MONTHLY REFERENCE EVAPOTRANSPIRATION MODELING USING GENE EXPRESSION PROGRAMMING FROM MINIMUM CLIMATIC DATA
Mattar, M. A.
Agric. Eng. Res. Inst. ARC, Egypt

ABSTRACT

Evapotranspiration is a key factor for water balance, irrigation scheduling, and crop yield. Even though, Penman-Monteith FAO-56 (PMF-56) equation had estimated the highest accuracy reference evapotranspiration (ET₀), it required complete climatic records, which may not be easily available. The present study is to develop and evaluate a gene expression programming (GEP) model for estimating mean monthly ET₀ using minimal number of climatic data. Climatic variables used to estimate ET₀ are maximum and minimum air temperature (T_max and T_min), mean relative humidity (RH), solar radiation (Rₛ), and wind speed at 2-m height (u₂). The data used in the analysis refer to 32 weather stations available at different locations in Egypt through the CLIMWAT database. The PMF-56 method was used as the reference standard for evaluating the developed GEP models based on statistical criteria such as: index of agreement (IA) and the root mean square error (RMSE). The results showed that the accuracy of the GEP model significantly improved when either RH or u₂ was used as additional input variables. The GEP model with the inputs: T_max, T_min, RH, and u₂ showed the highest IA (0.991 and 0.990) and the lowest RMSE (0.426 mm d⁻¹ and 0.430 mm d⁻¹) for training and testing sets, respectively. Comparing the results of GEP models with other empirical models showed that ET₀ values estimated by using the GEP models are more accurate. Accordingly, the GEP technique can be employed successfully in modeling ET₀ from the available climatic data and allowed for providing simple algebraic formulas.

Keywords: Evapotranspiration; Gene expression programming; Penman-Monteith; Empirical methods

INTRODUCTION

Over the last few decades, one of the critical problems encountered in water use scheduling was to decrease water availability in some parts of the world and to obtain accurate information on the agricultural demand. In Egypt the agriculture sector consumes about 85% of the available water resources (Ismail, 2002). Mismanagement of water resources through over-irrigation led to rapid land degradation due to the salinity, alkalinity, and water logging problems and then, reduced crops productively.

Overcoming these problems and improving the efficiency of water use is through an accurate irrigation schedule which identifies the required water at right time of irrigation. Irrigation scheduling aimed to replenish crop water requirements as quantified in evapotranspiration amounts under certain climatic conditions (Hunsaker and Pinter, 2003). Evapotranspiration is a term describing the transport of water into the atmosphere from surfaces, including evaporation from land surfaces and transpiration of vegetation (Allen et al., 1998; Hongjie et al., 2002; Kalluri et al., 2003).
Crop evapotranspiration ($ET_c$) can be estimated by multiplying reference evapotranspiration ($ET_o$) by the crop coefficient. This coefficient is determined based on the crop type, growth stage, canopy cover and density, and soil moisture content (Allen et al., 1998).

$ET_o$ values can be estimated or calculated using hydro-meteorological methods, which are either physically-based equations or empirical relationships between meteorological variables. One of these methods is Penman-Monteith method (physically-based), which estimates the monthly and daily $ET_o$ and was recommended by the Food and Agriculture Organization of the United Nations (FAO) as the standard equation (Allen et al., 1998; Naoum and Tsanis 2003 and Saghravani et al., 2009); it will be referred in the hereafter as FAO-56. Penman-Monteith FAO-56 (PMF-56) method is widely used in agricultural and environmental research to estimate the $ET_o$. PMF-56 coincides well with field observations and calibrates other models under various climates throughout the world (Allen et al., 1998; Kashyap and Panda 2001; Garcia et al., 2004; Popova et al., 2006).

A survey of the literature clearly indicated that the PMF-56 method is superior compared to all other commonly used empirical methods such as: Hargreaves-Semani (HS), Blaney-Criddle, Priestley-Taylor (PT), Jensen-Haise, Irmak (IR), and Turc (TR). Unfortunately, PMF-56 method requires complete climatic data, which may be unavailable or of low reliability in certain location, especially within developing countries and then is difficult to apply. In these cases, alternative methods that rely on a few climatic data are necessary.

In the last decade, intelligent computational techniques such as gene expression programming (GEP), artificial neural network (ANN), fuzzy logic (FL), adaptive neuro-fuzzy inference system (ANFIS), etc. have been proposed as alternative approaches. These models are computation techniques in diverse fields of hydrology engineering forecasting.

One alternative method is GEP that was invented by Ferreira (2001a). GEP is a computational technique that allows the solution of problems by automatically generating algorithms and expressions; this algorithm is used to implement symbolic regression in an attempt to find a mathematical function that fits a data set (Sakthivel et al., 2012). GEP is the natural development of genetic algorithms (GAs) (Goldberg, 1989) and genetic programming (GP) (Koza, 1992). In GAs, the individuals are linear strings of fixed length (chromosomes). In GP, the individuals are nonlinear entities of different sizes and shapes (parse trees). While, GEP involves computer programs (nonlinear entities) of different sizes and shapes (expression trees) encoded in linear strings of fixed lengths (chromosomes) (Ferreira, 2001a, b).

Various studies examined the applicability of GEP in hydrological and hydraulic modeling; however, only a few studies examine the application and duration of GEP in modeling the evapotranspiration process. Whigham and Crapper (2001) used GEP for rainfall-runoff modeling in Australia. Shiri and Kişi (2011a) compared GEP with ANFIS for predicting groundwater table depth fluctuations. Shiri and Kişi (2011b) compared GEP, ANFIS, and ANN to estimate daily pan evaporation by using recorded and estimated weather parameters. Shiri et al. (2012) applied GEP for modeling to estimate the daily
ET₀ for four climatic stations in Northern Spain over five-year period (1999–2003). They found that GEP model perform better than the ANFIS, PT, and HS models. Terzi (2013) compared GEP, ANFIS as an alternative approach to estimate daily pan evaporation in Turkey. Traore and Guven (2013) used GEP for modeling the ET₀ using routing weather data from tropical seasonally dry regions of West Africa in Burkina Faso. There is enough studies used GEP technique to estimate the ET₀ in Egypt.

