

Genetic Programming: An approach for Evapotranspiration Modelling

Shreenivas Londhe¹, Girija Mulay²

¹Department of Civil Engineering, V.I.I.T, Pune, snlondhe@gmail.com

²Department of Civil Engineering, V.I.I.T, Pune, mulaygirijas15@gmail.com

Abstract—Estimation of evapotranspiration (ET) requires knowledge of the values of many climatic variables, some of which require special equipment and careful observations. Although ET is an important component of water balance, the data required for its accurate estimation are commonly available only at widely spaced measurement stations. Genetic programming (GP) is the technique adopted in this paper for modeling of evapotranspiration using the data provided by 3 stations of Florida. Models were created using measured inputs (solar radiation, mean temperature, average relative humidity and wind speed) and observed output (evapotranspiration). The results obtained were compared with conventional methods of evapotranspiration estimation. The criterions of root mean square error and coefficient of correlation are used to measure the performance of the model developed by employing GP. The error measures indicate that the new equation produced by GP yield quite satisfactorily results and can be used as an alternative to the conventional models.

Keywords-Evapotranspiration; genetic programming; root mean square error; coefficient of correlation

I. INTRODUCTION

Determining ET is particularly required for water resources management, irrigation system design, hydrological analysis, drainage studies, crop production, and environmental impact assessment. Direct measurement of ET is difficult, and hence costly. Empirical methods are data intensive; they require measurement of many meteorological variables such as solar radiation, air temperature, wind speed and relative humidity. It is therefore necessary to try alternative approach to estimate ET with reasonable accuracy and avoiding excessive data measurement. The soft computing tool of Genetic Programming which is essentially classified as an Evolutionary Computation (EC) technique has found its foot in the field of Hydraulic Engineering in general and modeling of water flows in particular since last 12 years or so [1]. It has been observed that only a few studies exist in the literature related to the use of GP in the field of water resources engineering [2]. This paper presents an approach to estimate ET using the equation produces by GP. The results have been compared with the conventional empirical methods, Penman – Monteith equation, Hargreaves equation and Hamon equation.

The succeeding section will present brief information about GP and GP operations followed by study area and data. The model formulation will be discussed later followed by results and discussion.

II. STUDY AREA AND DATA

The data set used in this study was obtained from site [3] at three stations at Florida were selected for model development. The Station 1 is located at St Johns Rv Marsh at Blue Cypress Nr Fellsmere, Florida, USA (USGS data Station No. 274143080424100) has latitude of

27°41'43.0" N and longitude of 80°42'41.0" E. Station 2 is located at Duda Farms Nr Lake Winder Nr Rockledge, Florida, USA (USGS data Station No. 281626080463400) has latitude of 28°16'26.0" N and longitude of 80°46'34.0" E in the Brevard County, Florida and Station 3 is located at Ferris Farms Nr Floral City, Florida (USGS data Station No. 284541082163400) has latitude of 28°45'40.7"N and Longitude is 82°16'34.4" E. Figure 1 shows the study area.

Data are available in the database at daily time scales, and includes observed ET, net solar radiation, maximum minimum air temperature, maximum minimum relative humidity and wind speed.

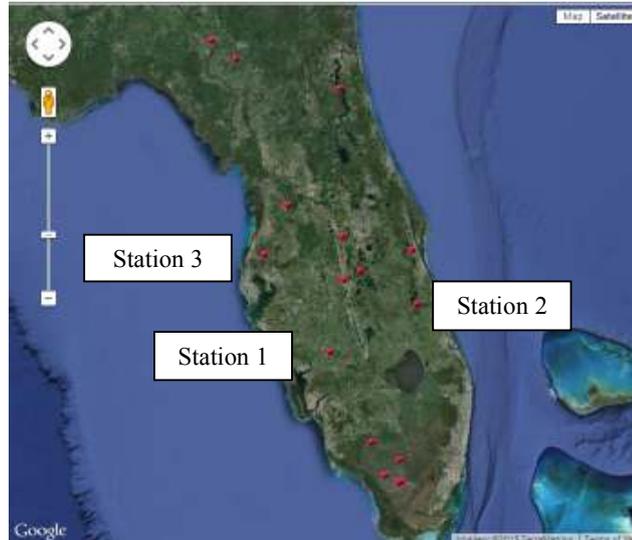


Fig.1 Study Area (Florida)

III. GENETIC PROGRAMMING

Like genetic algorithm (GA), the concept of GP follows the principle of ‘survival of the fittest’ borrowed from the process of evolution occurring in nature. But unlike GA, its solution is a computer program or an equation as against a set of numbers in the GA; and hence it is convenient to use the same as a regression tool rather than an optimization one like the GA. GP operates on parse trees rather than on bit strings as in a GA, to approximate the equation (in symbolic form) or computer program that best describes how the output relates to the input variables. A good explanation of various concepts related to GP can be found in [4]. In GP, a random population of individuals (equations or computer programs) is created, the fitness of individuals is evaluated and then the ‘parents’ are selected out of these individuals. The parents are then made to yield ‘offspring’s’ by following the process of reproduction, mutation, and crossover. The creation of offspring’s continues (in an iterative manner) till a specified number of offspring’s in a generation are produced and further till another specified number of generations are created. The resulting offspring at the end of all this process (an equation or a computer program) is the solution of the problem. The GP thus transforms one population of individuals into another one in an iterative manner by following the natural genetic operations like reproduction, mutation, and crossover. The algorithm for GP and details of genetic operations [1] are given as below. The freely available software “GP Kernel” with tree program structured was used to develop GP models in the present work.

