

## **MODELLING PAN EVAPORATION USING GENETIC PROGRAMMING**

**MOHAMMAD ALI GHORBANI<sup>1</sup>, VIJAY P. SINGH<sup>2</sup>, RASOUL  
DANESHFARAZ<sup>3</sup> and MAHSA HASANPOUR KASHANI<sup>1</sup>**

<sup>1</sup>Department of Water Engineering  
University of Tabriz  
Tabriz  
Iran  
e-mail: mahsakashani2003@yahoo.com

<sup>2</sup>Department of Biological and Agricultural Engineering  
Texas A and M University  
Scoates Hall, 2117 TAMU  
College Station, Texas 77843-2117  
USA

<sup>3</sup>Department of Civil Engineering  
University of Maragheh  
Maragheh  
Iran

### **Abstract**

Evaporation estimation is an important task in irrigation water demand estimation, water resources management, and determination of water budgets. In recent years, artificial intelligence models have been applied more successfully to model evaporation and other hydrological processes. In this study, daily evaporation, as a dependent variable, was modelled using genetic

---

2010 Mathematics Subject Classification: 62J02.

Keywords and phrases: artificial intelligence model, evaporation, genetic programming, modelling, pan evaporation.

Received April 5, 2012

programming (GP), which is one of the artificial intelligence models and an evolutionary data-driven modelling approach. We used data of four years (1987-1990) of five independent climatic variables including air temperature, solar radiation, wind speed, pressure, and humidity for two weather stations in California, USA. The GP models were trained and validated by using various combinations of the independent climatic variables. The performance of the selected GP model was compared with two artificial intelligence models, i.e., artificial neural network (ANN) and neuro-fuzzy (NF), and a traditional linear model by Stephen and Stewart (SS). The goodness of fit for the models was evaluated by using five performance criteria, namely, coefficient of determination, mean square error, mean absolute relative error, coefficient of efficiency, and index of agreement. The results obtained provided evidence that the GP model is capable of accurately modelling of evaporation and is a viable alternative to other artificial intelligence and traditional models.

## 1. Introduction

Evaporation is a significant concept in understanding the hydrological cycle, and its estimation is an important component in irrigation water demand, water resources management, and determination of water budgets, especially under arid conditions where water availability is scarce and fresh-water is limited. Direct methods of measuring evaporation rates employ evaporation pans but their provision at a point of interest is costly and therefore, there is a need for estimating potential evaporation from relevant parameters.

In recent years, artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), and genetic programming (GP) methods have been applied to model evaporation. GP was first proposed by Koza [23], as a generalization of genetic algorithms (GAs) (Goldberg [15]) and is particularly suitable, where interrelationships among relevant variables are poorly understood; theoretical analysis is constrained by assumptions and therefore their solutions are of limited use; and there is a large amount of data in computer readable forms requiring tedious processing. Owing to such restrictions, it is justifiable to test the application of GP to model evaporation rates in terms of directly measured independent variables of air temperature, solar radiation, wind

speed, pressure, and humidity. The focus of this technical note is to apply GP to model daily pan evaporation using other climatic data and compare its performance with those published by Kisi [19], using neuro-fuzzy and neural network techniques.

### 1.1. Literature review

Estimating reservoir evaporation using evaporation pans has long been popular. There are several pan methods, such as the use of direct measurements by Class A pan, as used by Salih and Sendil [31]; an adjusted form of those measurements according to the recommendation of Doorenbos and Pruitt [7]; or as suggested by Christiansen's model (Jensen [17]). A review by the Environment Agency for England and Wales (EA [9]) identified seven methods of estimating evaporation rates covering: pan evaporation, mass balance, energy budget models, bulk transfer models, combination models, the equilibrium temperature method, and empirical factor methods. The report ranked the methods against nine criteria (including accuracy, robustness, ease of use) for the purpose of the organisation and as a result the equilibrium temperature method was the first in the rank and empirical factors and the combination models were ranked equal second; but the recommended method did not include the pan evaporation method.

A number of attempts have been made to relate evaporation to climatic variables (Stephen and Stewart [34], Linarce [26], Burman [5], Gavin and Agnew [12]). Over the years, several linear relationships have been proposed by Stephen and Stewart (SS) [34] and this is employed in this note for comparison purposes. However, the process of evaporation is highly nonlinear because of the interdependence of climatic variables and therefore linear models are expected to perform poorly.

ANNs can approximate any nonlinear mathematical time series, so the estimation of pan evaporation has been investigated to model evapotranspiration by using ANNs. Kumar et al. [24] applied a multi-layer neural network using back propagation training algorithm; Sudheer

et al. [35] applied radial-based neural networks with limited climate data; Trajkovic [37] applied radial-based neural networks using temperature-based models; Kisi [20, 21] applied generalized regression neural networks and feed forward neural networks; Kisi [22] applied a neural computing technique using climatic data; Aytek [1] applied a co-active neuro-fuzzy inference system; and Moghaddamnia et al. [28] applied neural networks and an adaptive neuro-fuzzy inference system technique.

