A COMPARATIVE STUDY on PREDICTION of EVAPORATION in ARID AREA BASED on ARTIFICIAL INTELLIGENCE TECHNIQUES

MANSURA JASMINE

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the

MASTER OF APPLIED SCIENCE

in Civil Engineering

Department of Civil Engineering

Faculty of Engineering

University of Ottawa

Abstract

Estimation of evaporation from open water is essential for hydrodynamics, manufacturing industries, irrigation, farming, environmental protection and many other purposes. It is also important for proper management of hydrological resources such as reservoirs, lakes and rivers. Recent methods are mostly data-driven methods, such as using Artificial Intelligence techniques. Adaptive Neuro Fuzzy Inference System (ANFIS) is one of them and has been widely adopted in many hydrological fields for its simplicity. The current research presents a comparative study on the impact of optimization techniques such as Firefly Algorithm (FFA), Genetic Algorithm (GA), Particle Swarm Optimizer (PSO) and Ant Colony Optimization (ACO) on obtained results. In addition, a practical method named Multi Gene-genetic Programming (MGGP) is employed to propose an equation for the estimation of the Evaporation. Six different measured weather variables are taken, which are maximum, minimum and average air temperature, sunshine hours, wind speed and relative humidity. Models are separately calibrated with total data set collected over an eight-year period of 2010-2017 at the specified station "Arizona" in the United States of America. Ten statistical indices are calculated to verify the results. All optimizers were observed and compared to check if the results are better than ANFIS or not. The objectives of the adoption of different optimizer techniques was to verify the accuracy of the prediction by ANFIS model. Comparisons showed that ANFIS and MGGP are slightly better than the other models. MGGP model is different from other models in a way that it provides a set of equations instead of showing numerical values; therefore, the computational time is high. PSO, FFA, ACO and GA are considered as optimizers in the main model. Though PSO provided very similar results to the ANFIS model and MGGP gives even better results than basic ANFIS model. ANFIS is easier in terms of model formation. ANFIS is simpler to build and easy to operate. Since the prediction was quite identical in all cases, the ANFIS model was suggested due to its simplicity.

Dedication: I would like to dedicate this thesis to my parents, my family, my beloved husband Mehedi Hasan and my little princess Muntaha Hasan.

Acknowledgement

My heart-felt thanks and greetings to my supervisor, Professor Dr. Majid Mohammadian, without his continuous support and guidance this thesis would not have been possible. He is very generous and diligence. His encouragement and quick response to any problems I have faced; have been a great motivation for me to continue this valuable work. I must have to mention one thing that, whenever I sent him an email at late night, I got reply right away; furthermore, he replied my quires continuously in the weekends by sacrificing his leisure time. All of his enthusiasm inspired me to work hard for the thesis to make him happy. Thanks to my co-supervisor, Dr. Bonakdari for his kind assistance over the past two years, especially he helped me writing my manuscripts properly. I also would like to thank Dr. Ioan Nistor who helped me getting admission in university of Ottawa. My deepest gratitude to my husband for inspiring me for higher studies and helping with financial support along with other necessary support all over this journey. I am very much thankful to Dr. Trevor Hall for his moral support and contribution on the initial stage of my higher studies. Thanks to Canadian government for issuing my visa. Finally, I give a special thanks to my parents who sacrificed a lot for me throughout my life. I am indebted to all my relatives and friends for their continuous support, closely or remotely, during my study.

Table of Contents

Chapter 1 Introduction

Chapter 2 Literature Review

Chapter 3 ANFIS-Type Models for Prediction of Evaporation in Arid Climate.

Chapter 4 An integrated approach based on Multigene Genetics Programming for Prediction of Evaporation in Arid Climate.

Chapter 5 Conclusion

List of Figures

List of Tables

List of Abbreviations

ANFIS	Adaptive Neuro Fuzzy Inference System
FIS	Fuzzy Inference System
ANN	Artificial Neural Network
PSO	Particle Swarm Optimization
ACO	Ant Colony Optimization
FFA	Firefly Algorithm
GA	Genetic Algorithm
MGGP	Multiple Gene-Genetic Programming.
AET	Actual Evapotranspiration
WHO	World Health Organization
ET	Evapotranspiration
ENN	Evolutionary Neural Network
MLP	Multi-Layer Perceptron Neural Network
GRNN	Generalized Regression Neural Network
DE	Differential Evolution
RMSE	Root Means Square Error
MAE	Mean Absolute Error
SI	Scatter Index
MARE	Mean Absolute Relative Error
VAF	Variance Account For
RMSRE	Root Mean Square Relative Error
GEP	Gene Expression Programming
SC	Subtracting Clustering
MLP	Multi-layer Perception
FG	Fuzzy Genetic
LSSVM	Least square Support Vector Machine
GRNN	Generalized Regression Neural Network
MARS	Multivariate Adaptive Regression Spline
FCTP	Fixed-charge transportation problem

CSA	Cuckoo Search Algorithm
DL	Deep Learning
GLM	Generalized Linear Model
RF	Random Forest
GBM	Gradient-Boosting Machine
GI	Glycemic Index
FPA	Flower Pollination Algorithm
MRP	Manmade River Project
MF	Membership Function
CV	Coefficient of Variation
SV	Standard Deviation

List of Symbols

α	Movement Co-efficient
β	Attraction Co-efficient
γ	Light Absorption Co-efficient
ε_i	Random Number Vector
I	Light Intensity
I_0	Light Intensity at Zero Distance
w	Attractiveness
w_0	Attractiveness at Zero Distance
ω	Inertial Weight (Explore Variable)
$\boldsymbol{c_1}$	Cognition Variable (Explore Variable)
$\boldsymbol{c_2}$	Social Variable (Explore Variable)
k_{max}	k_{max} Highest Number of Repetition
k	k Current Number of Repetition
f_{best}	Finest Objective Function.
f_{worst}	Least Objective Functions
ε_{div}	Diversity Expression
ρ	Rate of Evaporation
τ_{kj}	Pheromone Concentration in side (k,j)
N_k^l	Feasible Neighbourhood of ant
η_{kj}	Visibility of Heuristic Matrix
$\Delta\tau^l$	Accretion in Pheromone Matrix

Chapter 1. Introduction

1.1 Background

Currently, water deficiency is increasing and becoming challenging for human being. For example, Libya has built one of the largest civil engineering groundwater pumping and transferring systems to overcome limitation of water and climate hinderance (high temperature and low rainfall). This project is also called the Manmade River Project (MRP), (Benzagtha, 2014). The purpose of this project was to supply water demand of Libya by pumping underground water underneath the Sahara Desert and transferring it using a network of huge underground pipes especially for irrigation. The high cost of water pumping, and the lack of appropriate planning are the main concerns. Water deficiency is increasingly becoming the most important environmental limitation which is limiting plant growth. It is also important to study this problem in Arid and semi-Arid climates. For example, over 30 arid and semi-arid countries are expected to have water deficiency in 2025 (Benzagtha, 2014). This will limit the development, threaten food supplies and inflame rural poverty. Estimation of water loss by evaporation is essential for integrated water resources management and modeling studies related to hydrology, agronomy, forestry, irrigation, flood and lake ecosystem. Evaporation is outlined as the loss of stored water due to change from liquid state to vapor state, which is influenced by the climate condition such as temperature, wind speed, and solar radiation. Few statistics can be drawn in order to understand the importance of calculating evaporation. According to The World Meteorological Organization (WMO), three quarters of the total input (inflow and over-lake rainfall) to Lake Victoria in the U.S. is lost due to evaporation, which results in relatively humid conditions (Benzagtha, 2014). In Australia, it is estimated that about 95% of the rain evaporates and has no contribution to runoff (Benzagtha, 2014). In Egypt's

Lake Nasser (located in arid area) where the Nile's water is stored, downstream water loss due to evaporation is estimated to be 3 meters in depth, or double that of Lake Victoria (Benzagtha, 2014). Therefore, estimation of evaporation is important for irrigation water management to know the amount of water needed for plants to grow, for rainfall-runoff modeling, and other water resources studies (Adeloye et al. 2012). Evaporation has a huge influence on the allocation of water in hydrological cycle, farming and water resource management. Evaporation refers to the motion of water to the atmosphere from soil, or waterbodies. According to the study of Kuo et al., 2011, accessible water sources are absorbed largely by irrigation and farming in Taiwan. Evapotranspiration (ET) is another term related to irrigation and farming. Evapotranspiration (ET) is the grand total of evaporation and plant transpiration from the soil surface and sea surface to the air. Estimation of Potential Evapotranspiration is also important. It plays an important role in estimating actual evapotranspiration (AET) in rainfall-runoff and ecosystem modeling.

Water deficiency is a global problem in all over the world. Governments of many countries and a few non-governmental organizations are working on water deficiency in order to manage water-saving irrigation system. 70% to 80% of water is being reused in developed countries. For example, due to water deficiency in China, 80% of total water consumption are utilized by the irrigation (Xuanrong. et al. 2017). Therefore, it is necessary to apply technological methods to encourage productive management of water resources, and successfully uplift the proficiency of irrigation water.

Around 61 percent of total rainfall is being evaporated from the ground (Chow et al. 1988 and Kim et al. 2010) and hence it is an essential element of the hydrological sector and its numerical analysis is a compulsory subject in water resources engineering field. Computing evaporation is also a very fundamental element for water demand and supply management and lake or river ecosystem. But it is not always easy to compute because of the uncertainty of the behavior of the atmospheric element.

There are many methods which have been practiced for many years. Usually two types of methods are very popular; direct and indirect methods (Tabari et. al, 2010). Example of a direct method is pan evaporation method. Pan evaporation is used to estimate the evaporation from lakes. Daily pan evaporation is an important term of estimation of water loss through evaporation. It is being used as an indicator of evapotranspiration, irrigation scheduling and for estimation of lake and reservoir evaporation. Evaporation pans have been adopted and compared with other methods (Choudhury 1999, Vallet-Coulombet al. 2001 and Kim et al. 2012). The U.S. Weather Bureau Class A pan is the highest adopted pan with 21 cm of diameter, 25.5 cm of depth, and situated on a timber which is 15 cm high from the ground. Pan coefficient is an important factor for pan evaporation and the size and state of upwind zone also have effect on it. It indicates the ratio of the evaporation volume from water reservoir to that calculated with evaporation pan. The range of it lies between 0.35 to 0.85 and this limit varies from condition to condition (Allen. et al., 1998 and Kim et al. 2012). Direct method is established on different types of field measurements and evaporimeter is also used. But this test is time consuming and needs precision in instrumental set up to get acceptable results. Pan efficiency can be influenced by handling issue and instrumental difficulties (Jensen. et al., 1990; Doorenbos; Pruitt 1977 and Tabari. et al., 2010). For example, it may have instrumentation errors, errors occur during operation and maintenance, watering of animals and weather effects which can cause an inaccurate measurement for the estimation (e.g. Sabziparvar. et al., 2009 and Tabari. et al., 2010). Evaporimeter cannot be placed anywhere, especially where a planned or existing lake or irrigation plan of action is already situated. Scaling is a problem, especially it is not usable for the large-scale experiments. Another limitation of this method is its expensive instrumental tools, which cannot be placed in a distant location. On the other hand, evaporation system is very non-linear (Kisi, 2006; Eslamian. et al., 2008; Tabari. et al., 2010). Therefore, it was becoming challenging to the researchers to compute evaporation where data are not available (Kisi, 2006). On the other hand, indirect methods are highly dependent on weather variables, and few of them demands data which is not always available (Rosenberry. et al., 2007 and Kim. et al., 2012). Usually, equation-based methods are considered as indirect methods, example of an indirect method is Penman-Monteith Benchmark Model. Two types of indirect methods are described by Adeloye. et al. (2011). Those are, Theoretical method and empirical method. Theoretical methods are fit for heat and mass transfer and energy formulation for evaporating surface and the result can be perfect if the data for climate variables (such as, temperature, humidity, wind speed and solar radiation) are available. For the different cases when data are not available, empirical method is required. This method mostly depends on the subset of available data (Adeloye. Et al., 2012). The limitation of this method is that, it is only valid for the source location of a data set and for the similar weather condition. Also, accuracy depends on the measurements; therefore, accuracy is the main issue for indirect methods. Thus, accuracy became the matter of concern to many researchers. Also, evaporation process is highly non-linear in nature. Hence, the importance of model-based computation is felt that can relate the non-linearities. Here artificial intelligence can be an alternative. Artificial Neural Network (ANN) is a perfect fit for this type of non-linear estimation. The potential of resolving problems that are complicated to formulate is the supremacy of data driven method (Sudheer et. Al, 2003).