The problem of incomplete or missing climatic data has a significant impact on the inaccurate estimation of the ET₀. Thus, ET₀ must be simulated using available minimal number of climatic variables. Therefore, the objectives of this study are to: (1) develop GEP models with limited climatic variables to predict the mean monthly ET₀, (2) evaluate the performance of GEP models developed with target PMF-56 set as the true reference values using statistical criteria, and (3) compare the accuracy of the results obtained from these models with the results of other empirical equations.

MATERIALS AND METHODS

1. Study area and data

Egypt is located between 22° and 31° 36’ 15” N, latitude and between 24° 41’ 49” and 36° 53’ 42” E, longitude; with area of about 1,002,450 km². The climate is characterized as Egypt essentially extremely dry all over the country except on the northern Mediterranean coast. In northern coast region, the average minimum temperature vary from 9.5 °C in winter to 23 °C in summer and average maximum temperature vary from 17 °C in winter to 32 °C in summer. In the central and the southern regions, the daytime temperature is higher especially in summers; the average maximum temperature exceeds 40 °C. Therefore, different regions were considered in the study to cover all the country areas.

The climatic data used in this study were collected from 32 weather stations with database known as CLIMWAT (Smith, 1993). The CLIMWAT database used in several studies of evapotranspiration, e.g., those reported by Allen (1993, 1996 and 1997), Temesgen et al. (1999), Droogers and Allen (2002), Valiantzas (2006), Trajkovic and Kolakovic (2009) and Todorovic et al. (2013). The spatial distribution of the selected stations is illustrated in Fig. 1. The data include the long-term mean monthly for the maximum and minimum air temperatures (T_max and T_min, °C), mean relative humidity (RH, %), solar radiation (R_s, Mj m⁻² d⁻¹), and wind speed at 2-m height (u_2, m/s) as well as ET₀ (mm d⁻¹) computed with the standard Penman-Monteith FAO-56 (PMF-56) equation.

In this study, gene expression programming (GEP) models take at most five input variables: T_max, T_min, RH, R_s, and u_2 while the ET₀ is the output variable. The input variables are divided into three sets: the training set consisted of 70% of data recorded by 27 of the weather stations; the remaining 30% of the data from the same weather stations was used as a set of test data to run the trained models. It is used to evaluate the generalisation abilities of the trained models. The third set will include the extracted data from the
remaining five weather stations (15% of the whole data); it will be used to validate and make the final check on the trained model's performance.

Fig. 1. Location of the selected weather stations in study area over Egypt

2. Gene expression programming

GEP is a new revolutionary member of the genetic computing family, benefiting from the genetic expression of the knowledge discovery technologies, owning to the merits of genetic algorithm (GA) and genetic programming (GP) that evolves computer programs. The basic difference among GEP, GA, and GP is due to the nature of the individuals (chromosomes). GEP has been recently introduced as a variant of GP (Ferreira, 2001b). In GEP, chromosomes are linear entities of fixed length that are expressing the genetic information encoded, and after that are converted to non-linear entities of varying sizes and length (expression trees or computer programs) at a later stage.

The chromosomes can have one or more genes, each gene encoding (genotype) a smaller subprogram. Each individual chromosome in the initial population through a random generation process is then expressed (phenotype). The fitness of each individual chromosome is evaluated against a set of fitness function equations (Ferreira, 2006). These chromosomes with
better solutions are then selected based on their fitness values to be reproduced with modification and re-evaluation by the genetic operators, such as mutation, inversion, transposition and recombination. This process is repeated until a correct solution (i.e., chromosomes) is found and the required accuracy is achieved (Ferreira, 2001a, b).

In GEP there are therefore two languages, the language of the genes and the language of expression trees. However, simple rules used to determine the structure of expression trees and their interactions; it is possible to infer immediately the phenotype given the sequence of a gene and vice versa. This bilingual and unequivocal system is called Karva language or K-expression.

3. Development of the GEP models

Computer software called GeneXproTools 5.0 program (GEPSOFT, 2014) was used to implement GEP. Once the training set is selected, one can say that the learning environment of the system is defined. The remaining data from the selected weather stations was used as the testing set. The major steps involved in developing GEP are illustrated in flow chart (Fig. 2). The first step is to identify the set of terminals to be used in the individual computer programs. The terminal set includes some of the independent variables: \( T_{\text{max}} \), \( T_{\text{min}} \), RH, \( R_s \), and \( u_2 \). The second step includes the set of functions. In this study, four basic arithmetic operators (+, −, ×, ÷) and some basic mathematical functions [\( \sqrt{\cdot} \), \( x^2 \), Power, \( \ln(x) \)] were used. These functions were applied to give better results for modelling evapotranspiration (Shiri et al., 2012). The terminals and the functions are the ingredients of the individual computer programs. The third step is to choose an appropriate fitness function. The root mean square error (RMSE) function was used to calculate the overall fitness of the evolved programs. The fourth step is selection of linking function. An addition (+) linking function was used to link the mathematical terms encoded in each gene (algebraic sub-trees). The next step is to choose the genetic operators. All genetic operators such as mutation, inversion, transposition and recombination or crossover were used. The parameters used in GEP modelling are summarized in Table 1. The final step is the determination of the criteria to terminate the run. The program was stopped when there was no improvement in the fitness function (RMSE) value. The final result is a mathematical expression that expresses the relationship between variables (inputs) and results (outputs), which can be used to predict the outcomes.