GP has been applied to a wide range of practical problems in artificial intelligence, engineering, and industry. However, it has been observed that the literature related to the use of GP in the field of water resources engineering is limited and includes: rainfall-runoff modelling studied by Drecourt [8], Savic et al. [32], Babovic and Keijzer [4], Muttil and Liong [29], Liong et al. [27], and Aytek and Alp [2]. Other hydrological and hydraulic applications include the determination of Chezy roughness coefficient in open channel hydraulics by Giustolisi [14]; flow forecasting in the River Nile, Northern Sudan by Sheta and Mahmoud [33]; a study of suspended sediment transport in streams by Aytek and Kisi [3], who found GP to perform better than conventional rating curves and multilinear regression techniques and sea level forecasting in an island environment of Indian Ocean, by Ghorbani et al. [13]. Only a few applications are cited on modelling evaporation, which comprise: Terzi and Keskin [36], who used the GP approach to evaporation estimation; and Guven et al. [16], who introduced a GP model for daily reference evapotranspiration.

## **2. Materials and Methods**

### **2.1. Genetic programming (GP)**

GP, proposed by Koza [23], is a generalization of genetic algorithm (GA), but they have one difference, where GP employs a “parse tree” structure for the search of its solutions, whereas GA employs byte strips. The technique is truly a “bottom up” process, as there is no assumption made on the structure of the relationship between independent and

dependent variables, but an appropriate relationship is identified for any given time series. The construction of the relationship is made possible by two components: (i) a parse tree, which is a functional set of basic operators and those selected in this study are

$$\{+, -, \times, \div, \sqrt{\quad}, \ln, e^x, x^2\}, \quad (1)$$

which emulates the role of RNA; and (ii) the actual components of functions and their parameters (referred to as the terminal set), which emulates the role of proteins or chromosomes in biological systems. When these two components work hand-in-hand, only then efficient emulation of evolutionary processes become possible.

The relationship between independent and dependent variables is often referred to as the “model”, the “program”, or the “solution”. Whatever, the terminology, the identified relationship in a particular GP modelling is continually evolving and never fixed. The evolution starts from an initially selected random population of models, where the fitness value of each model is evaluated by using the values of the independent and dependent variables. As the population evolves from one generation to another, new models replace the old ones by having demonstrably better performance.

There are various selection methods and the method used in this study is referred to as the gene expression programming (GEP) based on evolving computer programs of different sizes and shapes encoded in linear chromosomes of fixed lengths (Ferreira [10, 11]). The chromosomes are composed of multiple genes, each gene encoding a smaller subprogram. Furthermore, the structural and functional organization of the linear chromosomes allows the unconstrained operation of important genetic operators, such as mutation, transposition, and recombination. It has been reported that GEP is 100-10,000 times more efficient than GP systems (Ferreira [10, 11]) for a number of reasons, including: (i) the chromosomes are simple entities: linear, compact, relatively small, easy to manipulate genetically (replicate, mutate, recombine, etc.); (ii) the parse

trees or expression trees are exclusively the expression of their respective chromosomes; they are entities upon which selection acts, and according to fitness, they are selected to reproduce with modification.

Applying operators like crossover and mutation to the winners, “children” or “offspring” are produced, in which crossovers are responsible for maintaining identical features from one generation to another, but mutation causes a random change in the parse tree, although data mutation is also possible. This completes the operations at the initial generation and the process is repeated until termination. There are now various software applications for implementing GP models and the GeneXpro (Ferreira [10, 11]) was used in this study.

The procedure to model daily pan evaporation is as follows. The first step is the fitness function. The second step consists of choosing the set of terminals  $T$  and the set of functions  $F$ , to create the chromosomes. In the current problem, the terminal set includes climatic variables:  $\{T, SR, W, P, \text{ and } H\}$ . The choice of the appropriate function depends on the viewpoint of user. In this study, different mathematical functions were utilized ( $\{+, -, *, /, \sqrt{x}, \sqrt[2]{x}, \ln(x), e^x, x^2, x^3\}$ ). The third step is to choose the chromosomal architecture. Length of head,  $h = 7$ , and three genes per chromosomes were employed. The fourth step is to choose the linking function, which was “addition” for this study. The fifth and final step is to choose the genetic operators. The parameters used per run are summarized as follows:

Number of chromosomes: 30, head size: 7, number of genes: 3, linking function = addition, fitness function error type = root relative squared error, mutation rate = 0.044, inversion rate = 0.1, one point recombination rate = 0.3, two point recombination rate = 0.3, gene recombination rate = 0.1, gene transposition rate = 0.1, insertion sequence transposition rate = 0.1, root insertion sequence transposition = 0.1.