Evaporation has high dependency on climate variables, such as, temperature, wind speed, humidity and a few more. All the variables are independent. For example, evaporation rate is higher for low humidity. Again, temperature has impact on humidity and surface temperature. For example, evaporation rate is lower for low temperature. Thus, evaporation is affected by all these factors. After many studies, researchers discovered data analysis-based modelling with the accuracy performance calculation. Artificial neural network (ANN) is one of them. ANN is a well-received modeling method for different topology and weather conditions among many modeling methods. Good thing about ANN model is that, it needs less input variables to run. ANN method copies the cognitive response of the human brain. It is a biologically motivated computational model that contains processing elements (neurons) and links between them (weights). ANNs have an advantage over other models with regard to the data requirements that are less and good for long-term forecasting. ANN can learn and generalize the relationships in multiplex datasets which enhance the chance of their applicability. Fuzzy inference system (FIS) is another system that enhances this modeling. This system is based on fuzzy logic and has also become popular. Adaptive Neuro Fuzzy Inference System (ANFIS) is a combination of FIS and ANN model. This model can extract data from input to fuzzy value (0 to 1). ANFIS is being successfully used in many hydrological fields, such as streamflow (Yaseen et. al, 2017), prediction of water level in a reservoir (Chang et. al, 2006), runoff and rainfall forecasting (Yaseen et. al, 2018). Optimizers can be used to optimize the performance of basic ANFIS model. For example, Firefly Optimization system can be used with ANFIS model (Yaseen et. al, 2017). In this study, Adaptive neuro fuzzy inference system (ANFIS), ANFIS with few optimizers were also applied in some cases to overcome the limitation of the models. The optimizers are ANFIS with FFA, ANFIS with Genetic Algorithm (GA), ANFIS with Particle Swarm Optimization (PSO), ANFIS with Ant Colony Optimization (ACO) and Multi-gene Genetic Programming (MGGP). Moreover, statistical accuracy indicators were applied for better accuracy for each model. Statistical analysis has been performed to verify model results with higher accuracy. This analysis also helps to compare all results and suggests best one among all models. Overall, all studies found the success of application of all type ANFIS models on evaporation estimation.

1.2 Objectives of this study

This study is focused on estimation of evaporation using various ANFIS-type models. Main objectives of this study are to introduce modeling of ANFIS-type methods, ANFIS with Firefly Algorithm (ANFIS-FFA), ANFIS with Genetic Algorithm (ANFIS-GA), ANFIS with Particle Swarm Optimization (ANFIS-PSO), ANFIS with Ant Colony Optimization (ANFIS-ACO) and another concept, Multigene- Genetic Programming (MGGP). Then, statistical accuracy indicator tests are performed in order to verify results from model output. Finally, to compare all the results and find the best model is another prime goal of this study. This analysis makes a comparison among results from model output and results from statistical analysis. The goal is to provide an accurate evaporation prediction that can contribute on the agriculture and hydrological model in arid area.

This study explored the skill of ANFIS-type models to improve the accuracy of monthly evaporation estimation for arid environments of Arizona, United State. Comparisons helped here finding the best model for evaporation from the available atmospheric data. The main goal of this study is to develop and monitor the statistical performance of a novel ANFIS-types model to meet the evaluation condition. The performance accuracies of the ANFIS model with optimizers are compared with the performance of the traditional ANFIS model. In the training phase, the root means square error (RMSE) is investigated as an objective function to evaluate the accuracy of the ANFIS-FFA mode. RMSE method is monitored to evaluate the accuracy of ANFIS-FFA model. The implementations of ANFIS-types are successfully tried on evaporation forecasting problem for an aid climate, Arizona. This area has been chosen because of its specific weather conditions.

1.3 Novelty of the work

Various methods have been conducted on evaporation estimation forecasting, such as empirical method, climate-based method, pan evaporation method and data driven methods. Artificial Neural Network (ANN), Fuzzy logic, ANFIS, Evolutionary Neural Network (ENN), Multi-Layer Perceptron Neural Network (MLP), Generalized Regression Neural Network (GRNN), Differential Evolution (DE) and many more are being used as data-driven methods currently. But ANFIS model with any optimizers are comparatively less used and compared for evaporation problems.

This is the first time (to the best knowledge of the author) that Adaptive Neuro Fuzzy Inference System (ANFIS) is considered as a basic model and then Firefly Algorithm, Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization and Multigene- Genetic Programming (MGGP) are being applied together and compared in the prediction of evaporation estimation.

In this study, ANFIS-type models incorporated with different algorithms have been examined in order to determine the suitable model for the study in which various output parameters are investigated. Also, the models are compared with the statistical accuracy indicator test results as well in order to verify the results to find the best fit model.

1.4 Scope of work

In this study, ANFIS-type models incorporated with five different algorithms or optimizers are investigated with statistical indices test analysis. Also, MGGP model is analysed and compared with ANFIS type model. In order to do that, data were collected from an arid climate, Arizona, USA. Therefore, the research focus is on arid climates only. Though it is an independent study, further researches can be done for different climate conditions. Investigating the specific data driven models (ANFIS-type models and MGGP model) are within the scope of this study.

Empirical methods or any climate-based methods are not being investigated by this data set, which can also be done in the future.

1.5 Outline of the thesis

The thesis is organized in five chapters. The first chapter i.e. Chapter 1 is the Introduction, which contains introductory deliberation and outlines and objectives of the study. Chapter 2 is the Literature Review in which numerical studies on evaporation estimations by using Artificial Intelligence are discussed in one section and in another section, various studies conducted on different areas using different Artificial Intelligence models and algorithms are reviewed. Also, in this section comparative studies among many ANFIS-types models are shown with some of the studies comparing statistical analysis results.

Chapter 3 consists of a Technical paper. In this paper, the implementation of Adaptive Neuro Fuzzy Inference System (ANFIS), and ANFIS in combination with three optimizers are shown, and compares the results to find the best fit model. The optimizers are Firefly Algorithm (FFA), Genetic Algorithm (GA) and Particle Swarm Optimizer (PSO). Six different measured weather variables are taken, which are maximum, minimum and average air temperature, sunshine hours, wind speed and relative humidity for the area Arizona, USA.

Chapter 4 consists of another Technical paper. This paper presents the application of ANFIS-ACO and ANFIS-MGGP and compares the results with the traditional ANFIS model again. Climate variables remain the same as the first paper for the same arid area.

Chapter 5 consists of conclusion, final remarks and suggestions for the future work. It should be mentioned that the thesis is presented in a paper-based format and the author has tried to avoid the repetition as much as possible.

1.6 Contributions and Achievements:

The results demonstrated in this thesis were obtained during a period of study at the University of Ottawa to fulfil the requirement of getting an M.A.Sc. The work was carried out under the supervision of Professor Majid Mohammadian from January 2019 to January 2020. The following manuscripts related to this research work (thesis) were submitted for publication during this period;

[1] M. Jasmine, M. Mohammadian, and H. Bonakdari, "ANFIS-Type Models for Prediction of Evaporation in Arid Climate," (Submitted) March 2020.

[2] M. Jasmine, M. Mohammadian, and H. Bonakdari, "An integrated approach based on Multigene Genetics Programming for Prediction of Evaporation in Arid Climate," (Submitted) March 2020.

Chapter 2: Literature Review

2.1 Literature Search

The method of estimating evaporation has a significant influence on the accuracy of the results. Many studies are going on in order to achieve results with higher accuracy. Research studies have been started with Pan evaporation, which is very traditional way to compute evaporation. The estimation of evaporation can be performed by pan evaporation or modeling based on environmental data (Dogana et al., 2010). Pan evaporation is not always possible due to the location, weather and difficulties of instrumental set up. Researchers have worked on climate-based models (Stephens and Stewart, 1963; Lu et al., 2005; Kisi, 2013; Benzagtha, 2014) and have faced problems related to data collection. Data are not comfortably accessible and do not always follow linear equations as climate-based methods are based on formulations. To overcome this limitation, a better modeling approach such as Artificial Intelligence is required (Dogana et al., 2010). Artificial intelligence models are becoming increasingly popular for forecasting data instead of traditional models. ANFIS model is one of them, which is also called a data-driven model (Kisi et al., 2014; Kisi et al., 2015), that can be used for different measurements, such as rainfall, streamflow, evaporation, water quality and many others.

Related works on this modeling approach include synthetic streamflow generation municipal water consumption modeling, identification of unknown pollution sources in groundwater, flood management, and sediment loss prediction. Development can be recognized in two ways; it can be either a mechanistic model or a data-driven model. The AI models have been upgraded by advanced computing modeling, which show a high level of accuracy in the prediction of various problems such as sediment transport, rainfall pattern analysis, and water irrigation. Significant developments have been observed by several researchers in the hydrological field including improvements in evaporation estimation. A comparison has been investigated by Moghaddamnia et al. (2009) on evaporation evaluation using Artificial Neural Network (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS) and the result of this comparison found that ANN was slightly better than ANFIS. The ANFIS model was compared with the regression-based method by Dogana et al. (2010) and ANFIS was declared to be the finest. Some researchers, (Kisi et al., 2012) have worked on Generalized Neuro Fuzzy model and climate-based models (Stephen and Stewart, Penman). A group of researchers (Goyal et al., 2014) has published their work on ANN, LS-SVR, Fuzzy Logic, and ANFIS on daily pan evaporation with the conclusion of Fuzzy Logic as being the best performer. Artificial intelligence method has also demonstrated advantages over the others. Another study was done by the same authors, (Kisi et al., 2014) by comparing two different ANFIS models, ANFIS-SC (subtractive clustering) and ANFIS-GP (grid partitioning) on daily evaporation. The same group worked on monthly evaporation forecasting using ANN, ANFIS-GP, ANFIS-SC and gene expression programming (GEP). ANFIS-GP was superior in both studies in 2015. Wang et al. (2017) published an analysis on pan evaporation using six methods: MLP (Multi-layer Perception), GRNN (Generalized Regression Neural Network), FG (Fuzzy Genetic), LSSVM (Least square Support Vector Machine), MARS (Multivariate Adaptive Regression Spline) and climate-based model. This study claimed MLP to be the best. Currently, few researchers (Gharbani et al., 2018) successfully implemented Quantum-behaved Particle Swarm Optimizer algorithm on evaporation forecasting which was combined with Multi-layer perceptron method.

Artificial Intelligence appraoch, such as ANN, is a well-received modeling approach for different topologies and weather conditions among many modeling methodologies. This method (Shirsath and Singh, 2010) copies the cognitive response of the human brain. It is a biologically motivated computational model that contains processing elements (neurons) and links between them (weights). ANNs have a supremacy over other models in regard to fewer data requirements, which is good for long-term forecasting. Artificial Neural Network can learn based on multiplex datasets, which enhances reliability. Fuzzy inference system (FIS) is another term that enhances this modeling. This system is based on fuzzy logic and is also becoming popular. In this study, adaptive neuro fuzzy inference system (ANFIS), ANFIS with Firefly Algorithm (Yaseen et al., 2016; Yaseen et al., 2017), ANFIS with Genetic Algorithm (Ayvaz and Elci, 2018) and PSO (Karahan.H, 2012) were analyzed to suggest the best modeling approach for evaporation.

Single gene genetic programme and multiple gene genetic programme (MGGP) are concepts inspired by a heuristic algorithm and have been used in various areas of research. For example, Raj and Rajendran (2009) proposed a simple heuristic algorithm to resolve a single-stage Fixed-charge transportation problem (FCTP) and compared the result with the traditional famous method by using benchmark problem instances. Yan, et al. (2019) used this optimizer (Multigene Genetic-Programming-Based Models) to predict initial dilution of vertical buoyant jets. This study showed a comparison of the results of Single gene genetic programing and multiple gene genetic programing and demonstrated the superiority of MGGP.

Another optimization system that has become popular in different areas of researches, is Ant Colony Optimization (ACO). Many researchers work on this model for different purposes. For example, Silva et al. (2009) applied ACO first to solve different operational activities for supply chain management. A meta-heuristic algorithm, which can also be called ACO model, has been used to solve problems. Mausavi et al. (2017) also utilized this Ant Colony Optimizer in a statistics-based study on divorce rate reduction. This optimization system follows the way an ant finds a path to reach its food or goal by avoiding all obstacles. Later on, Hong et al. (2018) used ACO on a two-stage supply chain problem and achieved a satisfactory result.