4. The Input Combinations

The combinations were used to estimate the mean monthly \( ET_o \) using the GEP technique is represented in Table 2. Eight GEP models were developed to test the performance of different combinations of climatic data. The first combination has two input climatic data, specifically, \( T_{\text{max}} \) and \( T_{\text{min}} \). This combination was designed as a temperature-based model similar to the Hargreaves and Samani (HS) model. The second combination was performed by inserting RH into the first combination. The third combination was performed by inserting \( R_s \) into the first combination. The Irmak (IR)
model presents the same inputs as those found in the third combination. The fourth combination was performed by inserting $u_2$ into the first combination. The fifth combination was performed by inserting $R_s$ into the second combination. This combination's inputs are identical to those of the Turc (TR) model. The sixth combination was performed by inserting $u_2$ into the second combination. The seventh combination was performed by inserting $u_2$ into the third combination. The eighth combination included all climatic data except the temperature data.
Fig. 2. The flowchart for the basic representation of the GEP algorithms

Table 1. Parameter settings for the GEP modeling

<table>
<thead>
<tr>
<th>GEP parameter</th>
<th>Setting of parameter</th>
<th>GEP parameter</th>
<th>Setting of parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of chromosomes</td>
<td>30</td>
<td>Inversion rate</td>
<td>0.00546</td>
</tr>
<tr>
<td>Head size</td>
<td>7.8</td>
<td>IS transposition rate</td>
<td>0.00546</td>
</tr>
<tr>
<td>Number of genes</td>
<td>3</td>
<td>RIS transposition rate</td>
<td>0.00546</td>
</tr>
<tr>
<td>Function set</td>
<td>+, -, *, +, √, x²,</td>
<td>Gene transposition rate</td>
<td>0.00277</td>
</tr>
<tr>
<td></td>
<td>Power, ln(x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitness function</td>
<td>RMSE</td>
<td>One-point recombination rate</td>
<td>0.00277</td>
</tr>
<tr>
<td>Linking function</td>
<td>addition</td>
<td>Two-point recombination rate</td>
<td>0.00277</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.00138</td>
<td>Gene recombination rate</td>
<td>0.00277</td>
</tr>
</tbody>
</table>

Table 2. The input combinations used for implementing the GEP models

<table>
<thead>
<tr>
<th>Model</th>
<th>Input combination</th>
<th>Model</th>
<th>Input combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEP1</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;</td>
<td>GEP5</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;, RH, R&lt;sub&gt;s&lt;/sub&gt;</td>
</tr>
<tr>
<td>GEP2</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;, RH</td>
<td>GEP6</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;, RH, u&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>GEP3</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;, R&lt;sub&gt;s&lt;/sub&gt;</td>
<td>GEP7</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;, R&lt;sub&gt;s&lt;/sub&gt;, u&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>GEP4</td>
<td>T&lt;sub&gt;max&lt;/sub&gt;, T&lt;sub&gt;min&lt;/sub&gt;, u&lt;sub&gt;2&lt;/sub&gt;</td>
<td>GEP8</td>
<td>RH, R&lt;sub&gt;s&lt;/sub&gt;, u&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

5. Reference evapotranspiration models

- Penman-Monteith FAO-56 equation

  The United Nations Food and Agriculture Organization (FAO) adopted the Penman-Monteith method to estimate ET<sub>o</sub> from climatic data; the details were presented in the FAO’s Irrigation and Drainage Paper no.56 (Allen et al. 1998); it would be referred to as FAO-56 in the hereafter. The Penman-Monteith FAO-56 (PMF-56) equation is highly rated across a wide range of climates (Allen et al., 1998), and used as a reference standard to evaluate the results of mathematical ET<sub>o</sub> models (Irmak et al., 2002; Liu et al., 2006; Zanetti et al., 2007; Landeras et al., 2008; Jain et al., 2008; Dai et al., 2009; Traore et al., 2010). The PMF-56 equation is given by Allen et al. (1998) as:

\[
ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}
\]  

(2)

where ET<sub>o</sub> is the reference evapotranspiration [mm d<sup>-1</sup>], \(\Delta\) is the slope of the saturation vapour pressure-temperature curve at mean air temperature [kPa °C<sup>-1</sup>], \(R_n\) is the net radiation [MJ m<sup>-2</sup> d<sup>-1</sup>], \(G\) is the soil heat flux [MJ m<sup>-2</sup> d<sup>-1</sup>], \(\gamma\) is the psychometric constant [kPa °C<sup>-1</sup>], \(T\) is the mean monthly air temperature at 2-m height [°C], \(u_2\) is the wind speed at 2 m height [m s<sup>-1</sup>], \(e_s\) is the saturation vapour pressure [kPa], and \(e_a\) is the actual vapour pressure [kPa]. All aforementioned parameters were calculated using equations reported in Allen et al. (1998).

- HS equation

  Mean monthly ET<sub>o</sub> was estimated by Hargreaves and Samani (1985). The HS equation is given by:

\[575\]
\[ ET_0 = 0.0023 \frac{R_\lambda}{\lambda} \sqrt{(T_{\text{max}} - T_{\text{min}})} (T + 17.8) \]  

where \( ET_0 \) in [mm d\(^{-1}\)], \( R_\lambda \) is the extraterrestrial radiation [MJ m\(^{-2}\) d\(^{-1}\)], and \( \lambda \) is the latent heat of vaporization [MJ kg\(^{-1}\)].

- IR equation
  
  Irmaek et al. (2003) developed an equation for estimating \( ET_0 \) in [mm d\(^{-1}\)] in the form:

  \[ ET_0 = -0.611 + 0.149R_s + 0.079T \]

where \( R_s \) is the solar or shortwave radiation [MJ m\(^{-2}\) d\(^{-1}\)].

