose inputs have integrated both u2 and Rh produced closer estimates values to the target ETo values, and from the Fig. 4 the regression lines are also closer to the exact (45°) line with very high R^2 values ranged from 0.962 to 0.980. These results show the high predictive ability of the GEP_{code4} to estimate ETo in all regions. However; the models of GEP_{code5} and GEP_{code6} have also produced reasonably good results. Figure 5 shows the comparison scatters of GEP_{code5} and GEP_{code6} models between their estimated and targeted ETo values in the regions of investigation. The GEP_{code5} model integrating Tmin, Tmax, Ra, u2, Rh and sun gives the highest accuracies in Gaoua (R^2=0.972 ; RMSE=0.126 mmday^{-1}), Pô (R^2=0.952 ; RMSE=0.163 mmday^{-1}), Ouahigouya (R^2=0.982 ; RMSE=0.134 mmday^{-1}), Dori (R^2=0.980 ; RMSE=0.119 mmday^{-1}). The performance of GEP_{code5} is evidenced for Gaoua, Ouahigouya and Dori when regressing the targeted values to the estimated ETo values. The linear regression slopes are very close to 1 and the intercept almost reach 0. Although the models have good accuracies, they application may be limited in a data unavailability situation. Indeed, based on the statistical comparison results from this study, the performances for all models are ranked as $GEP_{code5}>GEP_{code4}>GEP_{code6}>GEP_{code2}>GEP_{code3}>GEP_{code1}$ for the region of Gaoua and Dori. While the performances rank are given as for Ouahigouya ($GEP_{code5}>GEP_{code6}>GEP_{code4}>GEP_{code2}>GEP_{code3}>GEP_{code4}$), Pô ($GEP_{code6}>GEP_{code5}>GEP_{code4}>GEP_{code2}>GEP_{code3}>GEP_{code4}$) and Bogandé ($GEP_{code4}>GEP_{code5}>GEP_{code6}>GEP_{code2}>GEP_{code3}>GEP_{code4}$). For Boromo region, GEP_{code4} comes at the top follows by $GEP_{code2}>GEP_{code5}>GEP_{code6}>GEP_{code3}>GEP_{code1}$. For the irrigation management purpose, the paper gives several data combination options to use, but the choice of any of these GEP models depends of the data availability context. Since the data unavailability is a main limitation of irrigation planning, a model with less data requirement is highly desired (Traore et al. 2010b). Noted that the relative humidity and sunshine are often unavailable or even available their records present a lot of missing values. In such circumstances, preference is given to the model of GEP_{code2}, although GEP_{code4}, GEP_{code5} and GEP_{code6} have also produced good attractive accuracies as showed in Fig. 5.

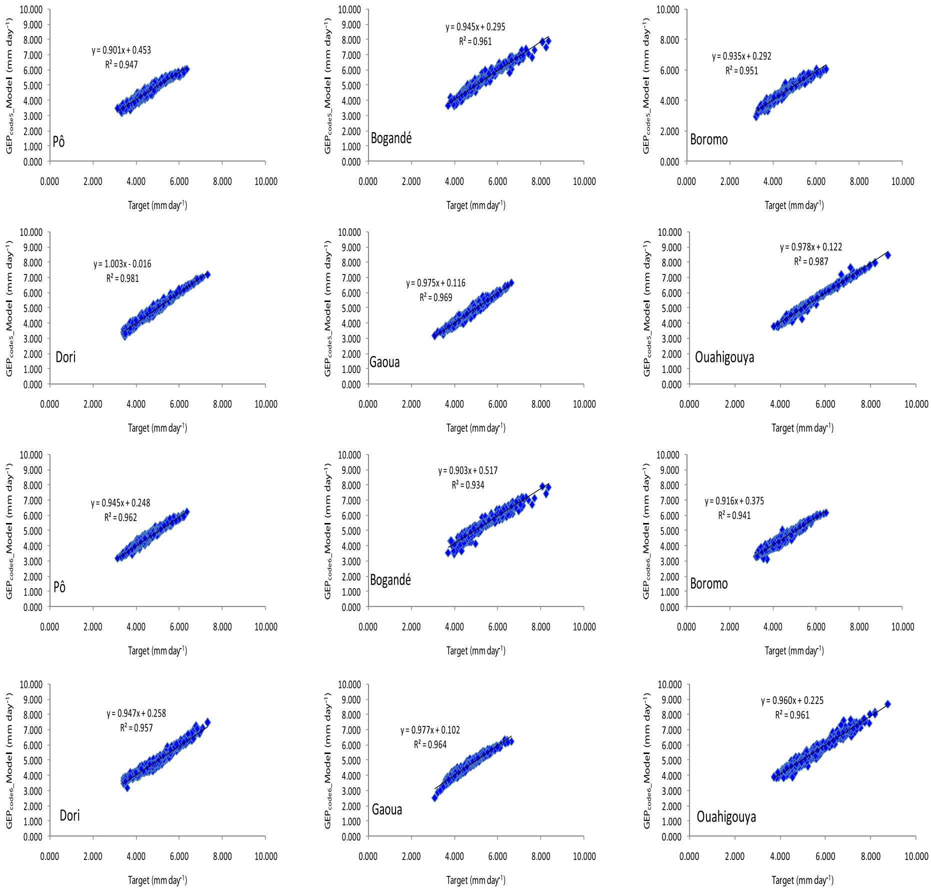

Fig. 5 Comparative scatter plots between estimated and targeted ETo values for the models of GEP_{code5} and GEP_{code6}

So, in Sahel only temperatures and wind may be sufficient to estimates accurately ETo when the others variables are missing. In overall, Gene-expression Programming ETo models are more adequate for computing ETo due to the simplicity of the numerical functions developed on regional data basis in Sahel. ETo is a sufficiently complex nonlinear process depending on several interactions between climatology factors, so, connecting variables in gene expression formulations according to the regional-specific data availability provide many advantages over traditional conventional approaches that are often used in Burkina Faso. These GEP regional-based ETo models are suggested in their originated regions for this Sahelian environment.