Khoob (2008) analyzed artificial neural network (ANN) of reference evapotranspiration from pan evaporation in a semi-arid area (In Iran). Maximum and minimum temperature were the weather variables for this study. In this study FAO-56 and Penman-Monteith methods were considered as standard and were compared with artificial neural network. Pan evaporation data set was used in ANN method. Comparison presented the best results which came from ANN model. Therefore, data-driven model performed better than traditional methodologies. A different methodology had tried by some researchers. Chang et al. (2010) examined the efforts of meteorological variables on evaporation estimation by self-organizing map neural network (SOMN). In this study authors proposed SOMN network to show the diversity of daily evaporation based on climate variables. Result showed the purposeful mapping of both weather variables presentation and presentation of evaporation estimation. This method can be successfully applied in daily, monthly and yearly evaporation estimation. Another study has been performed by Kuo et al., (2011). This was about evapotranspiration using backpropagation neural network (Penman–Monteith method versus pan evaporation method). Results showed that, this method was able to consider ten weather variables (maximum, minimum and average temperature, humidity, dew point, solar radiation, sunshine hour, wind speed, morning and afternoon earth temperature) and most effective weather variable combination was determined by Glycemic Index (GI) factor. This method worked well for three nodes and four weather variables (maximum and average temperature, humidity, dew point). It provides best result with higher accuracy under this combination of input. Tabari H. et al. (2010) estimated daily pan evaporation by artificial neural network (ANN) and multivariate non-linear regression analysis (MNLR). This study showed the importance of data driven methods when the data set are not available properly for pan evaporation. Humidity, temperature, solar radiation, wind speed and precipitation were considered to verify effectiveness of each variables on evaporation estimation. Finally, this study proved that the estimated data are more sensitive to temperature and wind speed, and ANN method performed better than MNLR method. Some researchers worked with daily pan evaporation by using soft computing models with limited climatic data from South- Western part of Iran in 2012. Kim et al tried to apply the multilayer perceptron-neural networks model (MLP-NNM), Kohonen self-organizing feature maps-neural networks model (KSOFM-NNM), and gene expression programming (GEP). Temperature, radiation, and sunshine duration-based input combinations were considered for calculating daily pan evaporation. Results achieved from the temperature-based 3 (TEM3) model showed the best among three models. The Mann-Whitney U test was performed in order to evaluate the rank of input combination for hypothesis analysis. Comparison revealed the superiority of soft computing methods over regression-based analysis. Another estimation using neural network for crop evapotranspiration was performed by Adeloye et al (in 2012). Self-organizing map (SOM) was performed and compared with traditional Penman-monteith method. As, comparison revealed superiority of SOM model, another comparison was done with feed forward neural network with back propagation in order to verify results. Again, SOM posed better result than another model. A comparison among extreme learning machine (ELM), Genetic algorithm neural network (GANN), Wavelet neural network (WNN), two temperature-based and three radiation-based models was examined by Feng et al. in 2016. This analysis was performed in a humid region of south-east part of China. ELM and GANN models performed with the best outcome. But temperature-based and radiation-based models can be applicable up to an acceptable boundary of precision. The same group of researchers, Feng et al. (in 2017) applied a different extreme learning machine-based model (ELM) and generalized regression neural network (GRNN) with only temperature-based data set. This experiment was performed in six station of south-west part in China. Two types of data set were calibrated; one from local source and another data set was taken from all six stations. However, ELM model came out with better result than GRNN model. Recently a study was performed on the successful application of hybrid fuzzy model with firefly optimization by Tao et al. (in 2018). This study also demonstrated the case study and importance of artificial intelligence in evaporation estimation. This model was applied in different station of Burkina Faso for different climate variables combinations. Result showed the successful application with higher accuracy in most station. In China Xuanrong et al., (2018) tried to apply back-propagation (BP) neural network algorithm to improve the uses of water resource. This study showed irrigation system solutions by predicting performances of model with different architectures. Another application of learning machine-based model was experienced by Wu et al, (2019) to predict Daily reference evapotranspiration with bio-inspired optimization algorithms. This hybridized extreme learning machine (ELM) model was Applied in China. Results promoted the ability of cuckoo search algorithm (CSA) and flower pollination algorithm (FPA) to upgrade the performance of the conventional ELM model in calculating daily evapotranspiration. An experiment had done in this current year (2019) by Yang et al. in order to predict irrigation water demand for a season. This speculation was estimated by the short-term daily evapotranspiration. A temperature-based method was designed with four combinations of wind speed data; wind speed by default, public weather forecasted wind speed, average wind speed on daily basis for long time period and yearly average wind speed. The Reduced-set Penman-Monteith (RPM) model was considered according to these four-dada set and results were compared with the Hargreaves Samani model. RPM model posed better result than Hargreaves Samani model. Though, different data combinations were proven different suitability for different weather conditions. For example, the model based on the data set provided by wind speed by default was suitable for mostly arid, humid, sub-arid and sub-humid areas.

Saggi et al. (2019) studied a H2O model framework for estimating evapotranspiration on daily basis at Northern Punjab, India. Deep Learning-Multilayer Perceptron's (DL) was applied and compared with other three models; Generalized Linear Model (GLM), Random Forest (RF), and Gradient-Boosting Machine (GBM). Results indicated the superiority of Deep Learning-Multilayer Perceptron's model among four. Random Forest (RF) and Gene-expression (GE) were also applied in order to calculate evapotranspiration with less data set in China. A group of researchers, Wang et al, proved a successful application of these models in the current year, (2019). Gene-expression was able to present relationship between dependent and independent variables, which is very important for calculating evaporation in irrigation. Results advocated suitable application of both methods, though Random Forest (RF) expressed slightly better result than Gene-expression.

A new approach was implemented by a group of researchers (Ferreira et al., 2019) in order to calculate evapotranspiration with utilizing less data set in Brazil. New approaches were; artificial neural network (ANN) and support vector machine (SVM). The FAO-56 Penman-Monteith (FAO-56 PM) equations were used as a benchmark for this study. Data were measured based-on temperature and relative humidity or only temperature. ANN presented best result with higher accuracy in all circumstances. But, ANN without clustering and four days as input was recommended.

Artificial neural network (ANN) is a well-established approach on predicting evaporation and evapotranspiration (ETo) modeling (Kumar et al. 2002; Sudheer et al. 2003; Trajkovic et al. 2003;

Keskin and Terzi 2006; Deswal and Pal 2008; Rahimi Khoob 2009). Bruton et al. (2000) examined the potential of ANN in estimating evapotranspiration by considering precipitation, air temperature, solar radiation, wind speed and relative humidity as input data. They proved the efficiency of ANN model with the reliable performance of estimating evaporation rate. Couple of research groups, Sudheer et al. (2002) and Terzi and Keskin (2005) applied ANN model successfully to predict Class A pan evaporation and daily evaporation respectively by adding meteorological variables such as relative humidity, air temperature, wind speed and sunshine hours as input. Kisi et al. operated the neuro-fuzzy model to estimate the daily evaporation using climate variables such as, air temperature, solar radiation, wind speed, air pressure, and relative humidity as the inputs of the neuro-fuzzy model in 2006. They advocated the capability of neuro-fuzzy technique on evaporation process from the available meteorological data. Another approach, support vector machines were compared with ANN model by Eslamian et al. (2008) to estimate monthly evaporation. Analysis revealed the similar results of ANN and support vector machines approaches. However, the support vector machines technique predicted monthly evaporation better than the ANN method in some cases. A comparison of ANN and genetic algorithm was performed by Kim et al. (2008), which confirmed the capabilities of ANN and genetic algorithm models as effective tools for estimation of evaporation and evapotranspiration. The neural networks using varied input combinations of meteorological variables are trained using various training algorithms and then tested. Therefore, it becomes very effective.

Artificial intelligence models are becoming very popular for predicting data instead of traditional models. The ANFIS model is one of them, which is also a data-driven model that can be used for various measurements, such as rainfall, streamflow, evaporation, water quality and many others. The ANFIS is a very useful model based on fuzzy logic and ANN. Several applications of this model can be found to control automatic trains, nuclear reactors, chemical reactors, as well as different purposes related to engineering, business, etc. For example, Chang, et al. applied the ANFIS model to predict reservoir water level in 2006. Identifying the relation between the input and output parameters without direct physical consideration is a characteristic of this model. Again, Kisi et al., showed the application of the ANFIS model in evaporation estimation daily in 2014. The ANFIS model was successfully applied in this study. For the modeling of reservoir performance, and to resolve the problem regarding data uncertainty or inexactness, fuzzy logic is a highly recommended system. It can work well during training sets which carry noise and/or measurement errors, and can also adapt to situations over time, even in changing environments. Information-processing quality is another characteristic of this model. This can also be described as a “feed forward neural network with back propagation training algorithm”, which is used for developing the ANN modeling approach because it is a commonly used and reliable approach in hydrological modeling. ANFIS can be employed for modeling numerous processes, such as motor fault detection and diagnosis, power system dynamic load, wind speed and forecasting systems, demonstrating its ability to create and extending and identifying the best fit data set or model. ANFIS allows the difficult conversion of human intelligence to fuzzy systems, and the extraction of fuzzy rules to numerical data (Chang et al., 2006).

Artificial intelligence has been successfully applied to calculate evaporation and evapotranspiration by using restricted climate variables in the last few years (Tao. et al., 2018). Combining the benefits of the fuzzy logic and neural network system in order to solve non-linear and ambiguous problems is the greatest success of ANFIS model (Jang, 1993). As ANFIS model is a combination of FIS and ANN, this type of fuzzy logic model can be beneficial to solve unclear and impressive data set issues. On the other hand, ANFIS model was exposed with few drawbacks because of the internal parameter optimization system. Therefore, combining ANFIS model with some nature-based optimizers or algorithms can be an alternative to elevate this situation. A hybrid approach was successfully applied recently in order to precise data for hydraulic modeling. The integration with optimizer showed the improvement of the performance of ANFIS model as it deducted the optimum output for a calculated problem, and decreased the computation time (Ghorbani et al., 2017). The nature-based optimizers or algorithms can attract researchers due to their capability of magnifying the performances of AI models. Firefly algorithm has drawn attention of the researchers as a few analyses were proved its efficiency on the accuracy of ANFIS model (Yaseen et al., 2017). According to these case studies, firefly algorithm is very efficient and trustable in an AI based model because of its coherency and solving ability for both local and global data set problems. Despite the dynamic nature of the FA, its application in evaporation estimation is yet to be examined. Therefore, to the best knowledge of the author, this is the first implementation of the ANFIS-FFA model for evaporation estimation over Arizona, USA. Also, few more optimization approaches are implemented with the traditional ANFIS model for the first time; particle swarm optimization, and ant colony optimization. Ant Colony Optimization (ACO) has become popular in different areas of researches. For example, ACO was first applied by Silva et al. to solve operational activities for supply chain management (2019). ACO model is considered as a meta-heuristic algorithm, which has a potential to solve problems. This optimization system follows the way an ant finds a path to reach its food or goal by avoiding all obstacles. Ant Colony Optimizer was also utilized by Mausavi, et al. in a statistics-based study on divorce rate reduction (2017). Later, Hong, et al. used ACO on a two-stage supply chain problem and achieved a satisfactory result (2018).

Particle swarm optimization approach is based on the nature of bird or fish swarm. This approach had already implemented successfully by Gharbani et al., (2018). This analysis revealed the Quantum-behaved Particle Swarm Optimizer algorithm on evaporation forecasting which was combined with Multi-layer perceptron method.

Genetic algorithm with ANFIS model and multi gene-genetic programming is successfully applied for the first time to improve the result with higher accuracy. Yan, et al. (2019). applied this optimizer (Multigene Genetic-Programming-Based Models) to measure initial dilution of vertical buoyant jets. This study compared the results of Single gene genetic programing and multiple gene genetic programing and advocated the superiority of MGGP model. Multiple gene genetic programing (MGGP) is a concept based on heuristic algorithm and draws attention to the researchers in different areas of interest. The finding of this research is significant for the evaporation estimation forecasting as an alternative modelling strategy and particularly within the context of arid climate in Arizona, USA, where data are available with six weather climate variables; relative humidity, wind speed, sunshine, minimum-maximum- average temperature. These six variables are the most important parameters essential for evaporation.

Statistical analysis test plays vital role on the accuracy analysis of this sort of models. Many accuracy indicators are being practiced performing these tests. Coefficient of determination and root mean square error are very popular among the indices. Coefficient of determination evaluates accuracy of model by its working pattern and measures its future output. Root mean square error (RMSE) counts differences between predicted and observed value. The deviations are basically predicational error that occurs during computational period. RMSE basically integrates the values of errors or deviations into a single measure of predicted power. Relative accuracy is an important statistical index that observes how close a measured value is to a standard value on relative terms.

Absolute accuracy test can determine how close a measured value is to a known absolute true value and mean absolute error is basically an average of absolute error. Biasness is a dominant index that measures how far the expected value is from the true value of the variables being evaluated. Variable account for (VAF) is necessary for acute analysis of a model. VAF needs to be counted in some cases where model looks simple but can go wrong by using wrong variables.

2.2 Methodologies

This section presents a short introduction to ANFIS, ANFIS-FFA, ANFIS-GA, ANFIS-PSO, ANFIS-ACO and MGGP models. Data were collected from the government datasets of United

Figure 2.1: Study area, (a) United State (b) Phoenix, Arizona (c) location of data (Source: maps.ie/coordinates).

State and these data sets based on pan evaporation. Input data was considered with six climatic variables for all six models: sunshine, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed. After collecting data, 70% of the total data set was used for the training purpose and 30% of the data set was used for verifying the efficiency of trained matrix.

To understand ANFIS model properly, FIS and ANN models are needed to understand, since Adaptive Neuro Fuzzy Inference System (ANFIS) is a combination of Fuzzy Inference System (FIS) and Artificial Neural Network (ANN).

2.2.1 Fuzzy Inference System (FIS)

Fuzzy Inference System (FIS) was first proposed by Chang et al. to solve ambiguous problems in 2006. The FIS has become a very popular and successful model because of the fuzzy logic which can convert fuzzy idea to a binary set. The applications of this approach are on control processing of machine systems, such as, automatic trains, nuclear reactors, chemical reactors, as well as different purposes related to engineering, business, etc. For the modeling of reservoir performance, and to resolve the problem regarding data uncertainty or inexactness, fuzzy logic is a highly recommended system (Chang, 2006). It also helps to provide a substructure to map input zone to output zone. For example, if a component number belongs to one fuzzy set, the range of the membership function would be 0 to 1, as identified by deleting the sharp boundary separating member from the non-members set (Goyal et al., 2014). This method provides a simple technique to eliminate any ambiguous, unclear or misleading data. Usually the Fuzzy Inference System (FIS) is built based on the consideration that the cluster system generates a cluster center during data collection period. This cluster center represents the behavior of this system. Although the cluster center is responsible for controlling behavior, the fuzzy logic toolbox of MATLAB is the library function which controls the creation, editing and execution of the FIS system. This toolbox can be used to develop modeling for evaporation systems.

The limitation of this system is not having proper methodical way to design a fuzzy controller (Chang et al., 2006). The ANN system is able to arrange input and output in pairs, manage the structure accordingly make it ready to calibrate. Therefore, an Adaptive Neuro-Fuzzy Inference System (ANFIS) is suggested to self-arrange and convert FIS data for forecasting the water or evaporation information from a reservoir system.

2.2.2 Artificial Neural Network (ANN)

The artificial neural network (ANN) relies on a hidden layer and is able to perform non-linear mapping between input patterns and target values. Nodes are the key elements of artificial neural networks and are also familiar as processing elements. Nodes are connected to each other by other elements known as weights, and they are distributed in the layers of the network. All processing nodes are arranged into layers, and every layer is completely interconnected to the next layer. There is no interrelation between the nodes of the same layer.