### 3. Study Area and Data Used

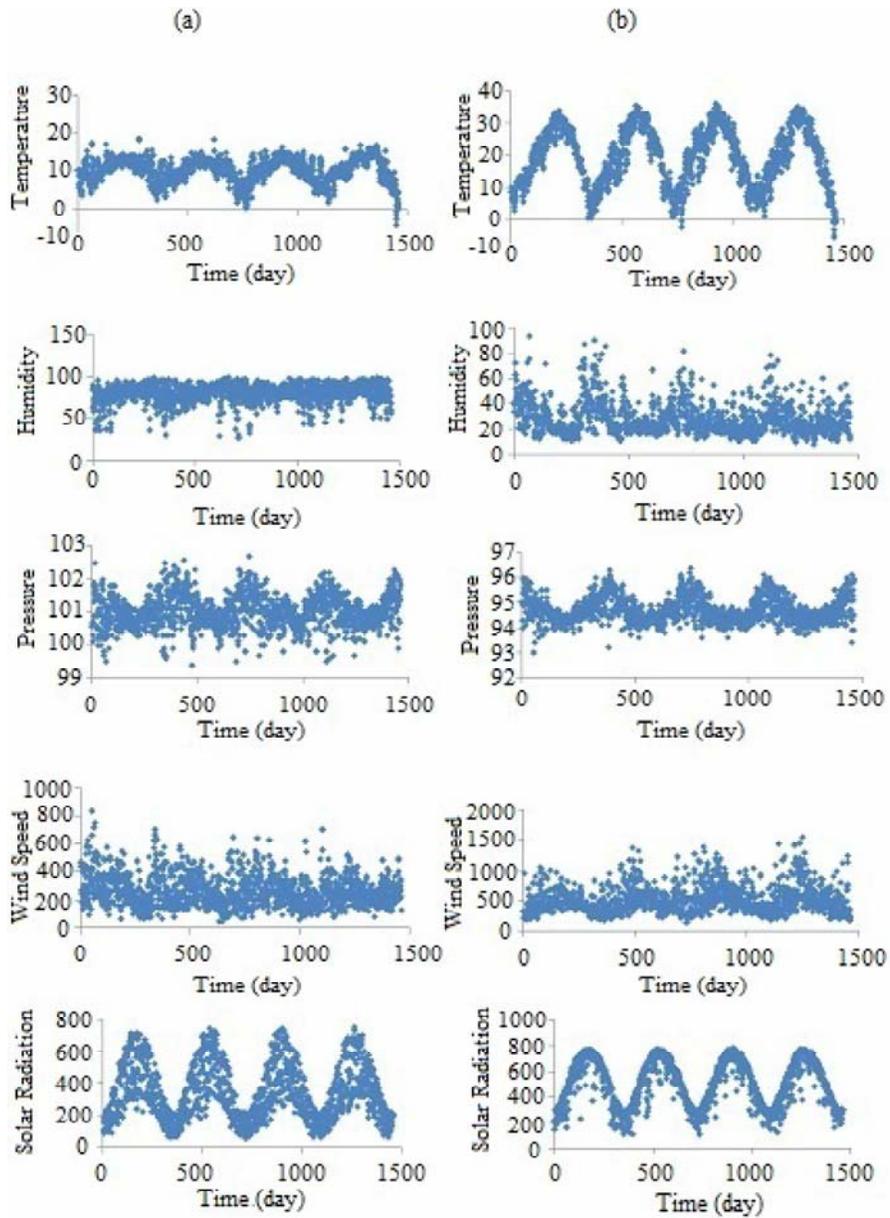
Climatic data from two stations are analyzed in this note: (i) Arcata-Eureka station (Latitude:  $40^{\circ}59'N$ , Longitude:  $124^{\circ}6'W$ , Elevation: 62m) and (ii) Daggett station (Latitude:  $34^{\circ}52'N$ , Longitude:  $116^{\circ}47'W$ , Elevation: 586m). Both are located in California, USA and cover a 4-year period from 1987 to 1990. Figure 1 shows their locations.

The climatic variables considered in this study include: daily temperature ( $T$ ), solar radiation ( $SR$ ), wind speed ( $W$ ), pressure ( $P$ ), and relative humidity ( $H$ ) and their values at the Arcata-Eureka and Daggett stations are shown in Figure 2.

In order to gain a preliminary insight into the variability of evaporation, recorded values at both sites are given in Figure 3. It is evident that the range at both sites are different (at Arcata-Eureka, the maximum evaporation is 7.8mm and the minimum is 0.1mm but at Daggett, the respective values are 26.9mm and 0.7mm). The difference in the variability of evaporation may be related to the Arcata-Eureka station being in a coastal area with moderately high humid and therefore the maximum amount of water losses in the form of evaporation is low; whereas high temperatures at the Daggett station cause moderately high evaporation. There is also a visual fluctuation associated with both time series, although its scale is more in the data for the Arcata-Eureka station than the Daggett station.

