

Comparison of three data-driven techniques in modelling the evapotranspiration process

I. El-Baroudy, A. Elshorbagy, S. K. Carey, O. Giustolisi and D. Savic

ABSTRACT

Evapotranspiration is one of the main components of the hydrological cycle as it accounts for more than two-thirds of the precipitation losses at the global scale. Reliable estimates of actual evapotranspiration are crucial for effective watershed modelling and water resource management, yet direct measurements of the evapotranspiration losses are difficult and expensive. This research explores the utility and effectiveness of data-driven techniques in modelling actual evapotranspiration measured by an eddy covariance system. The authors compare the Evolutionary Polynomial Regression (EPR) performance to Artificial Neural Networks (ANNs) and Genetic Programming (GP). Furthermore, this research investigates the effect of previous states (time lags) of the meteorological input variables on characterizing actual evapotranspiration. The models developed using the EPR, based on the two case studies at the Mildred Lake mine, AB, Canada provided comparable performance to the models of GP and ANNs. Moreover, the EPR provided simpler models than those developed by the other data-driven techniques, particularly in one of the case studies. The inclusion of the previous states of the input variables slightly enhanced the performance of the developed model, which in turn indicates the dynamic nature of the evapotranspiration process.

Key words | actual evapotranspiration, data driven techniques, eddy covariance, evolutionary polynomial regression, genetic programming, neural networks

I. El-Baroudy

A. Elshorbagy (corresponding author)
Department of Civil and Geological Engineering,
Centre for Advanced Numerical Simulation
(CANSIM),
University of Saskatchewan,
Saskatoon, Saskatchewan,
Canada S7N 5A9
E-mail: amin.elshorbagy@usask.ca

S. K. Carey

Department of Geography and Environmental
Studies, Carleton University,
Ottawa, Ontario,
Canada K1S 5B6

O. Giustolisi

Department of Civil and Environmental
Engineering, Engineering Faculty,
Technical University of Bari,
via Turismo, 8, Taranto 74100,
Italy

D. Savic

School of Engineering, Computer Science and
Mathematics, University of Exeter,
Exeter, Devon EX4 4QF,
UK

INTRODUCTION

Evapotranspiration is a term used to describe all the processes by which water moves from land or vegetation surfaces to the atmosphere. It is one of the main components of the hydrological cycle as it accounts for more than two-thirds of precipitation losses globally (Dingman 2002; Fisher *et al.* 2005). Therefore, reliable estimates of the actual evapotranspiration (AET) losses are crucial to a wide range of hydrological modelling problems (Xu & Singh 1998) and practical applications. Direct measurement of evapotranspiration losses are difficult, expensive and often unreliable (Gasca-Tucker *et al.* 2007). Moreover, the interdependence between different components of the soil-vegetation-atmosphere system hinders the proper estimation of the evapotranspiration losses, making it one

of the most challenging tasks in hydrological modelling (Xu & Singh 2005).

The evapotranspiration process depends on many factors related to the land, plant and atmosphere. Atmospheric factors include radiation, air temperature, humidity and wind speed, while the plant factors include plant type, variety and development stage. The state of land surface wetness (or near-surface soil moisture) is the main land factor that affects the quantification of evapotranspiration (Hornberger *et al.* 1998; Savenije 2004). There are a wide variety of methods used to estimate potential evapotranspiration (PET) based on atmospheric forcing and assuming unlimited water supply at the surface. These models assume that there is no limit to the water provided to the

atmosphere. However, accurate modelling of AET requires the consideration of different physical processes such as soil moisture availability, the moderating role of vegetation, the ability of turbulence to move eddies from the surface and available energy.

Measuring AET requires expensive and data-intensive systems, such as the eddy covariance system (EC) (Baldocchi *et al.* 1988), although less intensive aerodynamic and combination methods are often employed (Stull 1988). Considering instrumentation requirements and the complexities and inherent assumptions of process-based AET models, this research explores the utility of using data-driven techniques to estimate AET from micrometeorological data. This work is an extension of the research work performed by Parasuraman *et al.* (2007a), where two data-driven modelling methods were compared, i.e. Genetic Programming (GP) (Koza 1992) and Artificial Neural Networks (ANNs), with the classical Penman–Monteith (PM) method in modelling EC-measured Latent Heat (LE). Details of ANNs and GP methods are briefly discussed in the third section of this paper to avoid repetition of the information cited by Parasuraman *et al.* (2007a). More focus is given to the third data-driven method, i.e. Evolutionary Polynomial Regression (EPR).

Despite the wide literature on the application of the data-driven techniques to various hydrological processes, the work of Parasuraman *et al.* (2006, 2007a) is the first application of such techniques to the EC-measured AET. In this paper, an alternate data-driven technique referred to as Evolutionary Polynomial Regression (EPR) is used, which is capable of modelling time series or regression type data containing information about any physical processes through the combination of evolutionary algorithms and numerical regression. EPR is a data-driven technique that incorporates the main features of numerical regression together with symbolic regression (Giustolisi *et al.* 2007). It produces flexible structure polynomial models where each monomial can include user-defined functions, which in turn improves the physical interpretation of the considered phenomenon (Giustolisi & Savic 2006). EPR therefore makes use of the knowledge concerning the phenomenon under consideration (as opposed to ANNs) derived from known physical laws, which is incorporated into the training process. Moreover, EPR

tends to evolve more parsimonious structures over long complex structures (as opposed to GP). As a result, EPR is more powerful than GP in finding the model constants and avoids producing functions that grow in length over time (Davidson *et al.* 1999).

The method develops pseudo-polynomial models characterizing the process under consideration (Laucelli *et al.* 2005; Giustolisi & Savic 2006). EPR is used in this research to model evapotranspiration using the environmental variables: net radiation (NR), ground temperature (GT), air temperature (AT), wind speed (WS) and relative humidity (RH). This paper uses the same case studies used by Parasuraman *et al.* (2007a) in the Mildred Lake mine, located to the north of Fort McMurray, Alberta, Canada to provide a comprehensive comparison with the GP and ANN methods. The performance of the modelling tools (ANN, GP and EPR) is evaluated based on the same three error measures used: (i) root mean squared error (RMSE), (ii) mean absolute relative error (MARE) and (iii) correlation coefficient (R), which measure different aspects of model performance (Dawson *et al.* 2007).

The main objectives of this paper are (i) to compare the utility of the EPR to the other data-driven techniques in modelling the evapotranspiration process and (ii) to investigate the effect of using previous states (time lags) of the input variables on modelling the EC-measured AET.

EVAPOTRANSPIRATION PROCESS

Physics of the evapotranspiration process

Evapotranspiration is the conversion of liquid water into vapour, either from a free surface of water or soils (evaporation) or through plant stomata (transpiration). The theoretical basis of evapotranspiration is based upon the seminal work of Penman (1948), reviewed both scientifically and operationally by Brutsaert (1982).

The source of energy for evaporation is net radiation (NR), which is the sum of short and long-wave radiation components. The supply of water depends upon soil wetness and any moderating role that plants play through regulation of their stomata. Turbulent transport of water depends upon wind speed and thermal instability of the

surface layer (Obukhov 1946). The processes associated with the above components are well defined for the soil-vegetation-atmosphere interface, and are characterized by radiative fluxes, soil moisture, surface soil temperature and turbulent transfer into the atmosphere. Consequently, the rate of AET for a given supply of NR corresponds to a particular combination of these variables, of which there are many possible combinations to satisfy the surface energy balance defined here as (Hornberger *et al.* 1998):

$$\frac{dQ}{dt} = NR - G - H - LE \quad (1)$$

where dQ/dt is the change in energy storage per unit time, NR is net radiation, G is ground heat, H is sensible heat and LE is latent heat.