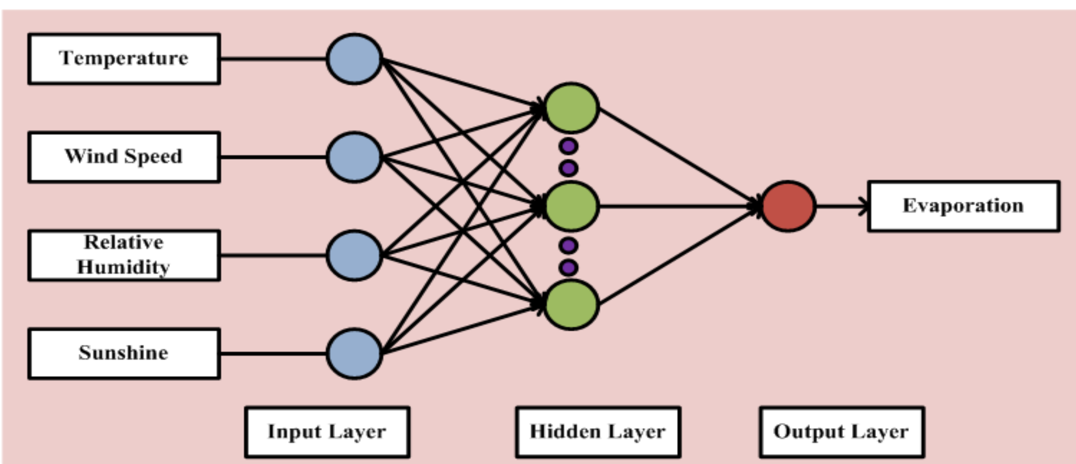


Figure 2.2: Structure of ANN model, (based on Benzaghta, 2014).

The input layer of the model acts as a distribution structure for the meteorological data that are presented to the network. The following processing layer, after the input layer, is called the hidden layer. The final processing layer is called the evaporation layer. This sort of ANN is called multilayer perceptron (MLP). A three-layer ANN structure is exemplified in Figure 3 which presents the structure of the ANN model which was employed in this analysis.

Each node in a layer collects and processes weighted inputs from a preceding layer and transfers its output to nodes in the following layer through connections. Each connection carries a weight, which is considered as a numerical estimation of the connection strength. During a training process (at each iteration), the initial assigned weight values are gradually corrected, and a comparison is done between forecasted outputs and known outputs. Proper weight adjustments are mandatory to reduce errors and back propagation is a way that can be used to find the proper weight. Weighted summation of inputs to a node is converted to an output according to a transfer function. The weights were adjusted according to the comparison of ANN output and the target until they matched. This transfer function is basically a sigmoid function and the model can be called a “feed forward neural network with back propagation (BPNN) training algorithm”, which is used for developing the ANN modeling approach because it is a commonly used and reliable approach in hydrological modeling. The following reasons support why the ANN model is well accepted as a computational tool:

(a) It identifies the relation between the input and output parameters without direct physical consideration.

(b) It is able to work well even when the training sets carry noise and/or measurement errors.

(c) It has the adaptability to maintain situations over time.

(d) It includes other basic information-processing qualities and is ready to use once trained.

2.2.3 Adaptive Neuro Fuzzy Inference System

The ANFIS model is a mixture of Fuzzy Inference System (FIS) and Artificial Neural Network (ANN). This model is adopted mainly due to its good capacity of extraction of data from input to fuzzy values in a range of 0 to 1. ANFIS is a multilayer feed-forward network, which is capable of arranging and converting data. This system can be employed for modeling numerous processes, such as motor fault detection and diagnosis, power system dynamic load, wind speed and forecasting systems, demonstrating its ability to create, extend and identify the best fit data set or model. ANFIS can allow the difficult conversion of human intelligence to fuzzy systems, and extraction of fuzzy rules to numerical data (Chang. et al., 2006). There are two basic components in this system, one is “node” and another one is “rule” where rule determines the relationship between input and output and node is membership Functions (MF). ANFIS requires feature extraction rules applied to the input data, such as “IF-THEN” rules.

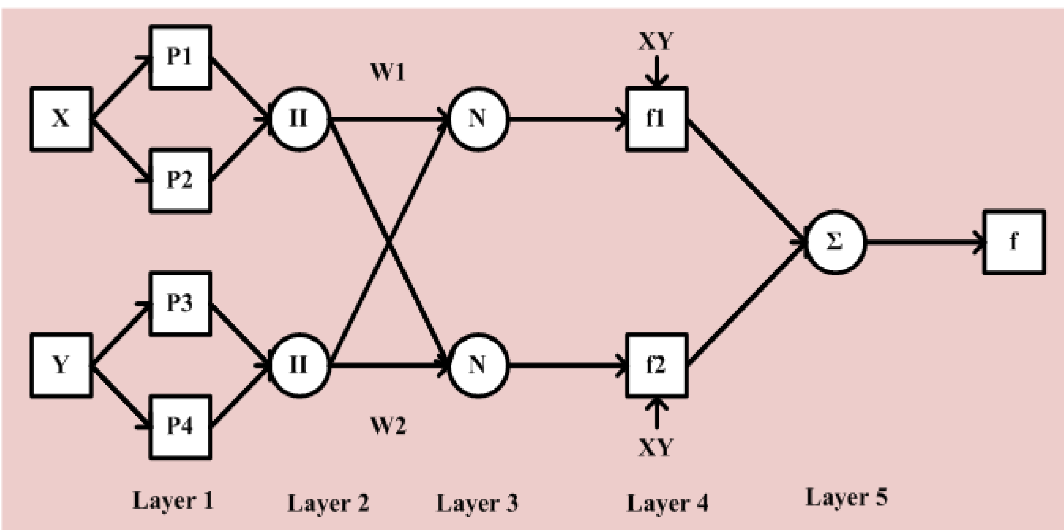


Figure 2.3: ANFIS architecture for two-inputs and 5 layers. Layer 1 (Fuzzy Rules), Layer 2 (Input MF), Layer 3 (Fuzzy Neurons), Layer 4 (Output MF), Layer 5 (Summation and Weight), (based on Yaseen et al., 2017).

The structure of ANFIS can be understood from figure 1. This figure demonstrates the details of an ANFIS model. Layer 1 represents the input nodes and each node of this layer creates membership grades; they belong to one of the fuzzy sets by utilizing the membership function (Chang et al., 2006). Layer 2 represents rule node where the AND operator is implemented to hold one output, which is the result of the previous layer. Therefore, the output of layer 1 becomes the input for layer 2. Layer 3 is considered an average node; whose main goal is to compute the ratio of single rule's strength to the total of all rule firing strengths (Chang et al., 2006). Firing strength is the degree to which the fuzzy rule from the previous segment is satisfied, and it forms the output function for the rule. Layer 4 is named the consequent node; whose main purpose is to calculate the efficiency of each rule with respect to the total output. Finally, layer 5 consists of output nodes. The output nodes result from adding up all the incoming signals. This layer also defuzzifies the system by modifying each fuzzy rule, which follows a crisp output in this layer.

The main limitation of this model is that it is a time dilated model during the training period, and parameters must be estimated.

2.2.4 Firefly Algorithm (ANFIS-FFA)

To overcome the limitation of the traditional model (ANFIS), ANFIS-FFA is adopted. The mechanism of FFA is based on the nature of the firefly (flashing behavior). ANFIS-FFA method is applied during the training phase to determine the best set of data. Figure 3 (Yaseen. et al., 2017) presents the flowchart that clearly demonstrates the working principal of Firefly algorithm.

This model (FFA) formation depends on three basic principles:

1. Each firefly is able to engage another firefly,
2. Attractiveness between two fireflies are calculated by the light intensity of each firefly,
3. Brightness is related to the light released by fireflies (Yaseen et al., 2017).

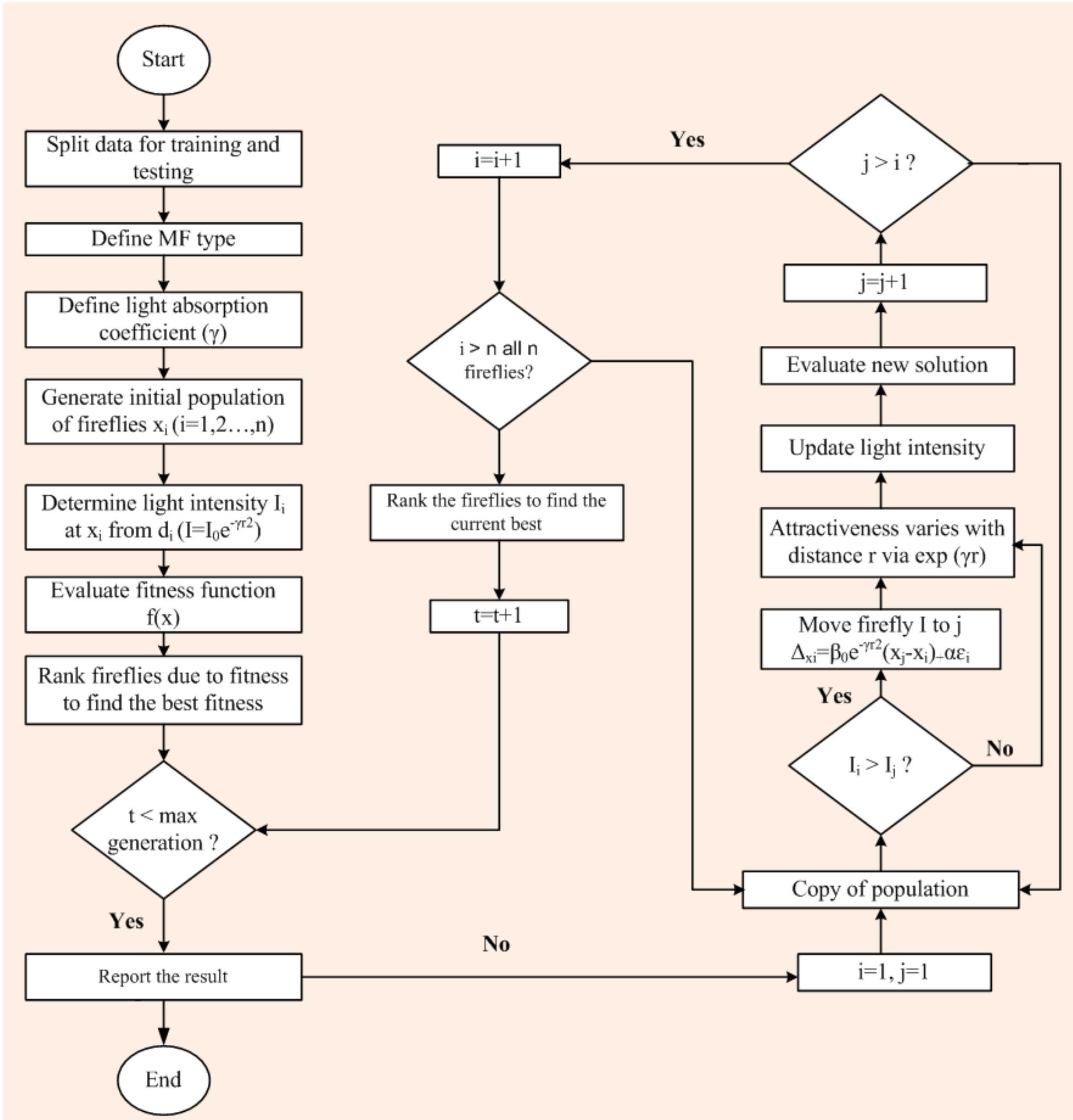


Figure 2.4: The flowchart of a ANFIS model optimized with FFA, (based on Yaseen et al., 2017).

Thus, the objective function of the FFA model is introduced by the intensity of the light produced by each firefly and the brightness of the firefly. To integrate the ANFIS model with FFA, these three coefficients α, β and γ are required to be adjusted. This adjustment can be done by trial and error process (Yaseen et al., 2017). As fireflies are attracted by each other, the movement of two fireflies can be calculated by empirical formulae.

2.2.5 Genetic Algorithm (ANFIS-GA)

This fuzzy-genetic algorithm consists of a genetic algorithm and an adaptive fuzzy inference system (AFIS), (Wang et al., 2017). The AFIS is established based on fuzzy logic and the genetic algorithm (GA) is based on the characters of natural genetics and its selection system. GA includes three major stages: (1) population initialization (2) GA operators (3) evaluation (Wang et al., 2017). This GA system can solve large space problems efficiently and optimize complicated functions. This model is highly useful for evapotranspiration calculations. Any hybrid model (hybrid ANFIS) can optimize the MF by using GA. This fuzzy-genetic algorithm has a potential to minimize model errors (Wang et al., 2017). In the classical form of GA, decision variables are hidden as binary strings. Chromosomes are formed by the consecutive combinations of genes, which are related to possible monitoring networks, i.e., candidate solutions. The development begins from a population of random chromosomes and thus, generations form. In each generation, the fitness of the whole population is estimated. Then, based on the fitness, multiple chromosomes are stochastically adopted from the current population and adjusted by utilizing genetic operators such as crossover and mutation to create a new population. The current population is applied in the following iteration of the algorithm (Tamer and Alper, 2018). The main objective is to define a small number of effective monitoring wells in the suggested approach by following this solution sequence.