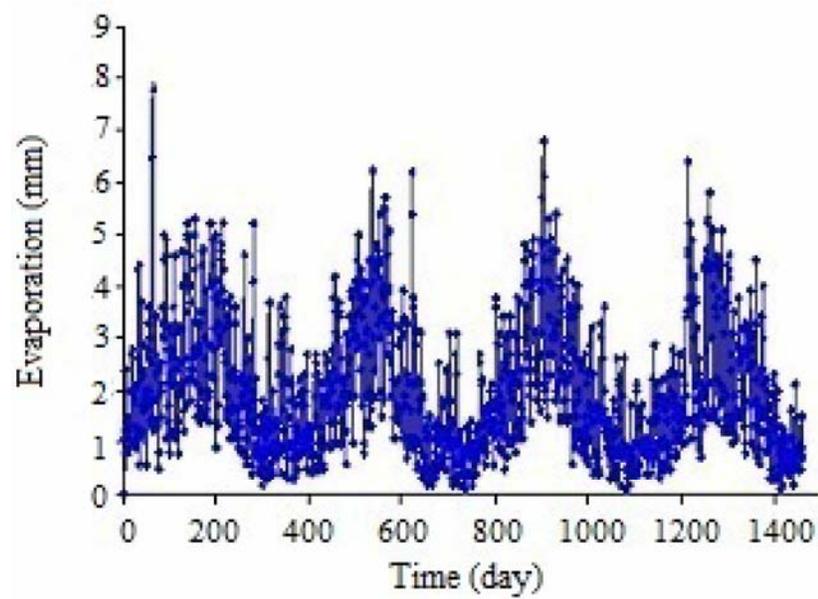


**Figure 1.** Location of the evaporation gauging sites.

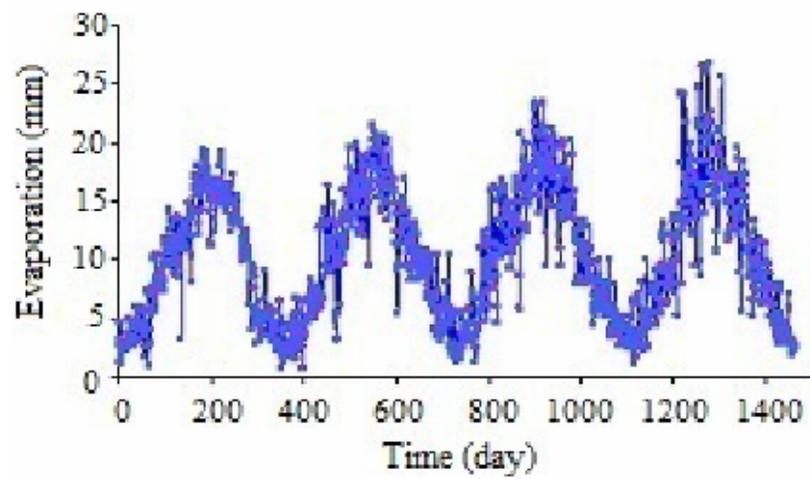


**Figure 2.** Daily temperature (°C), solar radiation (langley), wind speed (mil/h), pressure (kpa) and relative humidity (%) time series at (a) Arcata-Eureka station and (b) Daggett station.

(a)



(b)



**Figure 3.** Variability of pan evaporation at (a) Arcata-Eureka station and (b) Daggett station.

#### 4. Results and Discussion

The recorded data at two weather stations, the Arcata-Eureka station and the Daggett station, from 1987 to 1989 were used for training and the remaining data were used for testing the GP model using GeneXpro tools software (Ferreira [10, 11]). In order to understand the inherent variability in the data, the model was formulated by using five different input combinations, as defined by GP.1 to GP.5 as follows:

$$E = f(T); \quad (\text{GP.1})$$

$$E = f(T, SR); \quad (\text{GP.2})$$

$$E = f(T, SR, W); \quad (\text{GP.3})$$

$$E = f(T, SR, W, P); \quad (\text{GP.4})$$

$$E = f(T, SR, W, P, H). \quad (\text{GP.5})$$

Three statistical criteria are evaluated to assess the GP model performance: coefficient of determination ( $R^2$ ), mean square errors (MSE), and mean absolute errors (MARE). Legates and McCabe [25] emphasized that correlation based measures (such as above-mentioned parameters) should not be used for evaluating the model performance alone and other evaluation measures, such as adjusted coefficient of efficiency  $E_1$  and adjusted index of agreement  $d_1$ , should supplement model assessment. These two measures are defined as

$$E_1(\%) = 1 - \frac{\sum_{i=1}^n |E_{io} - E_{ie}|}{\sum_{i=1}^n |E_{io} - \bar{E}_o|}; \quad (2)$$

$$d_1(\%) = 1 - \frac{\sum_{i=1}^n |E_{io} - E_{ie}|}{\sum_{i=1}^n (|E_{ie} - \bar{E}_o| + |E_{io} - \bar{E}_o|)}, \quad (3)$$

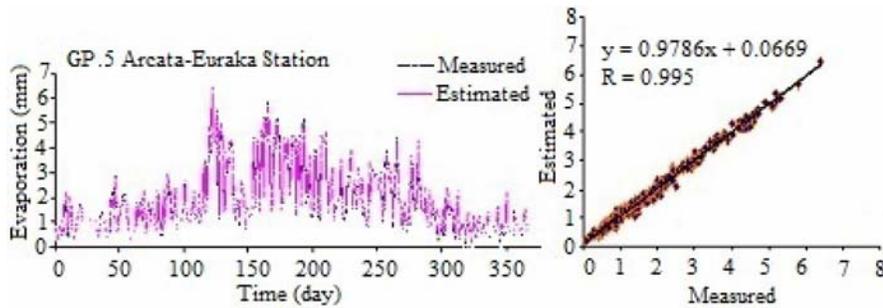
where  $E_{io}$ ,  $E_{ie}$ , and  $\bar{E}_o$  denote the observed, estimated, and mean of observed evaporation values, respectively.