The latent heat of vaporization is a measure of the energy required to change a unit mass of water from liquid to water vapour, under constant pressure and temperature. The value of the latent heat varies slightly over normal temperature ranges (around 20°C), therefore the value of $2.5 \times 10^6 \text{ J/kg}$ is used as the representative value of the latent heat of vaporization. As a result, as radiation expressed in $\text{J/m}^2 \text{ day}$, LE can be converted to equivalent evaporation in mm/day by using a conversion factor equal to the inverse of the latent heat of vaporization. Division of LE by the latent heat of vaporization provides AET, and these values can be considered equivalent.

In their work, Parasuraman *et al.* (2007a) developed GP and ANN models that showed better performance than the PM model while providing insight into the evapotranspiration process and investigated the contribution of different atmospheric and surface factors (Parasuraman *et al.* 2007a). In the present study, the authors extended this application to another prominent data-driven modelling method, named Evolutionary Polynomial Regression (EPR), proposed by Giustolisi & Savic (2006). In addition to modelling the evapotranspiration process using the different atmospheric and surface factors, the effect of lag time in input variables on the characterization of the evapotranspiration losses is explored.

The net radiant energy (NR) received by the land surface is partly absorbed by the water to enable the phase change from liquid state to vapour state (evapotranspiration) in the form of latent heat flux (LE). The other part of

the received energy is consumed in warming up the land surface and the near-surface atmosphere in the form of atmospheric sensible heat flux (H) and land surface heat flux (G). The partitioning of LE among plants, vegetation and soil is strongly affected by the wetting conditions of the surface soil (soil moisture) (Gutierrez & Meinzer 1994). Surface soil moisture also has an indirect correlation with the surface air temperature, which is supported by the findings of Wilson & Baldocchi (2000) and MingFeng & HuiJun (2007). Their results show the existence of 'decadal' and seasonal climate co-variabilities and significant interactions between surface air temperature and soil moisture, which in turn affect the evapotranspiration process.

Evapotranspiration measurement methods

Measurement of AET is complex with many assumptions, and there exist a variety of methods employed by hydrologists (Brutsaert 2005). Most measurements are indirect and of limited spatial extent. Additionally, complications arise when attempting to separate evaporation and transpiration components, or above- and below-canopy AET. The only direct method to measure AET is the eddy covariance (EC) method (Baldocchi *et al.* 1988), based on analysis of high-frequency wind and scalar (i.e. water, heat, CO_2) data. Although a direct method, EC suffers from recognized systematic errors related to sampling frequency and energy balance closure (Finnegan *et al.* 2003; Fisher *et al.* 2005). Regardless, EC is the adopted method for international flux research programs (i.e. Baldocchi *et al.* 2001; Margolis *et al.* 2006), and has a reasonable accuracy range from ± 15 to $\pm 20\%$ for hourly evapotranspiration measurements and up to ± 8 to $\pm 10\%$ for longer periods (Eichinger *et al.* 2003; Strangeways 2003).

DATA-DRIVEN TECHNIQUES

Physically based models are typically used to model hydrological processes. These models draw their names from the physical principles that are used to explain the behaviour of the different processes of the hydrological systems. Data-driven models (black-box models), on the other hand, are based on a limited knowledge of the

physical processes and depend on the information contained within the data to characterize the functional relationship between the system inputs and outputs, especially for complex systems (Giustolisi & Savic 2006). These methods use specific datasets representing limited realizations of the system states; however, they can be used to draw reasonable generalizations of the process under consideration.

The last two decades have witnessed the rise of a variety of data-driven techniques due to the fast growing computational power of personal computers. Solomatine (2002) provided an overview of the widely used data-driven techniques in the field of water resources management and control and showed the utility of techniques such as the ANNs, machine learning (ML), M5 model trees and fuzzy rule-based systems (FRBS) in river flow predictions, rainfall–runoff process, modelling channel networks and many other applications.

Artificial neural networks

ANNs are computational models imitating the behaviour of biological neural networks in the human brain. They consist of an interconnected group of artificial neurons which process the available information through these connections and the neurons' parameters to produce complex global behaviour of non-linear systems. Typically, an ANN model consists of an input layer which receives the problem inputs. Through hidden layers (which contain a certain number of nodes that are usually determined by the user), these inputs are processed to provide the output layer with the values predicted by the network (model output). The nodes within the adjacent layers are connected by links together with synaptic weight for each link to express the relative connection strength of two nodes in predicting the input–output relationship (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a).

Parasuraman *et al.* (2007a) used feed-forward neural networks (FF-NNs) architecture with J input neurons, K hidden neurons and L output neurons. This architecture uses the tan-sigmoidal activation function in the hidden layer and the linear activation function in the output layer. Typical ANN modelling is performed on two steps: training

and testing. In the training process, connection weights are minimized based on a certain cost function with normalization of inputs and outputs to enhance the training process. The mean sum of squares of the network errors (MSE) is the cost function used to minimize the connection weights. These weights are evaluated in the testing step using a different dataset.

Enhancement in the convergence in the training step is achieved using the Bayesian-regularization back-propagation algorithm developed by Demuth & Beale (2001). This algorithm develops networks with smaller weights and biases and, consequently, a smoother response. ANN architecture and algorithms are found in Parasuraman *et al.* (2007a).

The objective of this paper is to explore the utility and effectiveness of data-driven techniques in modelling actual evapotranspiration measured by the eddy covariance system. Therefore, identical datasets are used for training and testing the ANN and EPR. The main criterion for the data-driven technique effectiveness is its ability to capture the meteorological input variables that can be used to characterize actual evapotranspiration.

The ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b) provides a thorough survey of the different successful applications of ANNs to hydrological problems, e.g. rainfall–runoff modelling, modelling stream flows, water quality modelling, groundwater modelling, estimating precipitation, forecasting river stages and many other applications. The ability of the ANNs to recognize the relationship between the input and output variables without explicit physical consideration makes them attractive for modelling complex hydrological processes. A major drawback of the ANNs is the absence of clear physical concepts, which makes the choice of their structure, training and definition of efficient behaviour a subjective choice (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a).

Genetic programming

GP is another widely used data-driven technique. GP (Cramer 1985) is an evolutionary algorithm (an ML technique) that mimics the biological evolution process (of natural selection) in an effort to build computer

models capable of simulating complex physical processes, e.g. non-linear and spatially and temporally variable processes (Koza 1992). GP uses a tree-like structure such as decision trees to represent its 'concepts and its interpreter as a computer program' (Banzhaf *et al.* 1998). It is therefore considered a superset of all other ML representations, which enables GP to produce any solution that is produced by any other ML system.

Although GP is computationally intensive, especially for generating programs that are capable of simulating complex processes, it has a major advantage in that it handles symbolic expressions. Genetic symbolic regression (GSR) is a special application of GP which is similar to mathematical numerical regression, where it finds a mathematical expression that fits the training dataset. GSR uses two sets of variables: the functional set and terminal set (Koza 1992). The terminal set consists of the independent variables and the constants and the functional set consists of the basic mathematical operators {+, −, *, /, log, power, sin, etc.}. GSR starts by constructing a population of mathematical models based on different combinations of the functional and terminal sets. Each model (individual) in the population is a potential solution to the problem. The mathematical models are usually coded in a parse tree form.