2.2.6 Particle Swarm Optimization (ANFIS-PSO)

The PSO technique was invented by Kennedy and Eberhart (1995), based on the characteristics of bird and fish swarms in a multi-dimensional area; for example, looking for food and running away from hazards (Kennedy and Eberhart, 1995). Every element in this algorithm is identified as a “particle”, and particles create the density (population) and each density is identified as a “swarm”. Every particle is considered as a candidate for the answer to the question in this algorithm. The swarm and particle values of this technique depend on the chromosome and density (population) items, which are similar to the Genetic Algorithm (Karahan.H, 2012). PSO is a trial and error solution procedure that explores the characteristics of swarm particles in a multi-dimensional exploration zone. The extent of the exploration zone is similar to the number of the unnamed variables of the explored problem, and the number of individuals in the swarm indicates the density of the swarm. Highest and lowest probable values of the variables are selected according to the limit of the probable values of the variables prior to resolving the complication. The highest and lowest speeds are estimated by utilizing these parameters (Karahan.H, 2012). Mostly, the determination of the optimum number of hidden modes is determined via iteration method by progressively varying the number of modes in the hidden layer. But the computation process is different in the case of large data sets because of the higher expenses of developing a significant number of models. Hence, an improvement in computational efficiency is required to overcome this problem. PSO can be an alternative to optimize the number of hidden nodes. PSO is able to optimize with a large possibility and high meeting (convergence) rate. In this analysis, the aim of the PSO algorithm is to minimize the objective function. The levels of computation of this process using PSO are given below (Karahan.H, 2012):

Level 1. Starting with the search variables:

Level 2. Estimation of the particle speed (maximum and minimum) in all direction based on mathematical expression.

Level 3. Calculation of initial particle positions and velocities based on mathematical expression.

Level 4. Estimation of objective function based on mathematical expression.

Level 5. Upgradation of vector and dimension that carries the finest location established by the swarm.

Level 6. Estimation of inertial weight value.

Level 7. Modify the speed of the particle.

Level 8. When the permitted value is lower than absolute particle velocity can be determined by formulation.

Level 9. Modify the particle positions based on equation mathematical expression.

Level 10. The particle is positioned at the unlimited searching range when the particle position is not situated in the searching range.

Level 11. Controlling diversity based related formula.

Level 12. Study the ending requirements. If ending requirements are not fulfilled, repeat Step 4.

2.2.7 Ant Colony Optimization (ANFIS-ACO)

Ant colony optimization system is inspired by the nature of ants. This optimization system follows the way an ant finds a path to reach its food or goal by avoiding all obstacles. ACO consists of ant behaviour and a pheromone matrix. ACO algorithm builds a pheromone matrix which is an integrated record of optimization steps. This matrix is easily accessible during the optimization process. This approach functions in few steps:

1. Input data: data inserted with the consideration of input variables, sunshine hour, relative humidity, wind speed, maximum, minimum and, average temperature.

2. Data initialization: This step starts with the initialization of parameters,

3. Solution generation and evaluation:

An ant creates a path in each stage by following a simulation-oriented process, which is shown step by step below;

Step 1: Data selection.

Step 2: Number selection (zero to one).

Step 3: Accomplishment of cumulative probability for the output data.

Step 4: Output data distribution.

Step 5: Again, number selection (zero to one).

Step 6: Accomplishment of cumulative probability from stage 2, for step 4.

Step 7: Finding the upstream unit where the cumulative probability value meets the uneven number from step 5. If the number is zero, go to the 5^{th} step.

Step 8: Upgradation of all data. If demand is not satisfied repetition from the 5^{th} step is required.

Step 9: Termination. If demand is not satisfied repetition from the first step is required.

4. Probability and visibility function.

5. Pheromone matrix updating.

6. Termination and output.

When the stopping conditions (iterations) are fulfilled, the algorithm is terminated. Otherwise, the number of iterations is increased, and the entire process is redone. At the end, the best data set can be obtained by the completion of a heuristic run.

2.2.8 Multi Gene-Genetic Programming (MGGP)

MGGP is one of the modern alternatives to GP, which has an admired evolutionary skill that can be successfully applied to data-driven nonlinear models. This new approach is based on the multigene genetic-programming (MGGP) technique, which is an upgrade of single gene-genetic programming (SGGP). By using training data, MGGP can spontaneously develop a distinct model and it does not need to define the model structure ahead of time. This helps in developing a mathematical model and also ignore errors, such as errors regarding judgments related to the model structure. MGGP has two primary benefits: (1) it is able to create multiple genes and each gene of the MGGP model is a conventional GP gene. Therefore, the accuracy of MGGP is higher compared with the traditional GP approach. (2) The order of the nonlinear term of a single gene is less because, each gene of MGGP obtains only a few tree depths, and therefore, the MGGP model becomes more concise. Basically, this approach builds an imperial model in this process. The first generation is called parent genes, and more generations are required to obtain the best set of equations, because the fitness of the first generation is always lower. Generations are developed by following three steps: reproduction, crossover and mutation (Yan and Mohammadian, 2019). The second generation consists of child genes which are formed by switching the sub-trees of the first generation. Mutation starts, when the crossover part comes to an end. In this process, the sub-trees are substituted with a new component. To build an entire MGGP model, numerous generations are required.

2.2.9 Model Accuracy Indicator

The performances of all six models may be individually evaluated using statistical analysis to monitor accuracy with respect to the evaporation forecasting data. The accuracy indicators for ANFIS, FFA, PSO, ACO, GA and MGGP models are calculated in terms of coefficient of determination (R^2), Nash-Sutcliffe coefficient (NSE), root mean square error (RMSE), Mean Absolute Error (MAE), Variance Account For (VAF), Absolute Relative Error (MARE), Scatter Index (SI), Bias, Mean Absolute Error (MAE), Root Mean Square Relative Error (RMSRE). The root mean squared error (RMSE) represents a good measure of the goodness of fit at high parameter values, while the relative error (MARE) provides a more balanced idea of the goodness of fit at moderate and low values. These two indices have the same scale and units as the observational data. The R^2 coefficient measures the correlation of the predicted values with the observational data, whereby the closer the coefficient is to one, the greater the correlation. The value of this coefficient does not interfere with the data unit considered. The SI index is the relative form of RMSE. The performance factor of the model expressed as the Nash-Sutcliffe (Nash and Sutcliffe, 1970) error criterion (E_{NSC}) was used to evaluate the predictive power of the model. A value of unity for the E_{NSC} indicates optimum conformity between predicted and observed data. Both R^2 and E_{NSC} may be expressed in percentages. The closer their magnitude to 100, the better the performance of the model. All of them can be calculated from designed formulations, which are presented in the appendices section.

2.2.10 Model Information

"MATLAB" was considered as programming language. The Codes have not been developed by the author and are taken from available sources. The idea of related optimizers has been taken from different research fields where these optimizers have already been used. Therefore, the code for optimizers were collected from the available sources. Run time is an important factor to be considered which is shown in the following table.

Table 2.1: Run time of six models.

Model Name	Run Time
ANFIS	5 to 10 minutes
ANFIS-FFA	30 minutes to 3 hours
ANFIS-PSO	10 to 30 minutes
ANFIS-GA	10 to 30 minutes
ANFIS-PSO	10 to 30 minutes
MGGP	10 to 15 minutes + 30 minutes (time for solving equations)

Run time was long for ANFIS-FFA and MGGP, as MGGP model needs extra time for solving equations. Other than this, remaining four models took average ten to thirty minutes to run, which is reasonable.

Chapter 3

ANFIS-Type Models (ANFIS, ANFIS-FFA, ANFIS-GA, ANFIS-PSO) for Prediction of Evaporation in Arid Climate.

3.1 Introduction

Currently, water deficiency is increasing and becoming a challenge for human society. It is increasingly becoming the most important environmental limitation, which is limiting plant growth. For example, over 30 arid and semi-arid countries are expected to experience water deficiency in 2025 (Benzagtha.MA, 2014). This will limit agricultural development, threaten food supplies and inflame rural poverty. Evaporation estimations are essential to the controlling and modeling analysis of integrated hydrological resources connected to hydrology, agricultural business, arboriculture, irrigation, flooding and lake ecosystems. Evaporation is described as the reduction of deposited water due to the conversion of liquid phase to steam phase, which is influenced by the climate situation such as weather, wind velocities, relative humidity and sunshine. According to The World Meteorological Organization (WMO), more than half of the total inflow (rainfall or any other sources) to Lake Victoria in the U.S. is lost due to evaporation, which results in relatively humid conditions (Benzagtha.MA, 2014).

The evaluation of evaporation from reservoirs in arid and semi-arid areas is also important. For example, Libya has built one of the largest civil engineering groundwater pumping and transferring systems to overcome water limitations and climate hindrance (high temperature and low rainfall). This project is known as the Manmade River Project (MRP), (Benzagtha.MA, 2014). The purpose of this project was to supply the water demand of Libya by pumping underground water underneath the Sahara Desert and transferring it using a network of huge underground pipes, especially for irrigation. The high cost of water pumping, and the lack of appropriate planning are the main concerns. In Egypt's Lake Nasser (located in an arid area), where the Nile's water is stored, downstream water loss due to evaporation is estimated to be 3 meters in depth, or double that of Lake Victoria (Benzagtha.MA, 2014). In Australia, it is calculated that around 95% of the precipitation evaporates and has no contribution to runoff (Benzagtha.MA, 2014).

The estimation of evaporation can be performed by pan evaporation or modeling with environmental data (Dogana et al., 2010). Pan evaporation is not always possible due to the location, weather and difficulties of instrumental set up. Researchers have worked on climate-based models (Stephens and Stewart, 1963; Lu et al., 2005; Kisi.O, 2013; Benzagtha.MA, 2014) and have faced problems related to data collection. Data are not comfortably accessible and do not always follow linear equations as climate-based methods are based on formulations. To overcome this limitation, a better modeling approach such as Artificial Intelligence is required (Dogana et al., 2010). Artificial intelligence models are becoming increasingly popular for forecasting data instead of traditional models. ANFIS model is one of them, which is also called a data-driven model (Kisi et al., 2014; Kisi et al., 2015), that can be used for different measurements, such as rainfall, streamflow, evaporation, water quality and many others.

Related works on this modeling approach include synthetic streamflow generation municipal water consumption modeling, identification of unknown pollution sources in groundwater, flood management, and sediment loss prediction. Development can be recognized in two ways; it can be either a mechanistic model or a data-driven model. The model has been upgraded by advanced computing modeling, which shows a high level of accuracy in the prediction of sediment transport, rainfall pattern analysis, and water irrigation. Significant developments have been observed by several researchers in the hydrological field including improvements in evaporation estimation. A comparison has been investigated by Moghaddamnia et al. (2009) on evaporation evaluation using Artificial Neural Network (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS) and the result of this comparison found that ANN was slightly better than ANFIS. The ANFIS model was compared with the regression-based method by Dogana et al. (2010) and ANFIS was declared to be the finest. Some researchers, (Kisi et al., 2012) have worked on Generalized Neuro Fuzzy model and climate-based models (Stephen and Stewart, Penman). A group of researchers (Goyal et al., 2014) has published their work on ANN, LS-SVR, Fuzzy Logic, and ANFIS on daily pan evaporation with the conclusion of Fuzzy Logic as being the best performer. Artificial intelligence method has also demonstrated advantages over the others. Another study was done by the same authors, (Kisi et al., 2014) by comparing two different ANFIS models, ANFIS-SC (subtractive clustering) and ANFIS-GP (grid partitioning) on daily evaporation. The same group has worked on monthly evaporation forecasting using ANN, ANFIS-GP, ANFIS-SC and gene expression programming (GEP). ANFIS-GP was superior in both studies in 2015. Wang et al. (2017) has published an analysis on pan evaporation using six methods: MLP (Multi-layer Perception), GRNN (Generalized Regression Neural Network), FG (Fuzzy Genetic), LSSVM (Least square Support Vector Machine), MARS (Multivariate Adaptive Regression Spline) and climate-based model. This study claimed MLP to be the best. Currently, few researchers (Gharbani et al., 2018) successfully implemented Quantum-behaved Particle Swarm Optimizer algorithm on evaporation forecasting which was combined with Multi-layer perceptron method.

Artificial Intelligence, such as ANN, is a well-received modeling approach for different topologies and weather conditions among many modeling methodologies. This method (Shirsath and Singh, 2010) copies the cognitive response of the human brain. It is a biologically motivated computational model that contains processing elements (neurons) and links between them (weights). ANNs have a supremacy over other models in regard to fewer data requirements, which is good for long-term forecasting. ANNs can learn based on multiplex datasets, which enhances reliability. Fuzzy inference system (FIS) is another term that enhances this modeling. This system is based on fuzzy logic and is also becoming popular. In this study, Adaptive neuro fuzzy inference system (ANFIS), ANFIS with Firefly Algorithm (Yaseen et al., 2016; Yaseen et al., 2017), ANFIS with Genetic Algorithm (Ayvaz and Elci, 2018) and PSO (Karahan.H, 2012) were analyzed to suggest the best modeling approach for evaporation.

The main objectives of this study are to evaluate the performance of the ANFIS model, ANFIS with firefly optimization algorithm (ANFIS-FFA), ANFIS-GA and PSO. Then, all the results were compared using statistical analysis and the best model was determined. This study explored the ability of the ANFIS model to improve the accuracy of daily evaporation estimation for arid environments in the United States. Comparisons aided in finding the best model for evaporation from the available atmospheric data.