The results of the models performance are presented in Table 1, according to which it is clear that as the number of independent variables in the formulation increases,  $R^2$  values increase towards 1.0 and at the same time MSE and MARE values decrease towards zero, indicating that the more the number of independent variables are represented in the model, the less variable the model is and the more well-behaved in terms of regression performance measures. The results also show that the model variability is more at the Daggett station than at the Arcata-Eureka station. However, the improvements are not monotonous, as a minor exception is observable not warranting explanation.

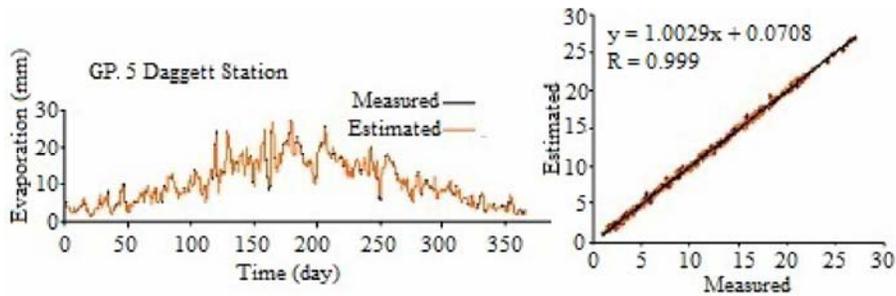
Figures 4 and 5 present the validated model results for GP.5 at both stations. The GP evidently reproduces evaporation ( $E$ ) in terms of directly measured independent variables of air temperature, solar radiation, wind speed, pressure, and humidity. However, at the Arcata-Eureka station, the modelled results suffer from more scatter against the observed values but the scatter is significantly less at the Daggett station. So, a systematic study is needed to gain an insight into the variability of the modelling results.

**Table 1.** GP results for different evaporation model formulations at both stations

Input combinations		Arcata-Eureka station			Daggett station		
		$R^2$	MSE	MARE	$R^2$	MSE	MARE
T	(GP.1)	0.287	1.15	79.41	0.766	8.14	21.42
T, SR	(GP.2)	0.890	0.19	34.6	0.874	4.41	15.7
T, SR, W	(GP.3)	0.926	0.12	26.0	0.980	0.68	8.75
T, SR, W, P	(GP.4)	0.924	0.12	25.79	0.981	0.65	8.5
T, SR, W, P, H	(GP.5)	0.968	0.000	0.021	0.953	0.016	0.088



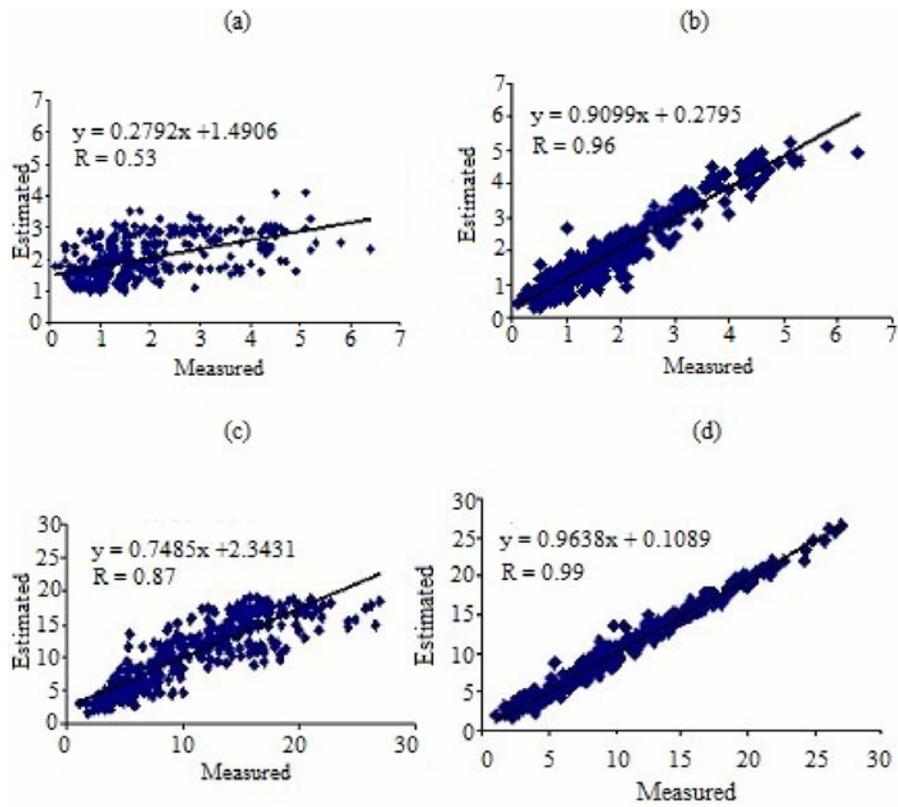
**Figure 4.** Modelled results at Arcata-Eureka gauging station using GP-5.



**Figure 5.** Modelled results at Daggett gauging station using GP-5.

As mentioned, the results showed that increasing the number of independent variables in the model improves its performance and this can further be visualized by Figure 6, displaying scatter diagrams of the results for GP.1 and GP.4 formulations at both gauging sites.