- TR equation
  
  TR equation (Turc, 1961) was developed in Western Europe, and has been used to some extent in the United States (e.g. Amatya et al., 1995). This equation is written as:

  \[ ET_0 = a_T 0.013 \frac{T}{T+15} + \frac{238856R_s + 50}{\lambda} \]

Value of \( a_T \) equals unity when the relative humidity is higher than 50% otherwise \( a_T \) is given by: \( a_T = 1 + \frac{50 + RH}{70} \)

6. Statistical evaluation

The \( ET_0 \) estimated using the GEP and the other empirical models were compared with \( ET_0 \) estimated using PMF-56 equation. The coefficient of determination \( (R^2) \), index of agreement \( (IA) \), root mean square error \( (RMSE) \), overall index of model performance \( (OI) \) and mean absolute error \( (MAE) \) were used for evaluating the performance of these models. Definitions of these parameters are as follows:

\[ R^2 = \frac{\sum (E_i - \bar{E})(C_i - \bar{C})^2}{\sum (E_i - \bar{E})^2 \cdot \sum (C_i - \bar{C})^2} \]  

\[ IA = 1 - \frac{\sum (E_i - C_i)^2}{\sum (C_i - \bar{E})^2 + \sum (E_i - \bar{E})^2} \]

\[ RMSE = \sqrt{\frac{\sum (E_i - C_i)^2}{n}} \]

\[ OI = 1 - \frac{1}{2} \left( 2 \cdot \frac{RMSE}{E_m - E_n} + \frac{\sum (E_i - C_i)^2}{\sum (E_i - \bar{E})^2} \right) \]

\[ MAE = \frac{\sum |E_i - C_i|}{n} \]

where \( E_i \) is the PMF-56 estimated \( ET_0 \) value, \( C_i \) is the GEP estimated \( ET_0 \) value, \( n \) is the number of data points, \( \bar{E} \) is the average PMF-56 estimated value, \( \bar{C} \) is the average GEP estimated value, and \( E_m \) and \( E_n \) are the maximum and minimum PMF-56 estimated value.

\( R^2 \) measures the degree of correlation among the PMF-56 estimated and GEP estimated \( ET_0 \) values with values close to 1.0 indicating good model performance. \( IA \) takes values from 0 (the worst fit) to 1 (the perfect fit) (Legates and McCabe, 1999; Moriasi et al., 2007). The lower the RMSE, the more accurate the GEP estimation is. \( OI \) ranges from \(-\infty \) to 1 and value of 1 denotes a perfect fit between the PMF-56 and GEP estimated \( ET_0 \) values (Alazba et al., 576
2012: Mattar et al., 2015; Mattar and Alamoud, 2015). MAE measures the average magnitude of the errors in a set of forecasts without considering their directions. MAE ranges from 0 to ∞, and its lower values are better.

RESULTS AND DISCUSSION

1. Performance analysis of GEP
For the analysis, 230 data points were collected from 27 weather stations, taken for the training process. The rest of the values (94 data points) are used for testing process for the generalization capability of the models. Several climatic variable combinations have been tried as input sets to the GEP for the models’ formulation. For developing the GEP-based empirical models, the program (GeneXproTools 5.0) was run until there was no longer significant improvement in the performance of the models. The final equations obtained from GEP are shown in Table 3 for eight combinations. Description for the expression trees of the GEP6 formulation (Table 3), and of the GEP6 model with the best performance after the training process is illustrated in Fig. 3. The four-input parameters to the GEP6 are: T_{max}, T_{min}, RH, and u2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Model equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEP1</td>
<td>ET_o = \left( \sqrt{T_{max} - 4.16} + \left( \frac{2.92}{T_{min}} \right)^{0.75} \right)^{0.75} + \ln \left( 9.31 \sqrt{T_{min} - 14.59} \right) + \frac{9.72 \times 10^{-11} (T_{max} - 4.9)}{T_{min}^{3.07}}</td>
</tr>
<tr>
<td>GEP2</td>
<td>ET_o = 3.84 + \left( \frac{13.29 - RH}{26.92 + T_{min}^{1.3}} \right) + \ln \left( \frac{T_{max}^{2}}{1.33^{1.3}} + \ln RH \right)^{-1} - \frac{T_{min}^{3.07}}{1.27^{0.71}}</td>
</tr>
<tr>
<td>GEP3</td>
<td>ET_o = \left( \ln \left( R_{h} \left( T_{min} - 1.66 \right)^{0.13} \right)^{0.75} + \left( \ln \left( \frac{4.43}{T_{max}} \right) \right) + \left( \ln \left( \frac{5.61 - T_{max}}{0.6} \right) \right) \right)^{-1} + \ln \left( \frac{6.07}{R_{h}} \right)</td>
</tr>
<tr>
<td>GEP4</td>
<td>ET_o = \left( \frac{2.5 - T_{max}}{9.72} \right)^{0.76} + \frac{T_{max}}{T_{min}^{0.85}} + \ln \left( \frac{R_{h}}{u_{2}} \right) + \left( \frac{T_{max} - 2.16 \times RH}{u_{2}} \right)^{0.85} + \ln \left( \frac{9.31}{R_{h}} \right)</td>
</tr>
<tr>
<td>GEP5</td>
<td>ET_o = \left( \frac{T_{max}^{0.75}}{\sqrt{T_{max} - 17.59}^{0.85} + 1.07^{RH}} \right) + \frac{1.89 T_{max}}{\sqrt{T_{max} + RH - 1.46 T_{min}}} + \ln \left( \frac{25.73 RH}{T_{max}^{0.75}} \right)</td>
</tr>
<tr>
<td>GEP6</td>
<td>ET_o = \left[ \frac{T_{min}}{\ln \left( RH + 0.46 u_{2} \right)} \right] + \frac{1.89 T_{max}}{\sqrt{1.21 T_{max} + RH}} + \frac{86.4 u_{2}}{25.73 RH + T_{min}}</td>
</tr>
<tr>
<td>GEP7</td>
<td>ET_o = \left[ \frac{21917.86 + 449.97 R_{h}^{2} + 449.97 R_{h} u_{2}}{0.01 T_{min}^{0.85} + 8.93 T_{max} - 8.93 T_{min}} \right] + \frac{u_{2}}{\sqrt{T_{min} - T_{max}}} - 13.73</td>
</tr>
<tr>
<td>GEP8</td>
<td>ET_o = \frac{113.08 + 86.4 u_{2}}{21.34 + R_{h}} + \frac{\sqrt{RH}}{R_{h}^{0.01}} + \ln \left( \frac{0.75 \sqrt{RH - 6.8}}{RH u_{2}} \right)</td>
</tr>
</tbody>
</table>

Table 3. Equations derived from the GEP models
Expression tree (ET)