3.2 Data Description

Arizona is the sixth biggest state of USA which is situated next to the state of California. The area of this state is 113,000 square miles and the weather condition is quite caustic with tropical summers and muggy winters. Phoenix is the capital of Arizona stare which is in the Northeastern part of Sonoran Desert; therefore, it has a hot desert climate condition. This city has an agricultural neighborhood which is closed to the confluence of the Salt and Gila river. This study is about the hot climate and being closed to an agricultural neighborhood is another reason of selecting this area as a study area. Figure 1 shows the study area, which is 355.7m higher from sea level, with 33.4258 latitude and -111.9217 longitude.

Figure 3.1: Study area, Phonix, Arizona (Source: maps.ie/coordinates).

Table 3.1 summarizes the statistical indices of test, training and all data used in this study. This table shows the calculation for all data set, training and test data set. This calculation was done in order to observe the nature of data set. In order to do that, Skewness, Kurtosis, Coefficient of Variation, Standard deviation, first and third Quarters were estimated, where, N is number of data, Min and Max is Minimum and Maximum of data, 1st Q and 3dr Q is first and third Quarters, Avg is average, SD is Standard Deviation, CV is Coefficient of Variation.

Table 3.1: Statistical indices of evaporation data with subsections; total, train, test.

Statistics	N	Min	1st Q	X50	3rd Q	Max	Avg	SD	CV (%)	Skewn ess	Kurtosis
All	85	44	82.5	158	254.5	331	172.30	89.48	51.93	0.0657	-1.446
Train	59	44	83	183	273.0	331	178.28	91.79	51.48	0.0145	-1.483
Test	26	49	74.75	154	247.25	298	158.73	82.40	51.91	0.1169	-1.500

Standard deviation shows the distribution nature of data set. For example, standard deviation of test data set is 82.40 and average value of data is 158.73. That means, most of the test data lies between 78.33 (158.73-82.40=78.33) to 241.13 (158.73+82.40=241.13). Coefficient of Variation shows the precision of data set in this table. That parameter presents the ratio of standard deviation and mean value in percentages. Two more statistical indices; Skewness and Kurtosis were calculated in order to complete the survey.

3.3 Methodology

This section presents a short introduction to ANFIS, ANFIS-FFA, ANFIS-GA and ANFIS-PSO models. Data were collected with the consideration of six climatic variables for all four models: sunshine, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed, as they are the most important parameters to calculate evaporation. After collecting data, 70% of the total data was used for the training purpose and 30% of the data was used for verifying the efficiency of the trained matrix. Two combinations of data set were trained and tested to check the results.

3.3.1 Adaptive Neuro Fuzzy Inference System

The ANFIS model is a mixture of Fuzzy Inference System (FIS) and Artificial Neural Network (ANN). This model is adopted mainly due to its good capacity of extraction of data from input to fuzzy values in a range of 0 to 1. ANFIS is a multilayer feed-forward network, which is capable of arranging and converting data. This system can be employed for modeling numerous processes, such as motor fault detection and diagnosis, power system dynamic load, wind speed and forecasting systems, demonstrating its ability to create, extend and identify the best fit data set or model. ANFIS can allow the difficult conversion of human intelligence to fuzzy systems, and extraction of fuzzy rules to numerical data (Chang. et al., 2006). There are two basic components in this system, one is “node” and another one is “rule” where rule determines the relationship between input and output and node is membership Functions (MF). ANFIS requires feature extraction rules applied to the input data, such as “IF-THEN” rules. Equation (9) and (10) show the rules for an ANFIS model for two inputs (x and y) and one output f.

Rule 1: IF x is P_1 and y is Q_1, then $f_1 = p_1 x + q_1 y + r1$ ..……………………..……..(1)

and

Rule 2: IF x is P_2 and y is Q_2, then $f_2 = p_2 x + q_2 y + r2$ ……………………..………..(2)

For a five-layer ANFIS structure, first layer defines the input nodes and each node of this layer creates membership grades; they belong to one of the fuzzy sets by utilizing the membership function (Chang et al., 2006). Second layer represents rule node where the AND operator is implemented to hold one output, which is the result of the previous layer. Therefore, the output of first layer becomes the input for layer 2. Third layer computes the proportion of single rule’s strength to the total of all rule firing strengths (Chang et al., 2006). Firing strength is the degree to which the fuzzy rule from the previous segment is satisfied, and it forms the output function for the rule. Fourth layer calculates the efficiency of each rule with respect to the total output. Finally, fifth layer consists of output nodes. The output nodes result from adding up all the incoming signals. This layer also defuzzifies the system by modifying each fuzzy rule, which follows a crisp output in this layer. The main limitation of this model is that it is a time dilated model during the training period, and parameters must be estimated. The flow chart has been drawn for this study, as shown below in Fig. 3.2.

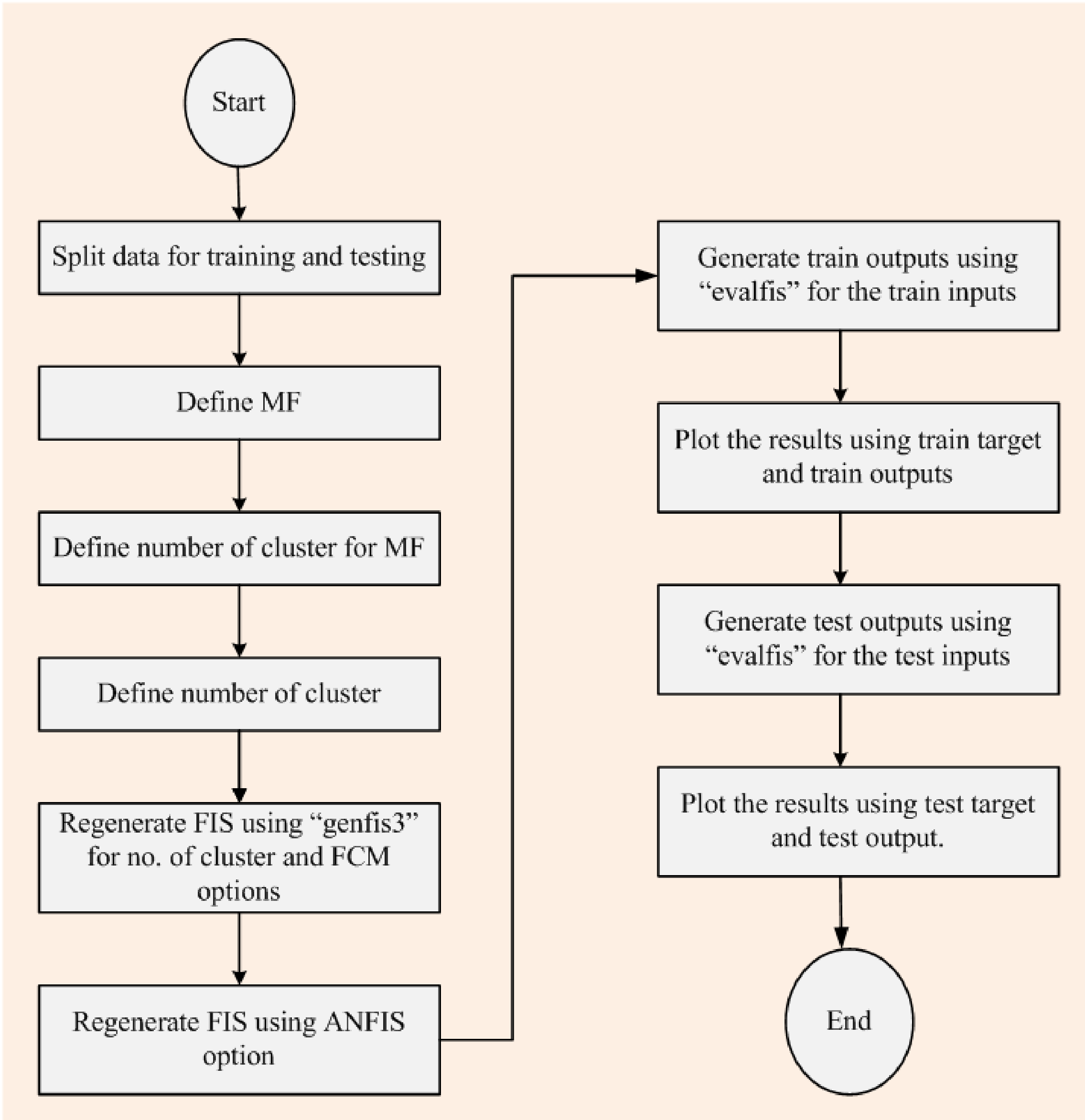


Figure 3.2: Flow chart of ANFIS model.

3.3.1.1 Fuzzy Inference System (FIS)

Fuzzy Inference System (FIS) has become very popular since it was first proposed to describe systems (Chang, 2006). This is a very successful model based on fuzzy logic, used to control processes like automatic trains, nuclear reactors, chemical reactors, as well as different purposes related to engineering, business, etc. For the modeling of reservoir performance, and to resolve the problem regarding data uncertainty or inexactness, fuzzy logic is a highly recommended system (Chang, 2006). statements (Goyal et al., 2014). It also helps to provide a substructure to map input zone to output zone. For example, if a component number belongs to one fuzzy set, the range of the membership function would be 0 to 1, as identified by deleting the sharp boundary separating member from the non-members set (Goyal et al., 2014). This method provides a simple technique to eliminate any ambiguous, unclear or misleading data. Usually the Fuzzy Inference System (FIS) is built based on the consideration that the cluster system generates a cluster center during data collection period. This cluster center represents the behavior of this system. Although the cluster center is responsible for controlling behavior, the fuzzy logic toolbox (MATLAB) is the library function which controls the creation, editing and execution of the FIS system. This toolbox can be used to develop modeling for evaporation systems.

There is a limitation with FIS model functions. The limitation is that no methodical way has been found for the design of a fuzzy controller (Chang et al., 2006). The ANN system is able to arrange input and output in pairs, manage the structure accordingly make it ready to calibrate. Therefore, an Adaptive Neuro-Fuzzy Inference System (ANFIS) is suggested to self-arrange and convert FIS data for forecasting the water or evaporation information from a reservoir system.

3.3.1.2 Artificial Neural Network

The artificial neural network (ANN) relies on one more hidden layer and is able to perform non-linear mapping between input patterns and target values. Nodes are the key elements of artificial neural networks and are also familiar as processing elements. Nodes are connected to each other by other elements known as weights, and they are distributed in the layers of the network. All processing nodes are arranged into layers, and every layer is completely interconnected to the next layer. There is no interrelation between the nodes of the same layer. The input layer of the model acts as a distribution structure for the meteorological data that are presented to the network. The following processing layer, after the input layer, is called the hidden layer. The final processing layer is called the evaporation layer. This sort of ANN is called multilayer perceptron (MLP).

The feed forward back propagation neural network algorithm uses forward propagation and back propagation to calculate all the variables throughout the training period. The activation style of a meteorological variable is propagated through the network to generate an output target (evaporation rate) for the forward flow. Every meteorological variable is multiplied by the adjoining weights before being fed to the processing element in the output layer. The sigmoid activation function is established as one of the most regularly used transfer functions.

The sigmoid function has the meteorological variables and compresses the output into the range 0–1. When the sigmoid activation function is used for the continuous and differential process for any meteorological variable, x, then the equation can be expressed as follows,

$$f\ (x) = \frac{e^x}{e^x+1} \qquad (3)$$

This function is a graphical representation where the curve does not meet any finite distance. In the forward propagation, the calculation of evaporation rate goes on layer by layer in the foreword direction. The difference between observed and predicted evaporation is calculated. The following reasons support why the ANN model is well accepted as a computational tool:

(a) It can identify the relation between the input and output parameters without direct physical consideration,

(b) It can work well even when the training sets carry noise and/or measurement errors,

(c) It can adapt to situations over time, even in changing environments,

(d) It has other basic information-processing qualities and is ready to use once trained.

3.3.2 Firefly Algorithm (FFA)

To overcome the limitation of the traditional model (ANFIS), ANFIS-FFA is adopted. The mechanism of FFA is based on the nature of the firefly (flashing behavior). ANFIS-FFA method is applied during the training phase to determine the best set of data. This model (FFA) formation depends on three basic principles:

1. Each firefly is able to engage another firefly,
2. The attractiveness between two fireflies are calculated by the light intensity of each firefly,
3. The brightness is correspondingly related to the light released by fireflies (Yaseen et al., 2017).

Thus, the objective function of the FFA model is introduced by the intensity of the light produced by each firefly and the brightness of the firefly. The following equations present the intensity (I) and attractiveness (w(r)) at distance r, respectively (Yaseen et al., 2017).

$$I = I_0 e^{-\gamma r^2} \qquad (4)$$

and

$$w(r) = w_0 e^{-\gamma r^2} \qquad (5)$$

Where, r is the distance between fireflies, I_0 is light intensity and w_0 is attractiveness at r=0 distance, γ is the light absorption coefficient. There are also two more coefficients, which are β and α. β is attraction and α is movement co-efficient. To integrate the ANFIS model with FFA, these three coefficients α, β and γ are required to be adjusted. This adjustment can be done by trial and error process (Yaseen et al., 2017). As fireflies are attracted by each other, the movement of two fireflies can be formulated as follows:

$$\Delta_{xi} = \beta_0 e^{-\gamma r^2}\left(x_j - x_i\right) + \alpha\varepsilon_i \qquad (6)$$

Where, $\alpha\varepsilon_i$ is a randomized term and $\beta_0 e^{-\gamma r^2}(x_j - x_i)$ is recognized as the attraction term. ε_i is the random number vector and mutation coefficient, α varies from 0 to 1. In this study, the value of α, β and γ is taken to be 0.2, 2 and 1 respectively.