Parasuraman *et al.* (2007a) used the ramped half-and-half method to initialize the population of parse trees with various sizes to provide a good coverage of the search space. The fitness of each individual (model) in the population is then evaluated based on certain objective function. High fitness models (individuals) have a high chance of being carried over to the next generation. Genetic operators (selection, crossover and mutation) are used to develop new sets of models from the parent generation (Koza 1992; Babovic & Keijzer 2000). Newly developed models are called offspring, as they represent the biological product of the parent generation and form the basis for the next generation. In this study, the fitness measure (objective function) is evaluated based on the root mean squared error (RMSE).

Selection of the candidate models based on their fitness value is carried out by several methods such as truncation selection, tournament selection, fitness proportional selection and roulette wheel selection (Koza 1992).

Parasuraman *et al.* (2007a) used the roulette wheel selection in developing their models. The roulette wheel is constructed by partitioning the roulette wheel space between different models based on the proportion of their fitness values. The selection process therefore ensures that the models with higher fitness have a higher chance of being selected to the next generation. The process of selection leads to the creation of a temporary population, called the mating pool. Genetic operators are applied to the models in the mating pool to produce the next generation of models. Evaluation of the new generation models is based on the same fitness function where the selection and mating are repeated until a certain number of iterations is reached or until a certain fitness value is achieved. Various GP operators used in this study are listed by Parasuraman *et al.* (2007a).

Babovic & Abbot (1997) demonstrated the utility of the evolutionary algorithms, e.g. GP and evolution strategies (ES), in extracting the information content of the hydraulic data to develop a representative model of complicated physical processes. In their work, four different hydrological applications of GP (rainfall–runoff modelling, sediment transport modelling, salt intrusion and steady flow studies) were presented. Similar to any data-driven technique, GP has its own limitations. The major problem is the deterioration of the prediction ability of the developed model with the increase in the prediction horizon, which is a common problem in any modelling method. The adverse consequences of this problem can be mitigated by incorporating the GP technique, with its capabilities in short-term forecasting, with knowledge-based techniques that use the accumulated knowledge of the process under consideration.

Several applications of the GP technique in hydrology exist in the literature. Parasuraman *et al.* (2007b), for example, explored the utility of the GP technique to develop pedotransfer functions (PTFs) for estimating the saturated hydraulic conductivity (K_s) from soil content (sand, silt and clay) and the bulk density. Babovic & Keijzer (2002) addressed the utility of GP in developing rainfall–runoff models on the basis of hydro-meteorological models as well as in combination with other conventional models, i.e. conceptual models. It was reported that the GP models gave more insights into the functional relationships between

different input variables resulting in more robust models. Similarly, Jayawardena *et al.* (2005) compared the GP technique in modelling rainfall–runoff processes to the traditional modelling approaches. They used the GP technique to predict the runoff from three catchments in Hong Kong and two catchments in southern China, and showed that the GP techniques involved simple models that enabled the quantification of the significance of different input variables for prediction.

Evolutionary polynomial regression

This paper utilizes EPR, a data-driven technique that models time series data containing information about physical processes (Giustolisi & Savic 2006). EPR combines the power of evolutionary algorithms with numerical regression to develop polynomial models combining the independent variables together with the user-defined function as follows (Lauccelli *et al.* 2005):

$$\hat{Y} = \sum_{i=1}^m F(X, f(x), a_i) + a_0 \quad (2)$$

where \hat{Y} is the EPR estimated dependant variable, $F(\cdot)$ is the polynomial function constructed by EPR, X is the independent variable matrix, $f(\cdot)$ is a user-defined function, a_i is the coefficient of the i th term in the polynomial, a_0 is the bias and m is the total number of the polynomial terms.

The user-defined function is included to enhance the characterization of the response (dependent) variable. EPR develops symbolic models in two main stages: (i) structure identification and (ii) parameter estimation, where it uses genetic algorithm (GA) simple search method to search in the model structure space. As for any other numerical regression, EPR uses the least squares (LS) method to estimate the parameters of the selected model structure based on the performed GA search. Applications of EPR are found in Savic *et al.* (2006), Doglioni *et al.* (2007) and Giustolisi *et al.* (2007). The search progresses by using the standard GA operators: crossover and mutation. This type of search is not exhaustive as it is practically impossible to conduct such a search on an infinite search space (Lauccelli *et al.* 2005).

This research makes use of the EPR toolbox (Lauccelli *et al.* 2005). It is a multi-objective implementation of EPR in the sense that it produces several models which are the best trade-off between fitness to training data and parsimony (see Giustolisi *et al.* 2006). The simplicity of the toolbox encouraged the authors to explore the utility of the EPR technique in modelling the actual evapotranspiration process using NR, GT, AT, WS and RH as input variables due to their physical controls on the evaporation process. The effect of previous time states of the input variables (time lag) is also investigated and compared to the static simulation models (models developed without considering a time lag in input variables).

The EPR tool performs three types of regression: dynamic, static and classification. Dynamic modelling is used to model systems that have memory or, in other words, systems in which the present state depends on the previous states of other input variables. On the other hand, static systems are those which are not influenced by the previous states of input variables. Classification modelling is a special type of static modelling in which the model output is an integer (Lauccelli *et al.* 2005). The readers may refer to the user manual for the details of the EPR toolbox and the different components of its simple interface (Lauccelli *et al.* 2005).

Optimal configuration of data-driven models

Optimal configuration of any data-driven model using ANN, GP or EPR and their key parameters is usually determined by a trial-and-error method. However, the significant input variables for characterizing a process are usually determined by linear cross-correlation (Bowden *et al.* 2005). Linear cross-correlation cannot be used for capturing the nonlinear dependence between the inputs and the output. Parasuraman *et al.* (2007a) did not report the occurrence of any pre-convergence problems. The authors of this study did not encounter pre-convergence problems in utilizing EPR. In this study, similar to Parasuraman *et al.* (2007a), the optimal EPR models are determined by performing a systematic search of different model configurations and user-adjustable parameters, based on the three selected error measures.

CASE STUDIES

The first case study is the South Bison Hill (SBH) and the second is the South West Sand Storage (SWSS), located several kilometres from each other. Both research sites are oil sands reclamation sites located at Syncrude Canada Ltd's Mildred Lake mine north of Fort McMurray, AB, Canada. The climate is a sub-humid continental, which is characterized by cold winters and warm summers. The mean annual precipitation is 456 mm, of which 342 mm occurs as rainfall mostly between June and August.

As outlined in [Parasuraman *et al.* \(2007a\)](#), the SBH was constructed with waste-rock material from oil sands mining in the period 1980–1996. The area of the SBH is 2 km². It rises 60 m above the surrounding landscape and has a large flat top several hundred metres in diameter. The underlying shale is covered by a 0.2 m layer of peat mineral mix on top of a 0.8 m layer of glacial till to reclaim the overburden soil surface. This area was then fertilized and seeded to agronomic barley and planted to white spruce and aspen in the summer of 1999 ([Carey 2006](#)). The top of the SBH is dominated by foxtail barley and other minor species of fireweed.

The second site SWSS is currently the largest operational tailings dam in the world, holding more than 46,000 m³ of material, covering 25 km² and standing approximately 40 m high with a slope of 20:1. Soils consist of mine tailings sand overlain with 0.4–0.8 m of topsoil which is a mixture of peat and secondary mineral soil with a clay loam texture. Both vegetation species and composition vary across the SWSS. Dominant groundcover includes horsetail, fireweed, sow thistle and white and yellow sweet clover in addition to different tree and shrub species ([Parasuraman *et al.* 2007a](#)). Data were obtained from micrometeorological towers located at each site, details of which can be found in [Parasuraman *et al.* \(2007b\)](#).