Fig. 3. Expression tree for the GEP6 formulation (d0, d1, d2, and d3 denote \(T_{\text{max}}, T_{\text{min}}, \text{RH}, \) and \(u_2\), respectively; \(c_0\) and \(c_1\) are constants)

In Figs. (4 and 5) the GEP model estimates were compared with the PMF-56 estimated data via scatter plots for the training and testing sets, respectively. Additionally, a linear regression is applied for evaluating the models’ performance statistically. The IA, RMSE, OI and MAE statistics of each GEP model during training and testing processes were given in Table 4. As shown in Figs. 4 and 5, the GEP2 and GEP4-GEP8 models results lie around the 45° straight line (perfect line) implying, that there are no bias effects than GEP1 and GEP 3 models. The GEP1 and GEP3 showed a bad correlation with PMF-56 estimated ET\(_o\). It can be also observed in the fit lines equations (assume that the equation is \(y = \alpha_0 x + \alpha_1\)) that the slopes (\(\alpha\)) are closer to one and the intercepts (\(\alpha_1\)) almost reach zero for all GEP models, except GEP1 and GEP3. The values of IA and OI for the GEP1 and GEP3 were lower than those for the rest of the models; however, the RMSE and MAE values were higher in the training and testing processes (Table 4).

The GEP1 (its inputs were the air temperatures only) showed the largest intercepts (1.106 and 1.065) and lowest slopes (0.767 and 0.78) with \(R^2\) of 0.768 and 0.784 in training and testing processes, respectively. The GEP3 in training and testing processes showed \(R^2\) values 5.6% and 3.19% higher than those resulted from the GEP1. The IA values for GEP3 increased by 1.5% and 0.85% for the training and testing processes, respectively, while the RMSE values decreased by 9.81% and 5.89%. The OI values increased by 2.76% and 1.3%; the MAE values decreased by 13.82% and 9.53%. Therefore, the ET\(_o\) estimation accuracy was slightly improved when \(R_s\) data were added to GEP1. This is similar to results reported by Traore and Guven (2013) for Burkina Faso. They
reported that the performance slightly vary for GEP models with $T_{\text{max}}$, $T_{\text{min}}$, and $R_s$ ($R^2 = 0.598$) than those from temperature data alone ($R^2 = 0.588$).

\[
y = 0.767x + 1.106 \quad R^2 = 0.768
\]

\[
y = 0.908x + 0.44 \quad R^2 = 0.912
\]

\[
y = 0.807x + 0.919 \quad R^2 = 0.811
\]

\[
y = 0.919x + 0.388 \quad R^2 = 0.92
\]

\[
y = 0.962x + 0.177 \quad R^2 = 0.951
\]

\[
y = 0.96x + 0.195 \quad R^2 = 0.963
\]

\[
y = 0.936x + 0.301 \quad R^2 = 0.941
\]

\[
y = 0.946x + 0.255 \quad R^2 = 0.948
\]

Fig. 4. Values of the ET$_o$ estimated by using the PMF-56 against the GEP models for the training data set.
Fig. 5. Values of the ET₀ estimated by using the PMF-56 against the GEP models for the testing data set.
Table 4. Performance statistics of the GEP models in the training and testing periods

<table>
<thead>
<tr>
<th>Model</th>
<th>Training period</th>
<th>Testing period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IA</td>
<td>RMSE, mm d(^{-1})</td>
</tr>
<tr>
<td>GEP1</td>
<td>0.931</td>
<td>1.070</td>
</tr>
<tr>
<td>GEP2</td>
<td>0.976</td>
<td>0.660</td>
</tr>
<tr>
<td>GEP3</td>
<td>0.945</td>
<td>0.965</td>
</tr>
<tr>
<td>GEP4</td>
<td>0.979</td>
<td>0.630</td>
</tr>
<tr>
<td>GEP5</td>
<td>0.988</td>
<td>0.490</td>
</tr>
<tr>
<td>GEP6</td>
<td>0.991</td>
<td>0.426</td>
</tr>
<tr>
<td>GEP7</td>
<td>0.984</td>
<td>0.540</td>
</tr>
<tr>
<td>GEP8</td>
<td>0.986</td>
<td>0.508</td>
</tr>
</tbody>
</table>

Conversely, replacing R\(_s\) data with RH or u\(_2\) data resulted in a better performance for GEP2 and GEP4 than GEP3. The intercept values for the GEP2 and GEP4 was close to one and the slop was close to zero. The R\(^2\) values were 0.915 and 0.925, on average, for the GEP2 and GEP4, respectively, during training and testing processes; increased by 12.96% and 14.2% than those from the GEP2. This is also confirmed by the IA, RMSE, OI, and MAE values in Table 4. These models showed that the effect of RH and u\(_2\) variable is significant, contributing in 48.58% and 18.07%, respectively. These results agreed with the results of Fisher et al. (2005) and of Xiaoying and Erda (2005); they found that ET\(_o\) is sensitive to u\(_2\).

It is also observed that the R\(_s\) data when incorporated together with RH or u\(_2\) data, as input parameters to GEP5 or GEP7, gave good estimates. In terms of R\(^2\) value, it was improved for GEP5 and GEP7, on average, by 17.47% and 16.91%, respectively, compared to GEP3 in the training and testing processes. The GEP5 for training and testing processes, on average, had an IA and OI values that were about 4.55% and 11.35% more accurate than those from the GEP3. The corresponding values of IA and OI for GEP7 increased by about 4.39% and 10.88% compared to GEP3. Moreover, RMSE and MAE for GEP5 or GEP7 also decreased to about half the value recorded for GEP3.