Mutation: One individual is selected as per the fitness. A sub-tree is replaced by another one randomly. The mutant is inserted into the new population. Individuals are increased by 1. The allowable range of mutation frequency parameter is 0 to 100%. (Fig.2)

Cross over: Two individuals (programs) are chosen as per the fitness called parents. Two random nodes are selected from inside such program (parents) and thereafter the resultant sub-trees are swapped, generating two new programs. The resulting individuals are inserted into the new population. Individuals are increased by 2. The parents may be identical or different. The allowable range of cross over frequency parameter is 0 to 100%. (Fig.3)

Reproduction: The best program is copied as it is as per the fitness criterion and included in the new population. Individuals are increased by 1. Reproduction rate = $100 - \text{mutation rate} - (\text{crossover rate} * [1 - \text{mutation rate}])$

Fig.4 shows general flowchart of GP as given by [4]

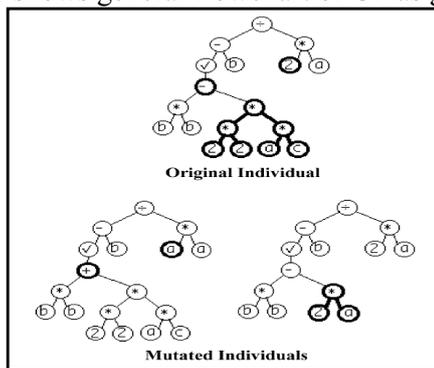


Fig.2 Mutation [5]

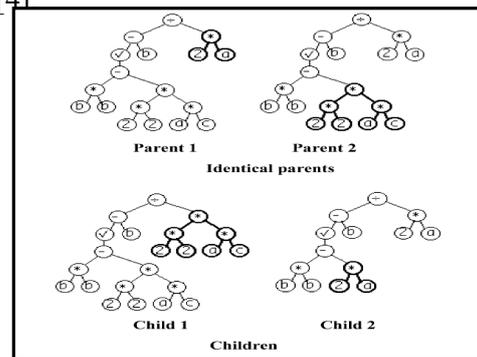


Fig.3 Crossover [5]

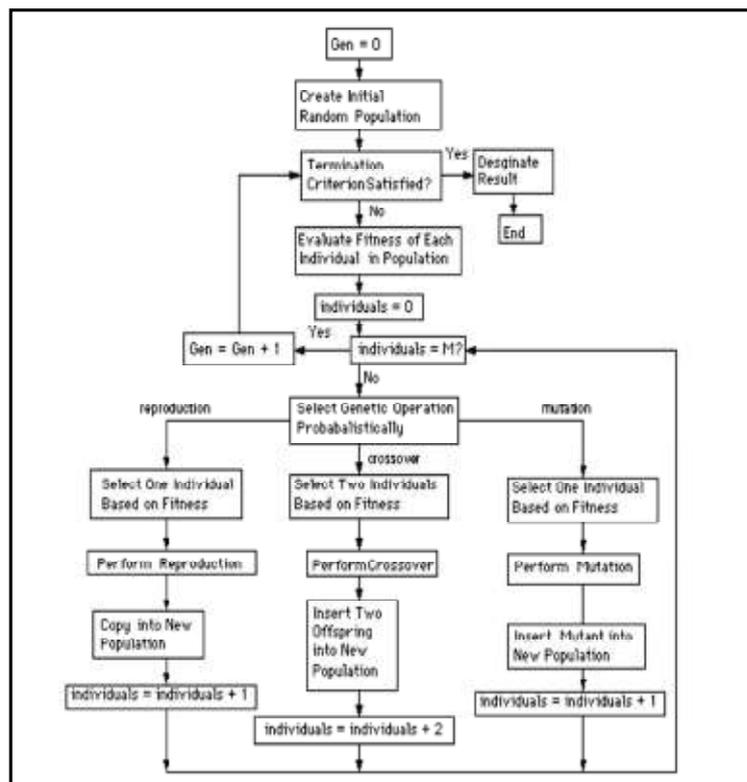


Fig.4. Flowchart of GP [4]