On the whole, the results provide evidence that GP can successfully be applied to modelling evaporation from climatic data, but the models contain inherent variability according to data quality at different sites and the number of independent variables used in the model. Although there is an emerging consensus that no single mathematical formulation is expected to outperform others and one expects that different models have different weaknesses and strengths, inter-comparison studies are still useful to obtain an insight for their comparative performances. The results produced in this study are therefore compared with those published by Kisi [19]. No detailed account of his model are presented here other than mentioning that the data was first studied by him using the three models of an artificial neural network (ANN) model, a neuro-fuzzy (NF), and a linear model by Stephen and Stewart [34] (the SS model). The corresponding results are reproduced in Table 2, which shows that generally the performance of GP is better or marginally better than the published results. However, differences may not be significant for practical problems. As the performance of the GP.5 model is better than other models, its statistics are provided in Table 3. Performance measures  $E_1$  (Equation (2)) and  $d_1$  (Equation (3)) have been proposed by Legates and McCabe [25] as the most appropriate relative error or goodness-of-fit measures. The value of  $E_1$  equal to 0.870 (for Arcata-Eureka station) confirms satisfactory estimation of evaporation by GP. Interpretation of  $d_1$  is similar to  $R^2$ , so high  $d_1$  values indicate a better agreement between the estimated and observed values.



**Figure 6.** Scatters in the GP.1 and GP.4 modelling results (the same results for GP.5 are given in Figures 4 and 5) – (a) Model: GP.1 station: Arcata-Eureka; (b) Model: GP.4 station: Arcata-Eureka; (c) Model: GP.1 station: Daggett; and (d) Model: GP.4 station: Daggett.

**Table 2.** Results of different evaporation models at both stations

Model	Input combinations	Arcata-Eureka station			Daggett station		
		$R^2$	MSE	MARE	$R^2$	MSE	MARE
GP.5	T, SR, W, P, H	0.968	0.000	0.021	0.953	0.016	0.088
ANN (Kisi [19])	T, SR, W, P, H	0.985	0.03	12.8	0.993	0.46	6.60
NF (Kisi [19])	T, SR, W, P, H	0.992	0.01	6.09	0.998	0.09	3.40
SS (Kisi [19])	T and SR	0.819	0.32	37.4	0.868	6.23	17.9

**Table 3.** Daily statistics of measured and estimated evaporation values (the statistics comprise:  $X_{mean}$  = mean,  $X_{max}$  = maximum,  $X_{min}$  = minimum,  $S_x$  = standard deviation,  $C_v$  = coefficient of variation,  $C_{sx}$  = skewness,  $E_1$  = adjusted coefficient of efficiency, and  $d_1$  = adjusted index of agreement)

Statistic	Arcata-Eureka station		Daggett station	
	$E_{io}$	$E_{ie}$ (with GP.5)	$E_{io}$	$E_{ie}$ (with GP.5)
$X_{mean}$	2.02	1.97	10.2	10.39
$X_{max}$	7.8	6.87	26.9	24.185
$X_{min}$	0.1	0.202	0.7	0.264
$S_x$	1.28	1.17	5.53	5.56
$C_v$	0.63	0.594	0.54	0.535
$C_{sx}$	0.91	0.934	0.25	0.224
$E_1$		0.870		0.822
$d_1$		0.888		0.909

As discussed earlier, the difference between the characteristics in the data between the two gauge stations have reflected expectedly in a significant difference between the performances of their respective models. For the Arcata-Eureka station, the maximum amount of water losses in the form of evaporation and the range of their variations are lower than those for the Daggett station. Therefore, the fitted GP.1 to

GP.5 models for the Arcata-Eureka station are considerably smoother than those for the Daggett station and this underlines the important role of data quality in the modelling results.

It may be remarked that the models investigated here are normally applied within deterministic frameworks in professional practices, which has encouraged the practice of comparing the actual with predicted values. However, this is a black-and-white approach for selecting the merits of a method and does not necessarily measure the impact on the decision. Such a technique is possible by risk-based approaches yet to emerge. Khatibi [18] discusses such a procedure for flood warning and more details are given in Defra [6] for flood risk management. Arguably, similar procedures are applicable to mathematical modelling, in general, but they are beyond the scope of this note.

It is also noted that total daily evaporation computed by GP.5 model is 3879mm for the test period, which is 0.37% lower than measured value (3907mm), while this value is 1.4%, 4.4%, and 12.5% for NF, ANN, and SS models, respectively.

One of the interesting aspects of GP is its “bottom-up” approach to derive the mathematical expression for time series, which goes beyond conventional functional expressions. This may be appreciated by considering the expressions derived by GP for GP.5 at both gauging stations. For the Arcata-Eureka station, the model formulation is

$$\begin{aligned}
 E = & (((d(2) - d(4)) + d(2)) - (d(4) - d(0))) * ((G1C1 / d(3)) / (G1C0 * d(4))) \\
 & + ((d(1) - d(4)) * (((d(0) + G2C1) / d(4)) / (d(3) * G2C0))) + ((d(1) - d(4)) * \\
 & (((d(0) + G3C0) / d(3)) / (G3C1 * d(3)))).
 \end{aligned}$$