GEP6 added RH and u\(_2\) to GEP1, significantly increased the performance, resulted in largest slope (0.96 and 0.95) and lowest interception (0.195 and 0.239) in the training and testing processes, respectively. In addition, low scattering of the data points was observed around the perfect line. The values of R\(^2\) for GEP6 increased drastically by 25.39% and 22.7% than that for the GEP1. The IA and OI values were very close to one, while the RMSE and MAE values were extremely low, indicating excellent agreement between the ET\(_o\) estimated from PMF-56 and GEP6. This implies that the RH and u\(_2\) (importance ratio of 23.52% and 22.96%, respectively) data are more effective in estimating the ET\(_o\). This result was consistent with Kìşi and Ozturk (2007), Traore et al. (2010), Ozkan et al. (2011), and Huo et al. (2012). While when T\(_{\max}\) and T\(_{\min}\) data are not available, the GEP8 is the most suited model with the slopes close to one (0.946 and 0.954),
interception of 0.225 and $R^2$ of 0.948. This was indicated by the higher values of IA (0.986) and OI (0.948) for the GEP8 were close to one for the training set, while the RMSE (0.508 mm d$^{-1}$) and MAE (0.421 mm d$^{-1}$) values were close to zero. Similarly, for the testing data set, the lower values of the RMSE and MAE were 0.546 mm d$^{-1}$ and 0.444 mm d$^{-1}$, respectively (close to zero), while the AI and OI values were 0.984 and 0.942, respectively (close to one). The GEP models behaved well with both the training and test data, as there were no significant differences in the $R^2$, AI, RMSE, OI, and MAE values given by the training and testing datasets.

2. Spatial validation of the GEP models

The validation was conducted in order to confirm the accuracy of the presented GEP models; the reference for this validation is the $E_{\text{T}}_o$ estimated by using the PMF-56 equation. Table 5 shows the statistical analysis of the $E_{\text{T}}_o$ for the validation performances with the five selected stations (Mersa-Matruh, Ismailia, Baharia, Asyut, and Kom-Ombo). In Table 5, for Mersa-Matruh station, the GEP5 estimates were closer to the corresponding PMF-56 estimated $E_{\text{T}}_o$ values with a higher IA and OI values (0.994 and 0.968), and lower PMSE and MAE values (0.165 and 0.143) than those of the other models. The performance of the GEP2 and GEP8 are ranked as the second and third best models, respectively. However, GEP4 and GEP7 showed low accuracy in estimating $E_{\text{T}}_o$ as they contain the temperature and $u_2$ data with no RH data. The IA, OI, RMSE, and MAE (Table 6) for the GEP4 and GEP7 confirmed the poor performance of them. It can be observed that the GEP2, GEP5, and GEP8 models, RH, is a significant variable affects the $E_{\text{T}}_o$ because Mersa-Matruh station is located in a humid climate. Shiri et al. (2013) also found that models including RH give more accurate results in humid region because of the higher influence of RH on $E_{\text{T}}_o$ in such regions. The GEP2 model seems to be slightly better than the GEP8 in robustness. The other GEP model overestimated mean monthly PMF-56 $E_{\text{T}}_o$ at Mersa-Matruh station with an IA from 0.738 to 0.835, RMSE from 1.114 to 1.899 mm d$^{-1}$, OI from 0.217 to 0.724, and MAE from 0.701 to 1.839 mm d$^{-1}$.

For the three stations: Ismailia, Baharia, and Asyut, the estimated $E_{\text{T}}_o$ values from GEP models showed the same trend of the training and testing data. It is clear from Table 5, it can be observed that the estimated $E_{\text{T}}_o$ values show much better accuracies in GEP6 than the other GEP models. The GEP6 gave values, on average, for IA of 0.993, RMSE of 0.353 mm d$^{-1}$, OI of 0.958, and MAE of 0.306 mm d$^{-1}$. The IA and OI values were very close to one while RMSE and MAE values were close to zero, indicating an excellent agreement between the PMF-56 results and estimated results from the GEP6.

As shown in Table 5, in Kom-Ombo station, all the GEP models presented large errors, except GEP4, GEP6, and GEP7. These models offer a very good performance, which may be due to the amalgamated effects of $u_2$. This station, however, exhibited high precision, with an IA of 0.959, RMSE of 0.688 mm d$^{-1}$, OI of 0.871, and MAE of 0.632 mm d$^{-1}$ for GEP4. GEP7 is ranked as the second best model. GEP2 and GEP5 provided the worst results because of the presence of RH. The GEP2 produced $E_{\text{T}}_o$ values with
40.77\% and 131.26\% lower accuracy in IA and OI values, than those from the GEP4. The RMSE and MAE for GEP2 increased by a factor of 3.3 (from 0.688 and 0.632 mm d\(^{-1}\), respectively, in GEP4 to 2.967 and 2.723 mm d\(^{-1}\)). The IA and OI values from GEP5 were 27.57\% and 89.80\% lower than those from GEP7. The MARE and MAE values for the GEP5 were almost 2.28 and 2.43 times that of the values for the GEP7, respectively. This shows the importance of \( u_\lambda \) data to calculate ET\(_{o}\) at a station located in arid region.

Table 5. Performance statistics of the GEP models in the validation periods for the stations: Mersa-Matruh, Ismailia, Baharia, Asyut, and Kom-Ombo