Modelling of LE was based on daytime hourly data (08:00–20:00) to exclude negative instances of ET. This is in addition to the recognized systematic error in night-time readings, caused by the inability of the sonic anemometer to resolve fine-scale eddies which occur during the night ([Fisher *et al.* 2005](#)). The split of data series, at each day, will cause some spurious effects in the case of using lag time. However, these effects are not expected to have any serious consequences on the developed models as they do not

exceed more than 8% of the daily lags. For one-hour lag, there are 12 one-hour lags a day, where one of these lags occurs at the split of each day. Therefore, it is one (1) lag out of the twelve (12) lags that would cause the spurious effect. The twelve-hour split at each day is not expected to affect the developed model as lagged LE values are not considered in the developed models. The model memory is taken into consideration by considering lagged values of other input variables. The SBH hourly data, which cover the period 20 May–25 August 2003, are divided into a calibration (or training) group which includes 658 instances and a validation (or testing) group which includes 381 instances. For SWSS, data were available from 15 May to 10 September 2005. Similar to the SBH data, the SWSS data are divided into a training group of 787 instances and testing group of 408 instances.

The authors of this study were careful to use exactly the same fixed datasets used by [Parasuraman *et al.* \(2007a\)](#) to provide a proper basis for comparing different data-driven modelling tools. The choice of similar datasets was based on preserving the information content in both datasets for the training and testing steps, in order to provide unbiased judgment on the prediction capability of EPR compared to ANN and GP. This study explores the effect of using different polynomial forms, functions and solution methods provided by the EPR tool to create a comprehensive model that can be used to reasonably characterize the evapotranspiration component of the hydrological cycle.

RESULTS AND DISCUSSION

Static AET modelling

SBH case study

The GP evolved model has the ([Parasuraman *et al.* 2007a](#)):

$$LE = -0.20NR^3 + 0.35NR^2 + 0.31NR - 0.20(NR^2 \cdot GT) + 0.26(NR \cdot GT) + 0.42GT - 0.18 \quad (3)$$

Equation (3) shows the dominance of NR and GT over other input variables. As mentioned, both GP and EPR use randomized evolutionary search methods due to the difficulty in performing an exhaustive search on the infinite

search space. It is therefore possible that such techniques may not find the optimum model. In this case, the EPR produced a very simple model that also includes both NR and GT for the same sets of training and testing data:

$$LE = 0.02 (NR \cdot GT) + 35.52 \quad (4)$$

The EPR produces several models (as a result of a multi-objective optimization approach which trades the fitness to training data against the parsimony of obtained expressions i.e. the number of polynomial terms) where the user of the EPR tool sets the maximum number of terms that can be used in the produced model. Other EPR models which include a higher number of terms are therefore also examined and the next best equation, among EPR models, is found to be:

$$LE = 0.22 NR + 4.62 \times 10^{-5} (NR \cdot AT \cdot GT^2) - 4.54 \\ \times 10^{-7} \left(\frac{NR \cdot AT^2 \cdot GT^2}{RH} \right) - 6.10 \\ \times 10^{-6} (NR^2 \cdot GT \cdot RH^2 \cdot WS) + 0.04 \left(\frac{GT^2}{RH} \right) + 3.95 \quad (5)$$

Table 1 lists a summary of results of the three data driven modelling tools based on the selected error measures. During the testing phase, there was a noticeable improvement in the results (see RMSE and MARE values) using EPR models (Equations (4) and (5)) relative to the ANN and GP models. Description of the ANN and GP model formulations can be found in Parasuraman *et al.* (2007a). This paper focuses on the application of EPR and the performance of EPR models in comparison to ANN

and GP models. The first EPR model, Equation (4), achieved an MARE value of 0.85 in the testing phase, which is lower than the other models. RMSE value shows similar improvement. The second EPR model, Equation (5), was poorer than the first model, particularly with regard to the RMSE values.

Evaluating all models from training to testing phases also reveals the better performance of the EPR models over other models. For example, in the case of ANNs, the value of RMSE increased by 17–39% of its training value while for EPR the increase is from 2–27%. The deterioration in the RMSE value using the GP model was 13% of its training value. More severe deterioration is evident in MARE values from training to testing phases, especially for ANN models. The deterioration in the MARE value in case of ANN models ranged from 76–104% of its training value. A 70% deterioration is realized with the GP model, while the deterioration range in EPR models was from 37–62%. Generally, it can be concluded that the EPR models provided better and more consistent characterization, in terms of performance of the training and testing phases, of the evapotranspiration process based on the metrological inputs. Less deterioration in the model performance during the testing phase, relative to the training phase, indicates better generalization ability.

A statistical test of significance of the testing datasets provides a reliable procedure to confirm or refute conclusions concerning the relative performance of different models. The main goal of this test is to determine if the difference between datasets is merely due to chance or not. The statistical test of hypotheses involving sample differences is used to test the null hypotheses that there is no

Table 1 | SBH case study error measures of the training and testing steps

| Model | Training | | | Testing | | |
|---------------------------------|----------|------|------|---------|------|------|
| | RMSE | MARE | R | RMSE | MARE | R |
| ANN (NR, GT)* | 56.7 | 0.53 | 0.82 | 66.1 | 0.93 | 0.77 |
| ANN (all inputs)* | 50.4 | 0.50 | 0.86 | 69.8 | 1.02 | 0.72 |
| GP (Equation (3))* | 57.8 | 0.54 | 0.82 | 65.5 | 0.92 | 0.77 |
| EPR (NR, GT) (Equation (4)) | 59.4 | 0.62 | 0.81 | 60.4 | 0.85 | 0.75 |
| EPR (all inputs) (Equation (5)) | 52.8 | 0.52 | 0.85 | 67.2 | 0.84 | 0.71 |

*Results based on Parasuraman *et al.* (2007a).

Table 2 | Statistical test of significance results of the SBH case study

| Model | Data mean (\bar{x}) | Data st. deviation (σ) | Sampling distribution st. deviation ($\sigma_{\bar{x}_1 - \bar{x}_2}$) | Sampling distribution statistic (Z) | Critical | |
|-------------------------------|-------------------------|---------------------------------|--|---|---------------------|---------------------|
| | | | | | $Z_{\alpha} = 0.05$ | $Z_{\alpha} = 0.01$ |
| GP (Equation (3)) | 152.6 | 75.6 | 5.50 | 2.18 | | |
| EPR (NR, GT)* | 140.6 | 76.1 | | | | |
| GP (Equation (3)) | 152.6 | 75.6 | 5.58 | 1.35 | | |
| EPR (all inputs) [†] | 145.1 | 78.3 | | | | |
| ANN (NR, GT) | 153.8 | 75.4 | 5.49 | 2.40 | 1.96 | 2.58 |
| EPR (NR, GT) | 140.6 | 76.1 | | | | |
| ANN (all inputs) | 151.2 | 77.6 | 5.65 | 1.09 | | |
| EPR (all inputs) | 145.1 | 78.3 | | | | |
| EPR (NR, GT) | 140.6 | 76.1 | 4.87 | 5.41 | | |
| EPR (Dynamic) [‡] | 114.3 | 56.9 | | | | |

*Equation (4).

†Equation (5).

‡Equation (10).

difference between any two testing datasets. Theoretical background and test procedures are available in many probability theory publications, while concise steps can be found in Schaum's outlines series (Spiegel *et al.* 2009).