III. METHODOLOGY

The daily measured data of ET from January 2000 to September 2004 years was used in the present work (Station 1 assessment). Four input variables (net solar radiation, mean temperature, average humidity and wind speed) and observed ET as the output used to develop GP model. Inputs and outputs are selected as per the data availability. Data (70 % of the total) taken from the station were divided as training and (30% of the total) testing sets for the GP. Initially the crossover frequency of 50% and mutation frequency of 68% are used. The output of GP Kernel is in the form an equation. The equation developed from the training set was used for assessment of testing set. The performance of testing set was evaluated by the coefficient of correlation (R) and root mean square error (RMSE). Lower value of RMSE and higher value of R indicate better accuracy [6]. The equations for determination of ET by traditional method (Penman – Monteith equation, Hargreaves equation and Hamon equation) are given below

Penman – Monteith Equation [7]:

$$ET_0 = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + C_d u_2)}$$

1

Where,

ET₀ = evapotranspiration (mm day-1)

R_n = net radiation (MJ m⁻² day-1)

G = soil heat flux density (MJ m⁻² day-1)

T = mean daily air temperature at 2 m height (°C)

u₂ = wind speed at 2 m height (m s⁻¹)

e_s = saturation vapour pressure (kPa)

e_a = actual vapour pressure (kPa)

e_s-e_a = saturation vapour pressure deficit (kPa)

Δ = slope vapour pressure curve (kPa °C⁻¹)

Γ = psychrometric constant (kPa °C⁻¹)

Hargreaves Equation [8]:

$$ET_0 = 0.0023(T_m + 17.8) \left(\sqrt{T_{\max} - T_{\min}} \right) R_a$$

2

Where,

T_m = daily mean air temperature (°C)

T_{max} = daily maximum air temperature (°C)

T_{min} = daily minimum air temperature (°C)

R_a = extra-terrestrial radiation (MJ m⁻² day-1)

Hamon Equation [9]:

$$ET_0 = \frac{2.1 * H_t^2 e_s}{(T_{mean} + 273.2)} \tag{3}$$

Where,

Ht (day) is average number of daylight hours per day

Tmean = daily mean air temperature (°C)

V. RESULT AND DISCUSSION

The equation yielded by GP kernel is as follows

$$E = (A / \exp((\tanh(\tanh((A / \exp(\sqrt{\tanh(\sqrt{(1.12261 + B))})))))) + \sqrt{(\exp((\tanh(C) + \tanh((A / A))) / D))})) \tag{4}$$

Where,

A = Net radiation (MJ/m²)

B = Wind Speed (m/s)

C = Average relative humidity (%)

D = Mean temperature (degree C)

E = ET by GP (mm/day)

Results of all the models are presented in Table 1. GP model shows RMSE as 0.21 while other conventional methods shows higher values. GP and PM model shows R as 0.95 which is better than other two methods. As coefficient of correlation is a measure of the strength and direction of the linear relationship between two variables RMSE is more reliable in terms of comparison between two methods. Even though R for GP and PM is same, GP has lesser value of RMSE and hence superior. In case of all the models GP shows better results and accuracy.

Table1. Result Table

ET Model	GP	PM	Hargreaves	Homan
RMSE	0.21	0.77	6.94	1.10
R	0.95	0.95	0.83	0.71

VI. CONCLUSION

Evapotranspiration was estimated using 4 methods, GP, PM, Hargreaves and Hamon. GP seem to be giving quite accurate results in comparison with other methods. An attempt to improve the present results can be made by making changes in the methodology. The changes could be in the population size, crossover and mutation frequency.

REFERENCES

- [1] Londhe S. N. Dixit P. R., “Genetic Programming: A Novel Computing Approach in Modeling Water Flows”, InTech Publications, editor: Sebastian Ventura Soto, ISBN 980-953-307-215-4, <http://dx.doi.org/10.5772/48179>, 199-224, 2012.

- [2] Guven A, “Genetic Programming- Based Empirical Model for Daily Reference Evapotranspiration Estimation”, *Clean* 2008, 36 (10 – 11), 905 – 912.
- [3] <http://fl.water.usgs.gov/et/etdata.html>, “USGS Florida Evapotranspiration Stations”
- [4] Koza J, “Genetic Programming: On the Programming of Computers by Means of Natural Selection”. A Bradford Book. MIT Press, 1992.
- [5] Londhe S. N, “Soft Computing Approach for Real Time Estimation Of Missing Wave Heights”, *Ocean Engineering*, 35, (11-12), 1080-1089, 2008
- [6] Dawson C. W. Wilby R. L, “Hydrological modelling using artificial neural networks”, *Progress in Physical Geography* pp. 80–108, 2001.
- [7] FAO Irrigation and Drainage Paper N0. 56, February 2006
- [8] Hargreaves, G. H., and Samani, Z. A, “Reference crop evapotranspiration from temperature”, *Applied Engineering Agric*, 1:96-99, 1985.
- [9] Haith, D. A., and Shoemaker, L. L, “Generalized watershed loading functions for stream flow nutrients”, *Water Resources Bulletin*, 23:471-478, 1987.