<table>
<thead>
<tr>
<th>Statistical parameters</th>
<th>GEP1</th>
<th>GEP2</th>
<th>GEP3</th>
<th>GEP4</th>
<th>GEP5</th>
<th>GEP6</th>
<th>GEP7</th>
<th>GEP8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mersa-Matruh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0.738</td>
<td>0.923</td>
<td>0.822</td>
<td>0.619</td>
<td>0.994</td>
<td>0.835</td>
<td>0.703</td>
<td>0.910</td>
</tr>
<tr>
<td>RMSE, mm d(^{-1})</td>
<td>1.219</td>
<td>0.576</td>
<td>1.114</td>
<td>1.899</td>
<td>0.165</td>
<td>0.868</td>
<td>1.620</td>
<td>0.753</td>
</tr>
<tr>
<td>OI</td>
<td>0.217</td>
<td>0.847</td>
<td>0.608</td>
<td>0.047</td>
<td>0.968</td>
<td>0.724</td>
<td>0.267</td>
<td>0.776</td>
</tr>
<tr>
<td>MAE, mm d(^{-1})</td>
<td>1.014</td>
<td>0.482</td>
<td>1.072</td>
<td>1.839</td>
<td>0.143</td>
<td>0.701</td>
<td>1.602</td>
<td>0.623</td>
</tr>
<tr>
<td>Ismailia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0.931</td>
<td>0.970</td>
<td>0.965</td>
<td>0.966</td>
<td>0.992</td>
<td>0.992</td>
<td>0.973</td>
<td>0.989</td>
</tr>
<tr>
<td>RMSE, mm d(^{-1})</td>
<td>0.946</td>
<td>0.645</td>
<td>0.714</td>
<td>0.716</td>
<td>0.353</td>
<td>0.328</td>
<td>0.643</td>
<td>0.403</td>
</tr>
<tr>
<td>OI</td>
<td>0.796</td>
<td>0.901</td>
<td>0.885</td>
<td>0.885</td>
<td>0.956</td>
<td>0.960</td>
<td>0.901</td>
<td>0.948</td>
</tr>
<tr>
<td>MAE, mm d(^{-1})</td>
<td>0.795</td>
<td>0.506</td>
<td>0.567</td>
<td>0.577</td>
<td>0.300</td>
<td>0.298</td>
<td>0.546</td>
<td>0.307</td>
</tr>
<tr>
<td>Baharia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0.954</td>
<td>0.974</td>
<td>0.975</td>
<td>0.973</td>
<td>0.982</td>
<td>0.994</td>
<td>0.984</td>
<td>0.981</td>
</tr>
<tr>
<td>RMSE, mm d(^{-1})</td>
<td>0.868</td>
<td>0.669</td>
<td>0.638</td>
<td>0.651</td>
<td>0.587</td>
<td>0.329</td>
<td>0.503</td>
<td>0.573</td>
</tr>
<tr>
<td>OI</td>
<td>0.833</td>
<td>0.902</td>
<td>0.908</td>
<td>0.906</td>
<td>0.918</td>
<td>0.962</td>
<td>0.934</td>
<td>0.921</td>
</tr>
<tr>
<td>MAE, mm d(^{-1})</td>
<td>0.684</td>
<td>0.567</td>
<td>0.500</td>
<td>0.514</td>
<td>0.485</td>
<td>0.285</td>
<td>0.397</td>
<td>0.490</td>
</tr>
<tr>
<td>Asyut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0.973</td>
<td>0.976</td>
<td>0.986</td>
<td>0.981</td>
<td>0.976</td>
<td>0.992</td>
<td>0.992</td>
<td>0.984</td>
</tr>
<tr>
<td>RMSE, mm d(^{-1})</td>
<td>0.696</td>
<td>0.745</td>
<td>0.501</td>
<td>0.591</td>
<td>0.755</td>
<td>0.402</td>
<td>0.380</td>
<td>0.572</td>
</tr>
<tr>
<td>OI</td>
<td>0.895</td>
<td>0.883</td>
<td>0.934</td>
<td>0.917</td>
<td>0.881</td>
<td>0.951</td>
<td>0.955</td>
<td>0.921</td>
</tr>
<tr>
<td>MAE, mm d(^{-1})</td>
<td>0.588</td>
<td>0.656</td>
<td>0.435</td>
<td>0.510</td>
<td>0.643</td>
<td>0.334</td>
<td>0.324</td>
<td>0.501</td>
</tr>
<tr>
<td>Kom-Ombo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0.793</td>
<td>0.568</td>
<td>0.764</td>
<td>0.959</td>
<td>0.691</td>
<td>0.899</td>
<td>0.954</td>
<td>0.758</td>
</tr>
<tr>
<td>RMSE, mm d(^{-1})</td>
<td>1.904</td>
<td>2.967</td>
<td>2.254</td>
<td>0.688</td>
<td>2.478</td>
<td>1.072</td>
<td>0.756</td>
<td>2.044</td>
</tr>
<tr>
<td>OI</td>
<td>-0.071</td>
<td>0.281</td>
<td>0.221</td>
<td>0.871</td>
<td>0.087</td>
<td>0.757</td>
<td>0.853</td>
<td>0.337</td>
</tr>
<tr>
<td>MAE, mm d(^{-1})</td>
<td>1.552</td>
<td>2.723</td>
<td>1.647</td>
<td>0.632</td>
<td>2.323</td>
<td>0.995</td>
<td>0.678</td>
<td>1.753</td>
</tr>
<tr>
<td>All stations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>0.909</td>
<td>0.900</td>
<td>0.913</td>
<td>0.926</td>
<td>0.933</td>
<td>0.972</td>
<td>0.947</td>
<td>0.940</td>
</tr>
<tr>
<td>RMSE, mm d(^{-1})</td>
<td>1.203</td>
<td>1.468</td>
<td>1.226</td>
<td>1.036</td>
<td>1.201</td>
<td>1.067</td>
<td>0.895</td>
<td>1.055</td>
</tr>
<tr>
<td>OI</td>
<td>0.741</td>
<td>0.667</td>
<td>0.756</td>
<td>0.813</td>
<td>0.762</td>
<td>0.904</td>
<td>0.852</td>
<td>0.807</td>
</tr>
<tr>
<td>MAE, mm d(^{-1})</td>
<td>0.927</td>
<td>1.014</td>
<td>0.840</td>
<td>0.814</td>
<td>0.779</td>
<td>0.522</td>
<td>0.709</td>
<td>0.735</td>
</tr>
</tbody>
</table>

For all the selected stations for the validation performance, the number of data points used were 60 in total. Validation of the GEP6 showed the best result compared to the other models during the validation of dataset (Table 5). This result was reflected in the IA, RMSE, OI, and MAE values. The IA and OI were very high (0.972 and 0.904), and values of RMSE and MAE
were the lowest (0.676 mm d⁻¹ and 0.522 mm d⁻¹) which means that the estimated \( ET_0 \) values by the GEP6 were closely related to those estimated by the PMF-56.