The testing datasets of GP and EPR models using NR and GT, Equations (3) and (4), are selected as both models show comparable performance measures and relatively simple structure. The test starts by calculating the basic statistics (means and standard deviations) of the testing datasets. These values are then used to calculate the sampling distribution of the difference of means standardized statistic (Z). This value is compared to the normal distribution standardized statistics at different levels of significance (α), critical Z value (Z_{α}). If the Z value exceeds the value of Z_{α} , then the difference in means between the two datasets is significant and not merely due to chance at the corresponding level of significance. For example in Table 2, comparing the Z value for the datasets for Equations (3) and (4) (Z value is 2.18) indicates that there is significant difference between GP and EPR testing datasets at 0.05 level of significance, but not at 0.01 level. This result indicates that there is a probable significant difference between the GP and EPR estimates of the AET.

Carrying out the test of significance for the GP and EPR results using all input variables (Table 2) shows that the

difference is not significant at any level of significance. Carrying out the test of significance between corresponding EPR and ANN testing datasets (Table 2) reveals that there is a probable significant difference using NR and GT, while there is no significant difference in the case of using all input variables. Generally, there is a probable significant difference in EPR testing dataset between NR + GT and GP + ANN datasets, although this difference is not significant in the case of using all input variables. However, EPR provides a relatively simple and robust modelling technique and better control over the model structure, as opposed to ANNs.

It should be noted that although Equation (5) contains more input variables than Equation (4), it did not improve the overall performance of the model. The contribution of the NR and GT ranges from 47 to 90% of the LE value, which can be attributed to the dominance of NR and GT over other variables for this particular ecosystem. Conversely, the inclusion of other variables resulted in higher model noise (over-fitting) that negatively affected the prediction ability of the model (note the deterioration of testing results compared to the training). This result supports the conclusion of Parasuraman *et al.* (2007a) that NR and GT can explain most of the variance in the LE flux and AET rates.

SWSS case study

A key distinction between the SBH and the SWSS sites is that the SWSS has a complete canopy cover of the surface with no exposed bare ground. In addition, the vegetation canopy is taller and more developed with a wide range of species. Equation (6) represents the GP model (after Parasuraman *et al.* 2007a):

$$LE = 0.48 NR + 0.29 GT^5 + 0.22 (GT \cdot AT \cdot WS) \quad (6)$$

The optimum models evolved for this case study, using the EPR tool, are represented by Equations (7) and (8):

$$LE = 0.02 NR \cdot GT + 0.01 (AT \cdot GT^2) + 39.19 \quad (7)$$

$$\begin{aligned} LE = & 2.24 (NR \cdot GT) - 0.68 GT + 1.41 (GT \cdot AT) \\ & - 0.36 (NR^{0.5} \cdot GT \cdot RH^2) - 0.29 (NR^2 \cdot AT^{0.5}) \\ & - 0.66 AT^2 - 1.59 (NR \cdot GT^2) \\ & + 0.73 (GT^2 \cdot AT \cdot RH \cdot WS) + 0.19 \quad (8) \end{aligned}$$

Table 3 shows the results of the three tools and the corresponding error measures. Unlike the first case study, the performance of the EPR is similar to the other ANN and GP models, especially in the testing phase. Using the testing phase results as criteria for evaluating the developed models, it is clear that the performances of all models are similar. Both GP and EPR models (using all inputs) have similar error measure values. The values of the error measures during both training and testing are quite comparable, indicating a stable ecosystem that was easier to model (in terms of model terms, data requirement

and computational data) by the data-driven techniques (Table 3). A stable ecosystem is one which has ability to return to equilibrium (steadiness) state after perturbations and is not in the developing stage (immature) or still developing (McCann 2000).

Unlike the SBH case study, EPR produced a more complex formula (compare Equation (8) which uses all variables to Equation (6)) than the GP. This indicates that random search and evolutionary-based techniques, such as GP and EPR techniques, do not guarantee consistent performance in all case studies e.g. good and/or bad performance for modelling AET. This is due to the practical impossibility of conducting exhaustive search, i.e. searching the entire solution space, to reach the optimal model. This limitation could be attributed to the adopted technique, the toolbox/software or to both of these, highlighting the importance of using multiple techniques and multiple tools to develop a robust data-driven model for hydraulic performance.

Dynamic AET modelling

SBH case study

The three data-driven techniques demonstrated comparable performance in characterizing actual evapotranspiration. EPR was therefore chosen to explore the effect of time-lagged inputs on estimating the actual evapotranspiration due to the simplicity of its toolbox and the ability to acquire the mathematical form of the evolved model. The tool was allowed to use lag time up to 7 hours back for all input variables. The choice of the 7 hours lag was based on preliminary trials, which included higher lag times. All EPR

Table 3 | SWSS case study error measures of the training and testing steps

| Model | Training | | | Testing | | |
|---------------------------------|----------|------|------|---------|------|------|
| | RMSE | MARE | R | RMSE | MARE | R |
| ANN (NR, GT)* | 41.7 | 0.29 | 0.85 | 38.8 | 0.37 | 0.85 |
| ANN (all inputs)* | 39.0 | 0.26 | 0.87 | 39.3 | 0.34 | 0.84 |
| GP (Equation (6))* | 42.2 | 0.27 | 0.84 | 39.0 | 0.32 | 0.86 |
| EPR (NR, GT) (Equation (7)) | 42.5 | 0.31 | 0.84 | 41.3 | 0.44 | 0.83 |
| EPR (all inputs) (Equation (8)) | 39.5 | 0.28 | 0.86 | 38.1 | 0.33 | 0.86 |

*Results based on Parasuraman *et al.* (2007a).

models excluded lagged input variables higher than 7 hours. Different forms of the resultant polynomial were explored including the use of logarithmic and exponential functions, as performed in the static AET modelling. Several models were produced by the EPR. The following equation produced the optimum performance based on the three error measures for the training and testing phases:

$$\begin{aligned} LE_t = & 0.20 NR_{t-1} + 0.02 (NR_t \cdot GT_{t-1}) - 1.38 \\ & \times 10^{-11} (NR_t^2 \cdot AT_{t-1}^2) \\ & \times \left(\frac{GT_t}{GT_{t-1}} \right)^2 \left(\frac{1}{RH_{t-2}^2 \cdot RH_{t-3} \cdot RH_{t-7}} \right) \\ & \times \left(\frac{WS_{t-1}^2 \cdot WS_{t-5}^2 \cdot WS_{t-6}^2}{WS_{t-2} \cdot WS_{t-7}^2} \right) + 11.15 \end{aligned} \quad (9)$$

Equation (9) is complex due to the existence of several input variables and their respective lag. By removing the third term from Equation (9) (due to its small coefficient), a simpler formula is produced without affecting its prediction ability (Table 4):

$$LE_t = 0.20 NR_{t-1} + 0.02 (NR_t \cdot GT_{t-1}) + 11.15 \quad (10)$$

Model simplification might imply that EPR is unable to achieve a good compromise between model simplicity and prediction ability. However, it should be noted that EPR managed to produce a relatively simple model with regard to the available input combinations. Simplification also partially affected the model performance, which applies to the training phase but not the testing phase (Table 4). EPR considers a training dataset in model development. A compromise, in both training and testing, should therefore be based on the subjective view of the modeller.