4 Conclusion

This study investigated the predictive ability of the Gene-expression Programming technique to build numerical and simple ETo formulations on the regional-specific data basis in the Sahelian country. When there is a water scarcity problem, adopting efficient irrigation systems is extremely desired as a suitable management option. The study has showed that with GEP technique, it is possible to formulate for each region an accurate numerical model ease to use by irrigation technicians for a manual computation of ETo in the environment of Sahel where the full meteorological data are often missing. The ETo formulations developed in the present study are simple explicit mathematical functions on a regional-specific data basis. Each GEP regional-based ETo model is suggested to be used in its originated development region. The factor of data availability should be considered when choosing the ETo GEP simpler equations proposed herein. This study provided convenient estimation tool ease agricultural extension agents for proper irrigation planning and water management purposes. Regional ETo accurate simpler models in Burkina Faso could be useful in other water resource analyses as well in poor weather data availability circumstance. Finally, GEP technique also provided an opportunity to apprehend the most sensitive variable involving in ETo Process.

References

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration, guideline for computing water requirements. FAO Irrigation drainage paper no 56, Rome, Italy

Aytek A, Kişi Ö (2008) Genetic programming approach to suspended sediment modelling. J Hydrol 351(3–4):288–298

Azamathulla HM, Ghani AA, Leow CS, Chang CK, Zakaria NA (2011) Gene-expression programming for the development of a stage-discharge curve of the Pahang River. Water Resour Manag 25(11):2901–2916

Cob AM, Juste MT (2004) A wind-based qualitative calibration of the Hargreaves ETo estimation equation in semiarid regions. Agr Water Manag 64:251–264

Dorado J, Rabunal JR, Pazos A, Rivero D, Santos A, Puertas J (2003) Prediction and modeling of the rainfall-runoff transformation of a typical urban basin using ANN and GP. Appl Artif Intell 17:329–343

Guven A (2009) Linear genetic programming for time-series modelling of daily flow rate. Hydrol Earth Syst Sci 118(2):137–146

Guven A, Aytek A (2009) New approach for stage–discharge relationship: gene-expression programming. J Hydrolog Eng 14(8):812–820

Guven A, Kisi Ö (2010) Daily pan evaporation modeling using linear genetic programming technique. Irrigat Sci 29(2):135–145

Guven A, Kisi Ö (2011) Estimation of suspended sediment yield in natural rivers using machine-coded linear genetic programming. Water Resour Manag 25:691–704

Guven A, Aytek A, Yuce MI, Aksoy H (2008) Genetic programming based empirical model for daily reference evapotranspiration estimation. Clean-Soil Air Water 36(10–11):905–912
Isidoro D, Quılez D, Aragüés R (2004) Water balance and irrigation performance analysis: La Violada irrigation district (Spain) as a case study. Agr Water Manag 64:123–142
Khoob AR (2008) Comparative study of Hargreaves's and artificial neural network's methodologies in estimating reference evapotranspiration in a semiarid environment. Irrigat Sci 26:253–259
Kisi O (2006) Generalized regression neural networks for evapotranspiration modelling. Hydrolog Sci J 51 (6):1092–1104
Kisi O, Shiri J (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resour Manag 25(13):3135–3152
Lopez-Urrea R, Olalla FMS, Fabeiro C, Moratalla A (2006) An evaluation of two hourly reference evapotranspiration equations for semiarid conditions. Agr Water Manag J 86:277–282
Popova Z, Kercheva M, Pereira LS (2006) Validation of the FAO methodology for computing ETo with limited data: Application to south Bulgaria. Irrigat Drain 55:201–215
Temesgen B, Allen RG, Jensen DT (1999) Adjusting temperature parameters to reflect well-water conditions. J Irrigat Drain Eng 125:26–33
Trajkovic S (2005) Temperature-based approaches for estimating reference evapotranspiration. J Irrigat Drain Eng 131(4):316–323
Traore S, Guven A (2011) New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa. Irrigat Sci. doi:10.1007/s00271-011-0288-y
Traore S, Wang YM, Kerh T (2008) Modeling reference evapotranspiration by generalized regression neural network in semiarid zone of Africa. WSEAS Trans Inf Sci & Appl 6(5):991–1000
Traore S, Wang YM, Kan CE, Kerh T (2010a) A mixture neural methodology for computing rice consumptive water requirements in Fada N'Gourma Region, Eastern Burkina Faso. J Paddy & Wat Env 8(2):165–173
Traore S, Wang YM, Kan CE, Kerh T (2010b) Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. Agr Water Manag J 97:707–714
Wang YM, Traore S, Kerh T (2009) Computational performance of reference evapotranspiration in semiarid zone of Africa. Sci Res Essays 6(4):577–583
Wang YM, Traore S, Kerh T, Leu JM (2011) Modeling reference evapotranspiration using feed forward backpropagation algorithm in arid regions of Africa. Irrigat Drain 60(3):404–417