Which the actual parameters are

$$d(0) = T, \quad d(1) = SR, \quad d(2) = W, \quad d(3) = P, \quad \text{and} \quad d(4) = H,$$

and the constant values are

$$G1C0 = -5.57016; G1C1 = -7.118103; G2C0 = 8.863892;$$

$$G2C1 = 5.663819; G3C0 = 8.031311; G3C1 = 6.25299.$$

After putting the corresponding values, the final equation becomes

$$E = 1.23 \left( \frac{T + 2(W - H)}{P.H} \right) + (SR - H) \left( \frac{T + 8.031}{6.253P^2} + \frac{T + 5.663}{8.863P.H} \right). \quad (4)$$

This is reflected in the results presented in Table 1 for GP.3 and GP.4. The difference between these two models is in the inclusion of pressure ( $P$ ) in GP.4. The table confirms that the inclusion of  $P$  make little improvement in its performance. For Daggett station, the formulation is

$$E = \frac{T}{P + H - 2T - 2.375} + \frac{0.098S_R - 6.971}{(\ln(S_R))^2} + \frac{-(S_R + 0.473)\sqrt[3]{H + S_R}}{W.S_R^{0.5}}. \quad (5)$$

In application of GP, it is usually observed that the program size (depth of parse tree) starts growing, which leads to producing nested functions. The ‘‘bloat’’ phenomena is a common problem with application of GP as it is the increasing in program size, which is not accompanied by any corresponding increase in model fitness. It has some practical effects, because the large programs are computationally expensive to evolve and later use can be hard to interpret. The nested functions give no sense about the physical basis of studied phenomena (Poli and McPhee [30]). For this, one should employ some penalization of complex models (limitation of the depth of the parse tree). There are several empirical models to control bloat, but the application of ‘‘parsimony coefficient’’ may be regarded as a proper method as described by Poli and McPhee [30]. In this way, the subtractive penalty term was selected and the new fitness function was defined as

$$f_p(x) = f(x) - w \cdot d, \quad (6)$$

which  $f_p(x)$  is the new fitness function (with penalty),  $f(x)$  is the fitness function,  $w$  is the parsimony coefficient, and  $d$  is the depth of parse tree.

In this paper, the “custom” user defined object in GeneXpro was tested and the penalty function was penalized by weighting the expression tree as detailed before. Although this method was helpful, but it could not remove un-necessary nesting. Therefore, the parsimony pressure tool of GeneXpro was applied to penalize tree depth. GeneXpro allows us to select models based not only on their raw fitness but also on their size by choosing a fitness function with parsimony pressure. Designing parsimonious solutions makes sense not only in terms of efficiency but also clarity, and modelers always try to balance both these factors. The fitness functions with parsimony pressure of GeneXpro allow us to achieve the same balance between efficiency and simplicity by favouring good solutions with more compact structures, as described in its manual.

The least that can be said about both of the above relationships is that they are counterintuitive and possibly beyond human perception. It may be that if the above are expanded by Taylor’s expansion theorem and manipulated through their definition in terms of mathematical series, the expressions derived by GP may possibly be simple deep down.

The fitted model described by both (4) and (5) are driven by the selections of basic operators, as detailed before. As the selection was successful, different ones were not tested. This possibly exposes a weakness that despite objectivity in science, the conventional models can be selections for convenience and other possibilities may be vast. Another interpretation is that mathematical modelers are not ideologically committed to any particular formulation and the data-driven GP techniques can find other possibilities.

## 5. Conclusion

A modelling study is reported here and develops a genetic programming (GP) model for simulating evaporation from daily climatic data for two gauging stations in California, USA. The overall results provide evidence that GP can successfully be applied for modelling

evaporation from climatic data, but the subsequent models vary according to data quality at different sites and the number of independent variables used in the model. The performance of the GP model is also compared with models fitted to the identical data and the inter-comparison study shows that generally the performance of the GP is better or marginally better than produced by ANN (artificial neural network), NF (neuro-fuzzy), and the SS (the linear model by Stephen and Stewart [34]) models. However, the differences may not be significant for practical problems.

### References

- [1] A. Aytek, Co-active neuro-fuzzy inference system for evapotranspiration modelling, *Soft Comput.* (2008). Doi: 10. 1007/s00500-008-0342-8
- [2] A. Aytek and M. Alp, An application of artificial intelligence for rainfall runoff modelling, *J. Earth Syst. Sci.* 117(2) (2008), 145-155.
- [3] A. Aytek and O. Kisi, A genetic programming approach to suspended sediment modelling, *J. Hydrol.* 351 (2008), 288-298.
- [4] V. Babovic and M. Keijzer, Rainfall runoff modelling based on genetic programming, *Nord. Hydrol.* 33 (2002), 331-343.
- [5] R. D. Burman, Intercontinental comparison of evaporation estimates, *J. Irrig. Drain Eng.* 102 (1976), 109-118.
- [6] Defra, Risk, Performance and Uncertainty in Flood and Coastal Defence: A Review; by P. B. Sayers, B. P. Gouldby, J. D. Simm, I. Meadowcroft and J. Hall, R&D Technical Report FD2302/TR1 (2003).
- [7] J. Doorenbos and W. O. Pruitt, Crop water requirements, FAO Irrigation and Drainage Paper, No. 24, FAO, Rome, 1977.
- [8] J. P. Drecourt, Application of Neural Networks and Genetic Programming to Rainfall Runoff Modelling, D2K Technical Report 0699-1-1, Danish Hydraulic Institute, Denmark, 1999.
- [9] EA, Estimation of Open Air Evaporation, R&D W6-048 (2001).
- [10] C. Ferreira, Gene Expression Programming in Problem Solving, Paper presented at the 6th Online World Conference on Soft Computing in Industrial Applications (invited tutorial), (2001a).
- [11] C. Ferreira, Gene expression programming: A new adaptive algorithm for solving problems, *Complex Sys.* 13(2) (2001b), 87-129.
- [12] H. Gavin and C. A. Agnew, Modelling actual, reference and equilibrium evaporation from a temperate wet grassland, *Hydrolo. Proc.* 18 (2004), 229-246.