3.3.3 Genetic Algorithm (ANFIS-GA)

This fuzzy-genetic algorithm consists of a genetic algorithm and an adaptive fuzzy inference system (ANFIS), (Wang et al., 2017). The AFIS is established based on fuzzy logic and the genetic algorithm (GA) is based on the characters of natural genetics and its selection system. GA includes three major stages: (1) population initialization (2) GA operators (3) evaluation (Wang et al., 2017). This GA system is able to solve large space problems efficiently and optimize complicated functions. This model is highly useful for evapotranspiration calculations. Any hybrid model (hybrid ANFIS) can optimize the MF by using GA. This fuzzy-genetic algorithm has a potential to minimize model errors (Wang et al., 2017). In the classical form of GA, decision variables are hidden as binary strings. Chromosomes are formed by the consecutive combinations of genes, which are related to possible monitoring networks, i.e., candidate solutions. The development begins from a population of random chromosomes and thus, generations form. In each generation, the fitness of the whole population is estimated. Then, based on the fitness, multiple chromosomes are stochastically adopted from the current population and adjusted by utilizing genetic operators such as crossover and mutation to create a new population. The current population is applied in the following iteration of the algorithm (Tamer and Alper, 2018). The main objective is to define a small number of effective monitoring wells in the suggested approach by following this solution sequence.

3.3.4 Particle Swarm Optimization (PSO)

The PSO technique was invented by Kennedy and Eberhart (1995), based on the characteristics of bird and fish swarms in a multi-dimensional area; for example, looking for food and running away from hazards (Kennedy and Eberhart, 1995). Every element in this algorithm is identified as a "particle", and particles create the density (population) and each density is identified as a "swarm". Every particle is considered as a candidate for the answer to the question in this algorithm. The swarm and particle values of this technique depend on the chromosome and density (population) items, which are similar to the Genetic Algorithm (Karahan.H, 2012). PSO is a trial and error solution procedure that explores the characteristics of swarm particles in a multi-dimensional exploration zone. The extent of the exploration zone is similar to the number of the unnamed variables of the explored problem, and the number of individuals in the swarm indicates the density of the swarm. Highest and lowest probable values of the variables are selected according to the limit of the probable values of the variables prior to resolving the complication. The highest and lowest speeds are estimated by utilizing these parameters (Karahan.H, 2012). Mostly, the determination of the optimum number of hidden modes is determined via iteration method by progressively varying the number of modes in the hidden layer. But the computation process is different in the case of large data sets because of the higher expenses of developing a significant number of models. Hence, an improvement in computational efficiency is required to overcome this problem. PSO can be an alternative to optimize the number of hidden nodes. PSO is able to optimize with a large possibility and high meeting (convergence) rate.

$$V_d^{Max} = (x_d^{Max} - x_d^{Min})/2 \quad \text{(7)}$$

$$V_d^{Min} = -V_d^{Max} \quad \text{(8)}$$

In this equation, the values of x_d^{Max} and x_d^{Min} are selected according to the limit of the variables and the starting position and velocities of the individuals that are irregularly calculated based on the following equations:

$$x_{prd}^{k} = x_d^{Min} + r(x_d^{Max} - x_d^{Min}) \quad (9)$$

$$v_{prd}^{k} = V_d^{Max}(2r - 1) \quad (10)$$

Where, p, d, v, x and r denote particle number, exploration direction, particle velocity, position of particle and irregularly created number close to unvaried distribution with the limit [0,1] respectively. The fitness value of every particle is estimated based on objective function identified for the particular issue and the position of the best particle is calculated. Each particle upgrades its own position until the position and velocity values face the stopping condition based on the earlier steps and the position of the finest particle in the entire swarm.

$$v_{prd}^{k+1} = \omega v_{prd}^{k} + c_1 r_1 (x_{prd}^{ind} - x_{prd}^{k}) + c_2 r_2 (x_d^{glo} - x_{prd}^{k}) \quad (11)$$

$$x_{prd}^{k+1} = x_{prd}^{k} + v_{prd}^{k+1} \quad (12)$$

Where, k indicates the number of repetitions needed for the trial and error process. ω, c_1 and c_2 are explore variables, r_1 and r_2 are two irregular numbers with an unvaried distribution with the limit [0,1]. x_{prd}^{ind} is the finest location defined by a particle, while x_d^{glo} is the finest location defined by the entire swarm. Variables c_1 and c_2 are the cognition and the social variables respectively (Kennedy and Eberhart, 1995). ω is known as inertial weight which was absent inside the main shape of the algorithm. Kennedy and Eberhart introduced ω as a coefficient, which is considered to be 1 in the PSO algorithm initially.

$$\omega = \omega_{max} - (\omega_{max} - \omega_{min}) \frac{k}{k_{max}} \quad (13)$$

k_{max} and k are considered the highest number of repetition and the current number of repetitions for the trial and error process respectively. Lack of diversity is a significant problem with this PSO iteration process. To overcome this problem, regeneration is chosen with the utilization of Linear Fitness Scaling (LFS).

$$f_{best} - f_{worst} < \varepsilon_{div} \quad \text{(14)}$$

Where; f_{best} and f_{worst} represent the finest and the least objective functions in the entire swarm and ε_{div} is the expression which is needed for diversity. The following equation presents the objective function.

$$MinF(x) = \frac{1}{N}\sum_{i=1}^{N}(I_i^{obs} - I_i^{est})^2 \quad \text{(15)}$$

Where, I_i^{obs} is observed and I_i^{est} is estimated evaporation intensity and N is observation number. This optimization process with Particle Swarm technique extends up to a required concluding situation. In this analysis, the aim of the PSO algorithm is to minimize the objective function. The levels of computation of this process using PSO are given below (Karahan.H, 2012):

Level 1. Starting with the search variables:

- "Niter" represents iteration number;
- "Npt" represents particle number;
- "Nd" denotes searched dimension number;
- x^{Min} and x^{Max} presents vectors of length Nd including searching limits;
- $\boldsymbol{c_1}$, $\boldsymbol{c_2}$, ω_{Min}, ω_{Max} are searching parameters of this technique;
- ε_{div} is diversity tolerance value;
- set k=0 which is known as iteration counter.

Level 2. Estimation of the particle speed (maximum and minimum) in all direction d, based on equations (7) and (8).

Level 3. Calculation of initial particle positions and velocities based on equations (9) and (10).

Level 4. Estimation of objective function based on equation (15).

Level 5. Upgrade this vector " x^{glo}" with dimension Nd which carries the finest location established by the entire swarm.

Level 6. Estimation of inertial weight value with the utilization of equation (13).

Level 7. Modify the speed of the particle for p=1…Npt; d=1…Nd with the utilization of equation (11).

Level 8. When maximum the permitted value is lower than absolute particle velocity, then:

$$v_{prd}^{k+1} = v_d^{Max} sign(v_{prd}^{k+1})$$

Level 9. Modify the particle positions based on equation (12).

Level 10. The particle is positioned at the unlimited searching range when the particle position is not situated in the searching range.

Level 11. Controlling diversity based on equation (14).

Level 12. Study the ending requirements. When ending requirements are not fulfilled, repeat Step 4.

3.3.5 Model Accuracy Indicator

The performances of all four models are individually evaluated using statistical analysis to monitor accuracy with respect to the evaporation forecasting data. The accuracy indicators for ANFIS, FFA, PSO and GA models are calculated in terms of coefficient of determination (R^2), Nash-Sutcliffe coefficient (NSE), root mean square error (RMSE), Mean Absolute Error (MAE), Variance Account For (VAF), Absolute Relative Error (MARE), Scatter Index (SI), Bias, Mean Absolute Error (MAE), Root Mean Square Relative Error (RMSRE). The root mean squared error (RMSE) represents a good measure of the goodness of fit at high parameter values, while the relative error (MARE) provides a more balanced idea of the goodness of fit at moderate and low values. These two indices have the same scale and units as the observational data. The R^2 coefficient measures the correlation of the predicted values with the observational data, whereby the closer the coefficient is to one, the greater the correlation. The value of this coefficient does not interfere with the data unit considered. The SI index is the relative form of RMSE. The performance factor of the model expressed as the Nash-Sutcliffe (Nash and Sutcliffe, 1970) error criterion (E_{NSC}) was used to evaluate the predictive power of the model. A value of unity for the E_{NSC} indicates optimum conformity between predicted and observed data. In this work, both R^2 and E_{NSC} are expressed in percentages. The closer their magnitude to 100, the better the performance of the model. All of them can be calculated from designed formulations, which are presented in the appendices section.

3.4 Results and Discussion

To assess efficiency, all models are separately calibrated with a total of 86 data points for an eight-year period of 2010-2017 at each selected station within the United States of America, with a one-month lead time. Data are collected from the government database of Arizona state in the US. Study area is humid and has an agricultural neighborhood. Two combinations of data set were studied to check the results and verify if they are similar in pattern or not. The data set is initially divided into two parts: training portion and test portion. Two-third, of the total data set has been selected as a training, and one third of the total data set has been considered as a testing data set. First two third of data set was taken for training and rest one third was for testing for first combination. On the other hand, middle one third was considered as the validation and the rest two third of total data was considered as calibration for second combination. As different combination of data set was tried to train to verify accuracy, and the best trained set of data is considered.

Observed data has been run through all four models (ANFIS, ANFIS-FFA, ANFIS-PSO, ANFIS-GA) using the same climate variables, and the predicted results are different for different models. Programming language "MATLAB" was used for coding and it shows the results in graphical form for training data and test data for better comparison.

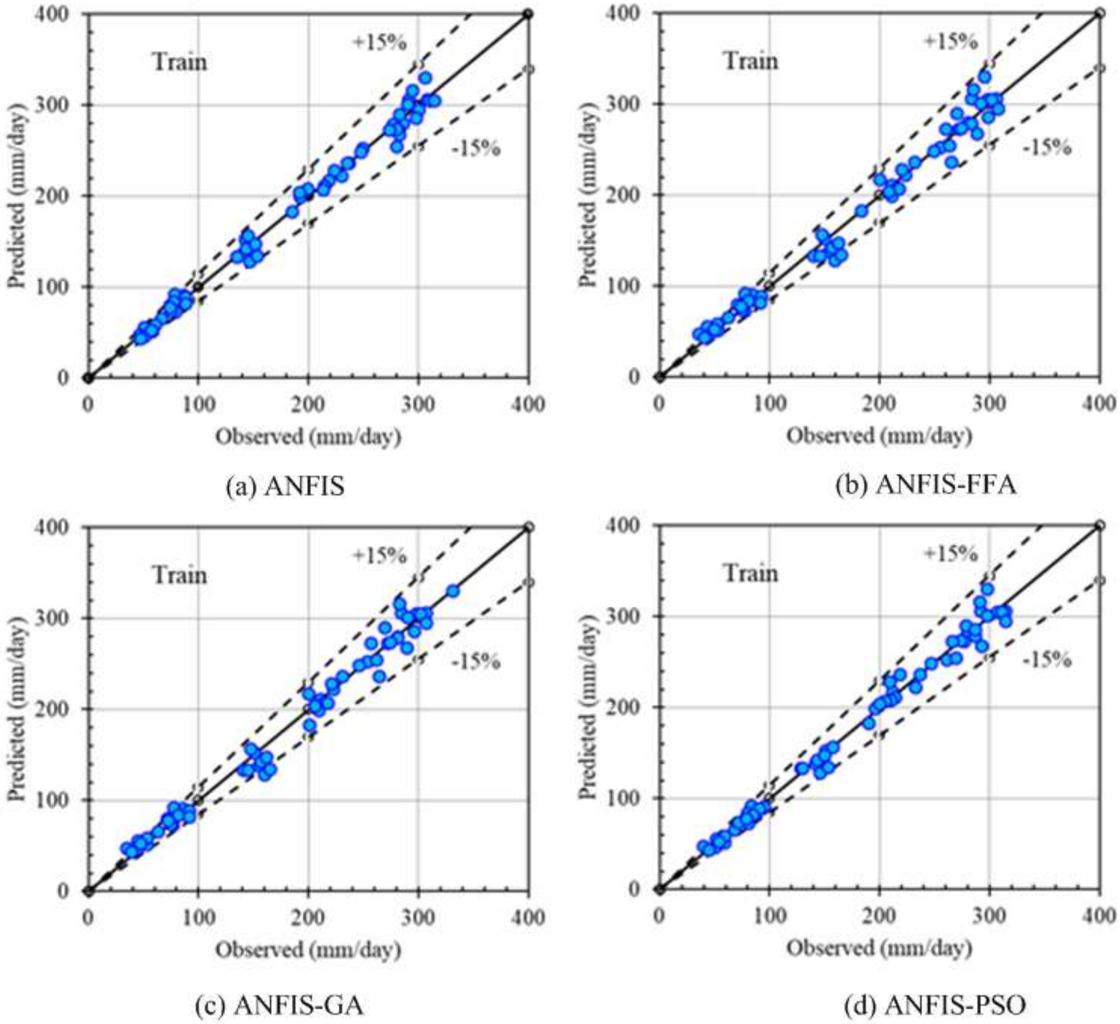
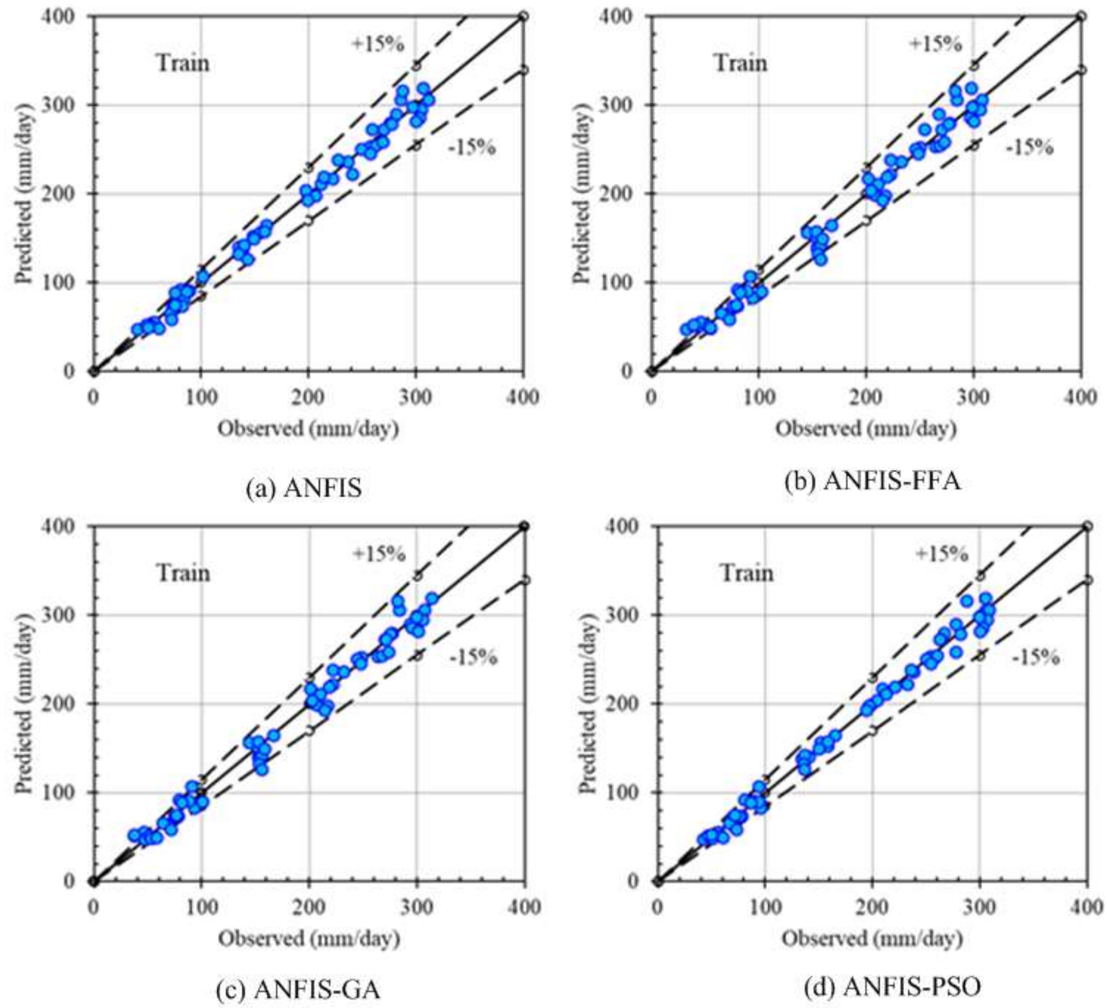