3. Performance of GEP models compared to empirical methods

Table 6 presented comparison for the performances of the GEP and the empirical models: HS, IR, and TR. It is obviously seen from the table that the GEP1 (temperature-based model) estimates has better accuracy than that of the HS model. According to R² values, it can be said that the GEP1 has the best estimates. This confirms the IA and OI statistics that were about 4.85% and 7.8%, respectively, more accurate than that from the HS. The RMSE and MAE values for the GEP1 model were 15.82% and 14.5% more accurate.

Table 6 shows that the GEP3 has a better generalization ability in \( ET_0 \) modeling when compared to the IR (radiation-based method). The IR model produced \( ET_0 \) values that were 14.26% and 16.53% less accurate in IA and OI values, respectively, than those from the GEP3 model. In terms of RMSE and MAE, the GEP3 was 25.78% and 24.51% more accurate than the IR model, respectively.

It is clear from Table 6 that the GEP5 is suited model with a higher R² value of 0.918 than those of the TR model. The TR model produced the poorest performance with IA and OI that decreases drastically from 0.978 to 0.535 and from 0.925 to -0.312 (i.e. 45.3% and 133.74% decrease), respectively. The MARE and MAE values, respectively, for the TR model were almost 5.09 and 7.25 times that of the values for the GEP5.

Comparing Table 6, the performance indices revealed that the GEP models were superior among the overall empirical models. This is because the GEP models had the highest IA and OI, and the lowest RMSE and MAE for all stations. Therefore, GEP technique can contribute for solving climatic data unavailability problem in Egypt.

### Table 6. Statistical performance of the developed GEP models and empirical models

<table>
<thead>
<tr>
<th>Models</th>
<th>Statistical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
</tr>
<tr>
<td>GEP1</td>
<td>0.758</td>
</tr>
<tr>
<td>HS</td>
<td>0.686</td>
</tr>
<tr>
<td>GEP3</td>
<td>0.789</td>
</tr>
<tr>
<td>IR</td>
<td>0.743</td>
</tr>
<tr>
<td>GEP5</td>
<td>0.918</td>
</tr>
<tr>
<td>TR</td>
<td>0.844</td>
</tr>
</tbody>
</table>

### CONCLUSIONS

It is difficult to calculate a reference evapotranspiration (\( ET_0 \)) using climatic data which are missing sometimes in the weather stations. Gene expression programming (GEP) technique for estimating the mean monthly \( ET_0 \) using climatic variables has been developed in this paper. Eight input combinations from the climatic data were collected 27 weather stations and
were used in training and testing processes. Other five stations were used for the validation process. The developed GEP models, Hargreaves and Samani (HS), Irmak (IR), and Turc (TR) were employed to study their performance comparing with the Penman-Monteith FAO-56 (PMF-56) that is the standard equation. The comparison was made according to various statistic measures.

The availability of climatic data such as mean relative humidity (RH), wind speed at 2 m height \((u_2)\), and solar radiation in some developing countries (e.g., Egypt) is sometimes difficult. Therefore, the GEP model including only temperatures (maximum air temperature, \(T_{\text{max}}\) and minimum air temperature, \(T_{\text{min}}\)) variables can be suited for estimating \(ET_o\). The RH and \(u_2\) are the most effective variables and highly recommended for modeling \(ET_o\). As \(T_{\text{max}}\) and \(T_{\text{min}}\) variables are excluded, the GEP model has good performance. On the other hand, the GEP models have a better performance when compared to the empirical HS, IR, and TR models. The TR produced significant underestimates. Consequently, GEP model is fairly a promising approach and a powerful tool to be used without recourse to the full set of climatic data requirement for accurately estimate \(ET_o\) in Egypt.

REFERENCES


588
نموذج البخار-نتح المرجعي الشهري باستخدام البرمجة التعبيرية الجينية أقل بيانات المناخية

محمد عبد العزيز مطر
معهد بحوث الهندسة الزراعية – مركز البحوث الزراعية

البخار-نتح عامل أساسي لتحقيق التوازن المائي، وجدولة الري، وانتاج المحصول. وعلى الرغم من قدرة معادلة بمان مونث فاود-56 (PMF-56) للبخار-نتح المرجعي ($ET_0$) بتقدير متوسط $ET_0$ الشهري باستخدام أقل عدد من البيانات المناخية، فإنها تتطلب بيانات عاجلة، والتي قد لا تكون متاحة بسهولة. هذه الدراسة هي تطوير تقييم نموذج البرمجة التعبيرية الجينية (GEP) لتقدير متوسط $ET_0$ الشهري باستخدام أقل عدد من البيانات المناخية. المتغيرات المناخية المستخدمة لتقدير $ET_0$ هي درجة حرارة الهواء العظمى ($T_{max}$) والصغرى ($T_{min}$)، ومتوسط الرطوبة النسبية ($RH$)، والأشعة الشمسية ($R_{s}$)، وسرعة الرياح ($u_2$). تشير البيانات المستخدمة في التحليل إلى 32 محطة رصد مناخية متاحة في مواقع مختلفة في مصر من خلال قاعدة البيانات CLIMWAT. تم استخدام معادلة PMF-56 كطريقة قياسية مرجعية لتقديم نماذج GEP المطورة على أساس معايير إحصائية مثل: مؤشر التوافق (IA)، وجدور متوسط مربع الخطأ (RMSE)، وأظهر النتائج تحسن دقة النموذج بشكل ملحوظ عندما يستخدم أي من $RH$ أو $u_2$ كمتغيرات مدخلة إضافية. أظهر النموذج GEP RMSE بشكل أ hại من النموذج $ET_0$ o 0.991 و 0.990 وأعلى IA ($0.426$ و $0.430$ مم يوم$^{-1}$ و يوم$^{-1}$) لمجموعات التدريب والاختبار، على التوالي. بمقارنة نتائج GEP نماذج GEP مع نماذج تجريبية أخرى، وجد أن النموذج المقدر باستخدام GEP عالي، يمكن استخدام تقنية GEP بنجاح في نسب $ET_0$ من البيانات المناخية المتوقعة. وكذلك تسمح تقنية GEP بتوفير الصيغ الجبرية البسيطة.