- [13] M. A. Ghorbani, O. Makarynsky, J. Shiri and D. Makarynska, Genetic programming for sea level predictions in an island environment, *Int. J. Ocean Climate Sys.* 1(1) (2010), 27-35.
- [14] O. Giustolisi, Using GP to determine Chezy resistance coefficient in corrugated channels, *J. Hydroinfo.* (2004), 157-173.
- [15] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, 1989.
- [16] A. Guven, A. Aytek, M. I. Yuce and H. Aksoy, Genetic programming-based empirical model for daily reference evapotranspiration estimation, *Clean* 36(10-11) (2008), 905-912.
- [17] M. E. Jensen (eds), *Consumptive Use of Water and Irrigation Water Requirements*, ASCE, New York, NY, 1974.
- [18] R. Khatibi, Barriers Inherent in Flood Forecasting and their Treatments, In: D. W. Knight and A. Y. Shamseldin (eds), *River Basin Management for Flood Risk Mitigation* (2006).
- [19] O. Kisi, Daily pan evaporation modelling using a neuro-fuzzy computing technique, *J. Hydrol.* 329 (2006a), 636-646.
- [20] O. Kisi, Generalized regression neural networks for evapotranspiration modelling, *Hydrol. Sci. J.* 51(6) (2006b), 1092-1105.
- [21] O. Kisi, Evapotranspiration estimation using feed forward neural networks, *Nord. Hydrol.* 37(3) (2006c), 247-260.
- [22] O. Kisi, Evapotranspiration modelling from climate data using a neural computing technique, *Hydrol. Proc.* 21(6) (2007), 1925-1934.
- [23] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, MA, 1992.
- [24] M. Kumar, N. S. Raghuvanshi, R. Singh, W. W. Wallender and W. O. Pruitt, Estimating evapotranspiration using artificial neural networks, *J. Irrig. Drain Eng.* 128(4) (2002), 224-233.
- [25] D. R. Legates and G. J. McCabe, Evaluating the use of "goodness - of - fit" measures in hydrologic and hydroclimatic validation, *Water Resour. Res.* 35(1) (1999), 233-241.
- [26] E. T. Lincee, Climate and evaporation from crops, *J. Irrig. Drain Eng.* 93 (1967), 61-79.
- [27] S. Y. Liong, T. R. Gautam, S. T. Khu, V. Babovic, M. Keijzer and N. Muttil, Genetic programming: A new paradigm in rainfall runoff modelling, *J. Amer. Water Resour. Assoc.* 38(3) (2002), 705-718.
- [28] A. Moghaddamnia, M. Ghafari Gousheh, J. Piri, S. Amin and D. Han, Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques, *Adv. Water Resour.* 32 (2009), 88-97. DOI 10.1016/j.advwaters.2008.10.005

- [29] N. Muttill and S. Y. Liong, Improving Runoff Forecasting by Input Variable Selection in GP, Paper presented at the Proceedings of World Water Congress, ASCE, 2001.
- [30] R. Poli and N. F. McPhee, Covariant Parsimony Pressure for Genetic Programming, Technical Report CES-480, ISSN: 1744-8050, (2008).
- [31] A. M. A. Salih and U. Sendil, Evapotranspiration under extremely arid climates, *J. Irrig. Drain Eng.* 110(3) (1984), 289-303.
- [32] A. D. Savic, A. G. Walters and J. W. Davidson, A genetic programming approach to rainfall runoff modelling, *Water Resour. Manage* 13 (1999), 219-231.
- [33] A. F. Sheta and A. Mahmoud, Forecasting using Genetic Programming, Paper presented at the 33-rd Southeastern Symposium on System Theory, (2001), 343-347.
- [34] J. C. Stephen and E. H. Stewart, A comparison of procedures for computing evaporation and evapotranspiration, Publication 62, International Association of Scientific Hydrology, International Union of Geodynamics and Geophysics, Berkeley, CA, (1963), 123-133.
- [35] K. P. Sudheer, A. K. Goasin and K. S. Ramasastri, Estimating actual evapotranspiration from limited climate data using neural computing technique, *J. Irrig. Drain Eng.* 129(3) (2003), 214-218.
- [36] O. Terzi and M. E. Keskin, Evaporation estimation using gene expression programming, *J. Appl. Sci.* 5(3) (2005), 508-512.
- [37] S. Trajkovic, Temperature-based approaches for estimating reference evapotranspiration, *J. Irrig. Drain Eng.* 131(4) (2005), 316-323.