Table 4 | SBH case study best equations based on different error measures

| Equation | Training | | | Testing | | |
|----------|----------|------|------|---------|------|------|
| | RMSE | MARE | R | RMSE | MARE | R |
| (9) | 55.1 | 0.56 | 0.84 | 55.4 | 0.71 | 0.78 |
| (10) | 57.6 | 0.58 | 0.82 | 56.8 | 0.71 | 0.78 |
| (11) | 55.0 | 0.57 | 0.84 | 62.9 | 0.85 | 0.75 |
| (12) | 70.9 | 0.58 | 0.75 | 50.5 | 0.69 | 0.74 |
| (4) | 59.4 | 0.62 | 0.81 | 60.4 | 0.85 | 0.75 |

The second-best equation produced by the EPR has the form:

$$\begin{aligned} LE_t = & 3.21 \times 10^{-6} (NR_{t-1} \cdot NR_{t-2}) \left(\frac{AT_t^3}{AT_{t-1} \cdot AT_{t-2}^2} \right) \\ & \times \left(\frac{GT_{t-2}^3}{GT_{t-1}^2} \right) RH_t \left(\frac{1}{WS_{t-2}} \right) + 0.01 NR_t \left(\frac{GT_t^3}{GT_{t-1}^2} \right) + 42.35 \end{aligned} \quad (11)$$

Similarly, removing the first term with the relatively small coefficient and other input variables other than NR and GT yields:

$$LE_t = 0.01 NR_t \left(\frac{GT_t^3}{GT_{t-1}^2} \right) + 42.35 \quad (12)$$

Here, the simplification of Equation (11) adversely affected its performance in the training phase while it improved its performance in the testing phase. From the produced equations, NR_t , GT_t , NR_{t-1} and GT_{t-1} have the dominant effect relative to other inputs, as they are always included in the polynomial terms with the relatively higher coefficients. Other inputs such as AT, WS and RH have less contribution to the value of the current LE_t . Higher lags of GT and NR do not have a similar effect to the 1-hour lag, as they are found in polynomial terms with small coefficients.

Table 4 shows the error measures for Equations (9)–(12) in both training and testing phases, relative to the best EPR model from Table 1 (Equation (4)). This indicates the importance of including previous input states, especially NR_{t-1} and GT_{t-1} , in addition to their current values to improve the performance of the models in characterizing the AET process. It can be concluded that both Equations (9) and (10) are good representatives of the SBH case study.

The results of the statistical test of significance (Table 2) for EPR static and dynamic models (Equations (4) and (10)) indicate that there is a significant difference between the two models (at all levels of significance for EPR dynamic model) and a probable significant difference for EPR static model (significant at 0.05 level, but not at 0.01 level). This difference is shown in Figure 1, which presents the scatter plots of the observed and computed LE values for the EPR models. Figure 1 indicates that the dynamic EPR model has

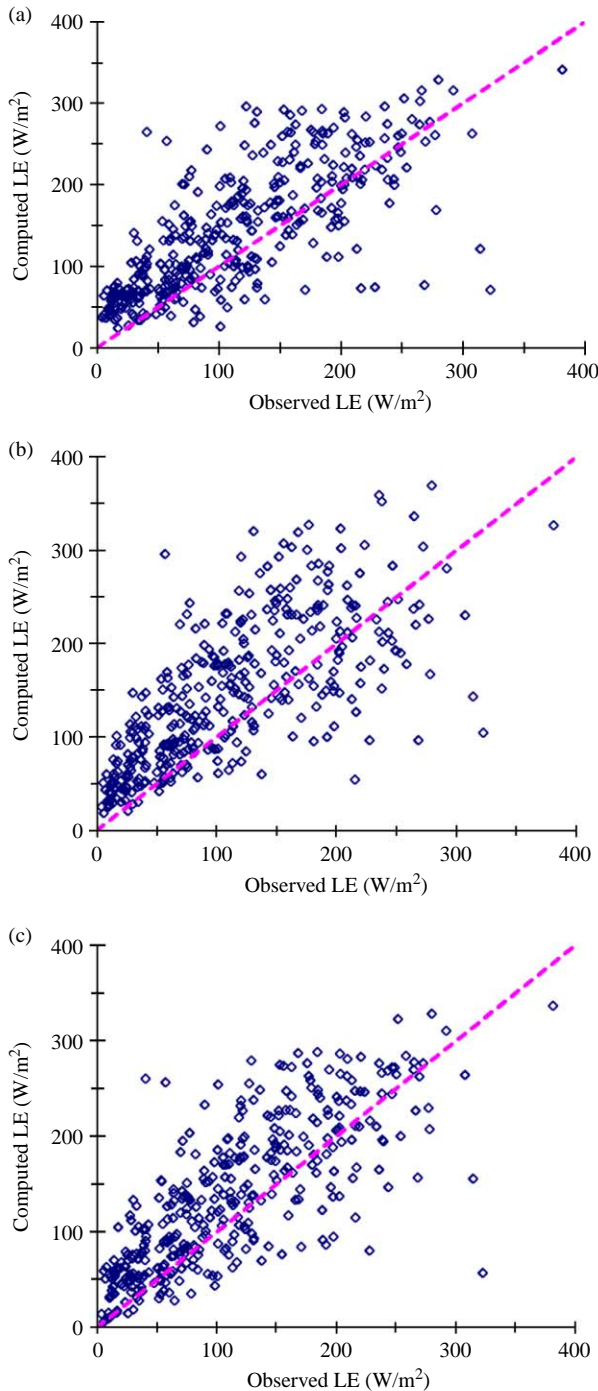


Figure 1 | The scatter plot of the SBH case study using the EPR models: (a) with two inputs (Equation (4)); (b) all inputs (Equation (5)); and (c) dynamic model (Equation (10)).

Table 5 | SWSS case study best equation based on different error measures

| Equation | Training | | | Testing | | |
|----------|----------|------|------|---------|------|------|
| | RMSE | MARE | R | RMSE | MARE | R |
| (13) | 35.5 | 0.24 | 0.89 | 34.4 | 0.32 | 0.88 |
| (14) | 37.4 | 0.32 | 0.84 | 38.4 | 0.45 | 0.86 |
| (8) | 39.5 | 0.28 | 0.86 | 38.1 | 0.33 | 0.86 |

a better performance than the other static models due to the inclusion of previous states of the NR and GT variables.

SWSS case study

Some of the equations evolved by EPR showed marginally improved performance over the static GP and ANNs models (see RMSE and R values in Table 3), especially in the training phase (Table 5). The simplest case relative to the other EPR models is:

$$\begin{aligned}
 LE_t = & 0.26 NR_t + 0.30 GT_t^2 + 4.09 \\
 & \times 10^{-7} (NR_{t-1}^2 \cdot AT_{t-2}^2 \cdot RH_{t-3}^4 \cdot WS_{t-1}) + 1.89 \\
 & \times 10^{-12} (NR_{t-1} \cdot AT_{t-2}^3 \cdot GT_t^3 \cdot GT_{t-1} \cdot RH_{t-3}^2 \cdot WS_{t-2}^2 \cdot WS_{t-3}^2) \\
 & - 0.001 (GT_t^4 \cdot RH_t^4)
 \end{aligned} \quad (13)$$

Simplification by removing the last two terms with small coefficients and relatively small value based on the range of the input variables used in each term, results in:

$$\begin{aligned}
 LE_t = & 0.26 NR_t + 0.30 GT_t^2 + 4.09 \\
 & \times 10^{-7} (NR_{t-1}^2 \cdot AT_{t-2}^2 \cdot RH_{t-3}^4 \cdot WS_{t-1})
 \end{aligned} \quad (14)$$

Simplification of Equation (13) adversely affected its performance compared to its complicated form and the static EPR model. The MARE measure is the most sensitive error measure to such simplifications (Table 5). The SWSS case study represents a well-vegetated cover where the effect of vegetation gives rise to the effect of other inputs such as AT, RS and WS in addition to their previous states. The inclusion of previous states of input variables (in the SWSS case study) did not have a similar improving effect as for the SBH case study.

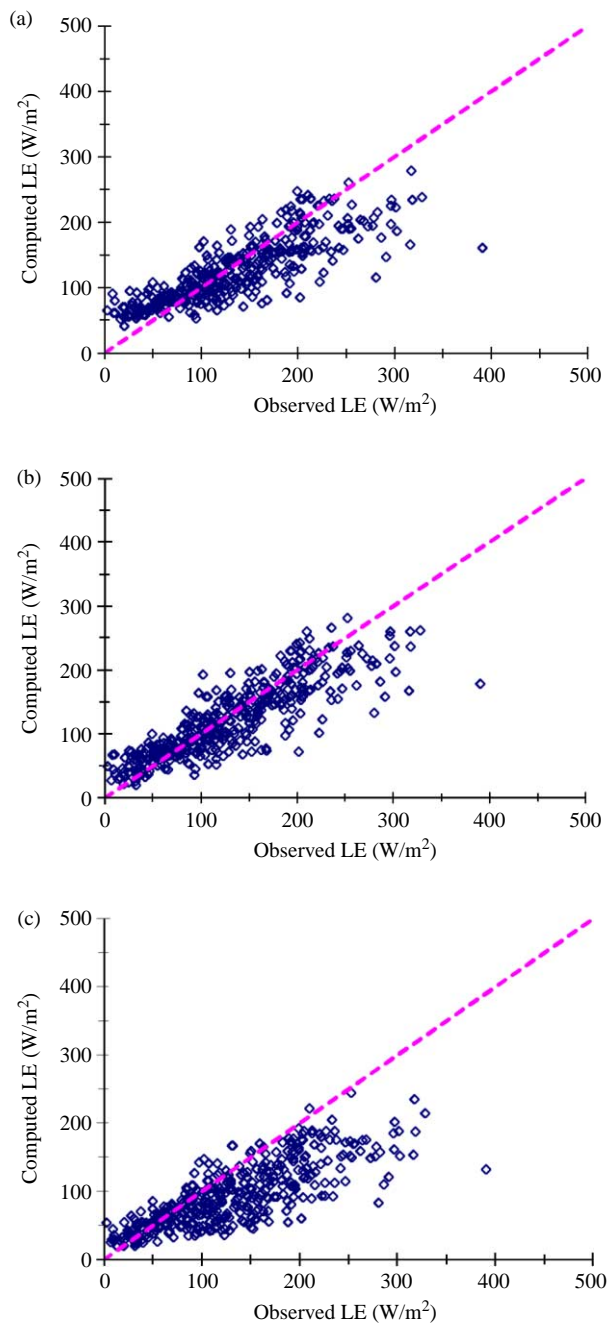


Figure 2 | The scatter plot of the SWSS case study using the EPR models: (a) with two inputs (Equation (7)); (b) all inputs (Equation (8)); and (c) dynamic model (Equation (13)).

A statistical test of significance of the predicted (testing) datasets was carried out on this case study and produced relatively similar results to the results of the SBH case study. It was therefore not necessary to repeat the conclusions of

the SBH case study, which indicated a probable significant difference between the EPR results and the other techniques (GP and ANN). This model also supports the conclusion that the contribution of the NR and GT is significant relative to other inputs, as seen in Equations (4), (10) and (12) where the effects of NR, GT and their time lags are present relative to other input variable. These inputs account for more than 96% of the total evapotranspiration value. The current values of NR and GT have the major role in characterizing AET, while previous time lags of NR and GT, such as NR_{t-1} and GT_{t-1} , do not have the same effect. The dynamic EPR model, Equation (13), has a poorer performance than the other static models due to the simplification of its form (Figure 2).

CONCLUSION

The evapotranspiration process is a significant component of the hydrological cycle. It also affects the other components of the hydrological cycle. The lack of reliable, inexpensive and practical methods to measure the actual evapotranspiration requires the development of alternative methods that provide close estimates of the actual evapotranspiration using readily available meteorological parameters.

The measurements obtained from the eddy covariance method were used in modelling the eddy covariance-measured latent heat flux (LE) to estimate the evapotranspiration losses. The models evolved using the EPR tool, based on the two case studies of Mildred Lake mine, i.e. the SBH and the SWSS case studies, showed comparable performance to the models of the GP and the ANN models. Moreover, the EPR provided simpler models, especially in the SBH case study, than those developed by the other data-driven techniques.

The EPR models supported the conclusion of Parasuraman *et al.* (2007a) that GT and NR can explain most of the LE variation. This is also confirmed when considering previous states of the input variables, as the contribution of the NR and GT ranges from 47 to 90% of the LE value in the SBH case study and 96% in the SWSS case study.

The three data-driven techniques demonstrated a comparable performance in characterizing actual evapotranspiration. The EPR was therefore chosen to explore the effect of time-lagged inputs on estimating the actual evapotranspiration due to the simplicity of its toolbox and the ability to acquire the mathematical form of the evolved model. The dynamic EPR models, for both case studies, resulted in predictions of AET that are significantly different than the static EPR models. The effects of the GT and NR, which is pronounced in the evolved models, is due to the effect of soil moisture and net energy flow conditions on the evapotranspiration process. Other meteorological inputs such as RH, WS, and AT do not have the same effect on the evapotranspiration process at these sites, and may create a less reliable estimate of its value when included in data-driven models.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by the Natural Science and Engineering Research Council (NSERC) of Canada for its financial support through the Discovery Grant and the CRD programs, and also the grant provided by the Cumulative Environmental Management Association (CEMA).

REFERENCES

- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000a *Artificial neural networks in hydrology; I: preliminary concepts*. *J. Hydrol. Eng.* **5** (2), 115–123.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000b *II: Hydrologic applications*. *J. Hydrol. Eng.* **5** (2), 124–137.
- Babovic, V. & Abbot, M. B. 1997 *The evolution of equations from hydraulic data part II: applications*. *J. Hydraul. Res.* **35** (3), 411–430.
- Babovic, V. & Keijzer, M. 2000 Genetic programming as model induction engine. *J. Hydroinformatics* **2**, 35–60.
- Babovic, V. & Keijzer, M. 2002 *Rainfall-runoff modelling based on genetic programming*. *Nord. Hydrol.* **33** (5), 331–346.
- Baldocchi, D. D., Hicks, B. B. & Meyers, T. P. 1988 *Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods*. *Ecology* **69** (5), 1331–1340.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, Ch., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K. & Wofsy, S. 2001 *FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities*. *Bull. Am. Meteorol. Soc.* **82**, 2415–2434.
- Banzhaf, W., Nordin, P., Keller, R. E. & Francone, F. D. 1998 *Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and its Applications*. Morgan Kaufmann Publishers, Inc., San Francisco, CA.
- Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 *Input determination for neural network models in water resources applications. Part 1—background and methodology*. *J. Hydrol.* **301**, 75–92.
- Brutsaert, W. 1982 *Evaporation into the Atmosphere: Theory, History, and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Brutsaert, W. 2005 *Hydrology: An Introduction*. Cambridge University Press, New York, NY.
- Carey, S. K. 2006 *Energy and water exchange from a saline-sodic overburden reclamation soil cover: Fort McMurray, Alberta*. *CLRA Conference*, Ottawa, Ontario, August 20–23.
- Cramer, N. L. 1985 *A representation for the adaptive generation of simple sequential programs*. *Proceedings of an International Conference on Genetic Algorithms and their Applications*, Pittsburgh, PA, July 24–25, pp. 183–187.
- Davidson, J. W., Savic, D. A. & Walters, G. A. 1999 *Method for identification of explicit polynomial formulae for the friction in turbulent pipe flow*. *J. Hydroinformatics* **1** (2), 115–126.
- Dawson, C. W., Abraham, R. J. & See, L. M. 2007 *Hydrotest: a web-based toolbox of evaluation metrics for the standardized assessment of hydrological forecasts*. *Environ. Model. Softw.* **22**, 1034–1052.
- Demuth, H. & Beale, M. 2001 *Neural Network Toolbox Learning. For Use with MATLAB*. The Math Works Inc., Natick, MA.
- Dingman, S. L. 2002 *Physical Hydrology*. Prentice Hall, Inc., Upper Saddle River, NJ.
- Dogliani, A., Giustolisi, O., Savic, D. A. & Webb, B. W. 2007 *An evolutionary approach to stream temperature analysis*. *Hydrol. Process.* **22** (3), 315–326.
- Eichinger, W. E., Nichols, J., Prueger, J. H., Hipps, L. E., Neale, C. M. U., Cooper, D. I. & Bawazir, A. S. 2003 *Lake evaporation estimation in arid environments*. Final Report, IIHR—Hydroscience & Engineering. The University of Iowa, Iowa City, IA, 52242–51585.
- Finnegan, J. J., Clement, R., Malhi, Y., Leuning, R. & Cleugh, H. A. 2003 *A re-evaluation of long-term measurement techniques. Part I: averaging and coordinate rotation*. *Boundary-Layer Meteorol.* **107**, 1–48.
- Fisher, J. B., DeBiase, T. A., Qi, Y., Xu, M. & Goldstein, A. H. 2005 *Evapotranspiration models compared on a Sierra Nevada forest ecosystem*. *Environ. Model. Softw.* **20** (6), 783–796.

- Gasca-Tucker, D. L., Acreman, M. C., Agnew, C. T. & Thompson, J. R. 2007 Estimating evaporation from a wet grassland. *Hydrol. Earth Syst. Sci. J.* **11** (1), 270–282.
- Giustolisi, O. & Savic, D. A. 2006 A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinformatics* **8** (3), 207–222.
- Giustolisi, O., Kapelan, Z. & Savic, D. A. 2006 Multi-objective evolutionary polynomial regression. In *7th International Conference on Hydroinformatics* (eds Gourbesville, P., Cunge, J., Guinot, V. & Liong, S. Y), pp. 725–732. HIC 2006, Research Publishing 1, Nice, France.
- Giustolisi, O., Doglioni, A., Savic, D. A. & Webb, B. W. 2007 A multi-model approach to analysis of environmental phenomena. *Environ. Model. Softw.* **22** (5), 674–682.
- Gutierrez, M. V. & Meinzer, F. C. 1994 Energy balance and latent heat flux partitioning in Coffee hedgerows at different stages of canopy development. *Agric. Forest Meteorol.* **68** (3–4), 173–186.
- Hornberger, G. M., Raffensperger, J. P., Wiberg, P. L. & Eshleman, K. N. 1998 *Elements of Physical Hydrology*. The Johns Hopkins University Press, Baltimore, MD.
- Jayawardena, A. W., Muttill, N. & Fernando, T. M. K. G. 2005 Rainfall-Runoff modelling using genetic programming. In: *International Congress on Modelling and Simulation*. (eds Zerger, A. & Argent, R. M), pp. 1841–1847. MODSIM 2005, Modelling and Simulation Society of Australia and New Zealand, December 2005.
- Koza, J. R. 1992 *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press, Cambridge, MA.
- Laucelli, D., Berardi, L. & Doglioni, A. 2005 Evolutionary polynomial regression toolbox: version 1.SA. Department of Civil and Environmental Engineering, Technical University of Bari, Bari, Italy. Available from: <http://www.hydroinformatics.it/prod02.htm> (accessed November 2007).
- Margolis, H. A., Flanagan, L. D. & Amiro, B. D. 2006 The Fluxnet-Canada research network: influence of climate and disturbance on carbon cycling in forests and peatlands. *Agric. Forest Meteorol.* **140**, 1–5.
- McCann, K. S. 2000 The diversity–stability debate. *Nature* **405**, 228–233.
- MingFeng, S. & HuiJun, W. 2007 Decadal co-variability of the summer surface air temperature and soil moisture in China under global warming. *Chin. Sci. Bull.* **52**, 1559–1565.
- Obukhov, A. M. 1946 Turbulentnost' v temperaturnoj–neodnorodnoj atmosfere (turbulence in an atmosphere with a non-uniform temperature). *Trudy Inst. Theor. Geofiz.* **1**, 95–115.
- Parasuraman, K., Elshorbagy, A. & Carey, S. K. 2006 Spiking–modular neural networks: a neural network modeling approach for hydrological processes. *Water Resour. Res.* **42**, W05412.
- Parasuraman, K., Elshorbagy, A. & Carey, S. K. 2007a Modelling dynamics of the evapotranspiration process using genetic programming. *Hydrol. Sci. J.* **53** (3), 563–578.
- Parasuraman, K., Elshorbagy, A. & Si, B. C. 2007b Estimating saturated hydraulic conductivity using genetic programming. *Soil Sci. Soc. Am. J.* **71**, 1676–1684.
- Penman, H. L. 1948 Natural evaporation from open water, bare soil, and grass. *Proc. R. Soc. Lond. A* **193**, 120–146.
- Savenije, H. H. G. 2004 The importance of interception and why we should delete the term evapotranspiration from our vocabulary. *Hydrol. Process.* **18**, 1507–1511. (Invited Commentary).
- Savic, D. A., Giustolisi, O., Berardi, L., Shepherd, W., Djordjevic, S. & Saul, A. 2006 Sewers failure analysis using evolutionary computing. *Water Manage. J.* **159** (2), 111–118.
- Solomatine, D. P. 2002 Applications of data-driven modelling and machine learning in control of water resources. In *Computational Intelligence in Control* (ed. M. Mohammadian, R. A. Sarker & X. Yao), pp. 197–217. Idea Group Publishing, Hershey, PA.
- Spiegel, M. R., Schiller, J. & Alu Srinivasan, R. 2009 *Schaum's Outline Series: Probability and Statistics*, Third edition. McGraw Hill, New York, NY.
- Strangeways, I. 2003 *Measuring the Natural Environment*. Cambridge University Press, Cambridge, UK.
- Stull, R. B. 1988 *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, Dordrecht.
- Wilson, K. B. & Baldocchi, D. D. 2000 Seasonal and inter-annual variability of energy fluxes over a broadleaved temperate deciduous forests in North America. *Agric. Forest Meteorol. J.* **100**, 1–18.
- Xu, C. Y. & Singh, V. P. 1998 A review on monthly water balance models for water resources investigations. *Water Resour. Manage.* **12**, 31–50.
- Xu, C.-Y. & Singh, V. P. 2005 Evaluation of three complementary relationship evapotranspiration models by water balance approach to estimate actual regional evapotranspiration in different climatic regions. *J. Hydrol.* **308** (1–4), 105–121.

First received 5 May 2009; accepted in revised form 30 July 2009. Available online 26 March 2010