Figure 3.3: Comparison of Target and output sample index of training data for (a) ANFIS, (b) ANFIS-FFA, (c) ANFIS-GA and (d) ANFIS-PSO respectively (first combination of data set).



(a) ANFIS

(b) ANFIS-FFA

(c) ANFIS-GA

(d) ANFIS-PSO

Figure 3.4: Comparison of target and output sample indexes of training data for (a) ANFIS, (b) ANFIS-FFA, (c) ANFIS-GA and (d) ANFIS-PSO respectively (second combination of data set).

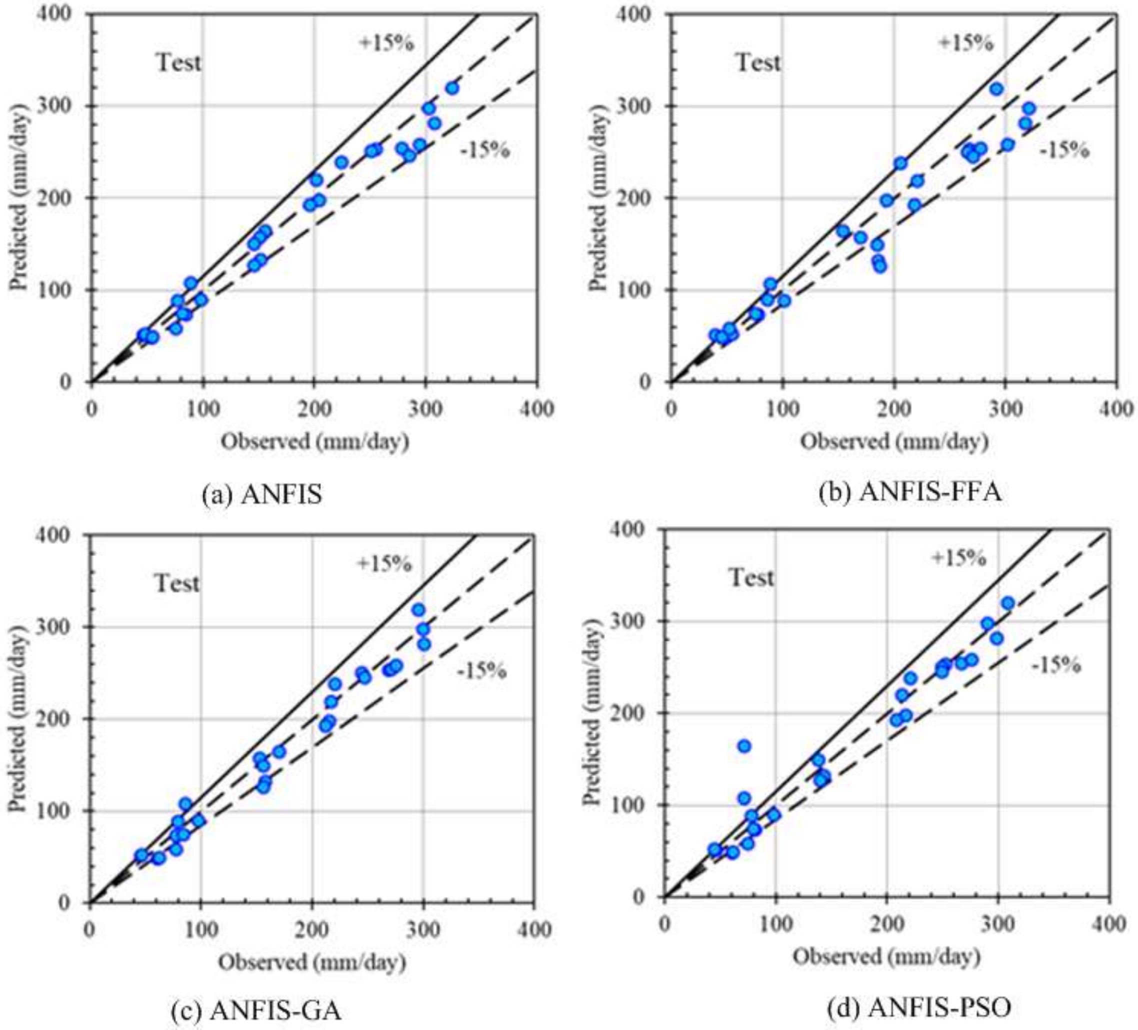


Figure 3.5: Comparison of target and output sample index of test data for (a) ANFIS, (b) ANFIS-FFA, (c) ANFIS-GA and (d) ANFIS-PSO respectively (first combination of data set).

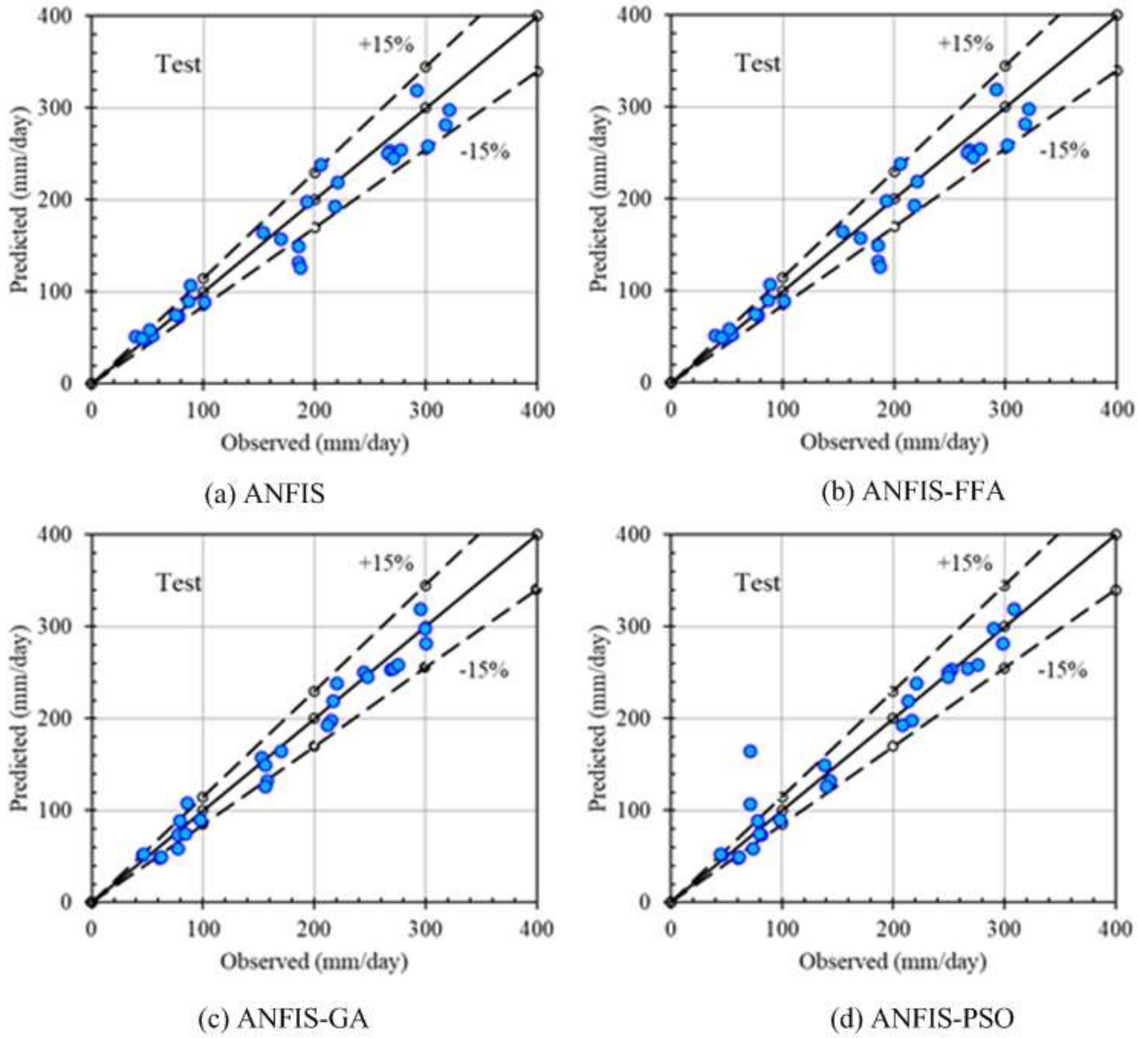


Figure 3.6: Comparison of target and output sample indexes of test data for (a) ANFIS, (b) ANFIS-FFA, (c) ANFIS-GA and (d) ANFIS-PSO respectively (for second combination of data set).

To verify the overall performance of the observed models, the observed and predicted evaporation values were plotted together for both combinations. Figure 3.3, 3.4, 3.5 and 3.6 show the pattern of observed data and predicted data for all four models. Figure 3.3 and 3.4 show the training data pattern for both combinations. These figures show the comparison of target and output sample index of trained data for (a) ANFIS, (b) FFA, (c) GA and (d) PSO models. Similarly, figure 3.5

and 3.6 show test data pattern of all models and those presents the comparison of target and output sample index of test data for (a) ANFIS, (b) FFA, (c) GA and (d) PSO respectively. According to the graphs, both training and test data set lies between -15% to +15% of perfect line. Graphical presentation also demonstrates that the data set are trained well. After analyzing all the graphs, it is clear that all the models are suitable for the evaporation estimation, and the models were trained properly. The pattern for Figure 3.5(a) ANFIS (first combination) and 3.6 (a) ANFIS (second combination) were the best fit and the pattern for Figure 3.5 (b) ANFIS-FFA (first combination) and 3.6 (b) ANFIS-FFA (second combination) show the worst model among the four models. Figures for ANFIS-PSO and ANFIS-GA for both combinations were very similar to each other. A few accuracy tests were performed in order to obtain a better understanding for both training and test data set. Some statistical indices tests have been performed and summarized in Tables 3.2, 3.3, 3.4, 3.5 and 3.6.

Table 3.2: Summary of model accuracy indicator test for training data set, (for the first combination), which was calculated in Excel; AN (ANFIS), FF(FFA), PS(PSO).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
AN	0.99	99.04	8.928	0.050	-0.0008	0.0439	0.0008	0.0009	0.0008	0.99
FF	0.97	94.08	24.38	0.140	8.976	0.1100	0.0792	0.0181	8.976	0.92
GA	0.98	97.50	14.38	0.084	4.569	0.0952	0.0238	0.0272	4.569	0.97
PS	0.99	98.85	9.73	0.054	-0.167	0.040	0.0009	-0.001	-1.687	0.98

Table 3.3: Summary of model accuracy indicator test for test data set (for the first combination), which was calculated in Excel; AN (ANFIS), FF(FFA), PS(PSO).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
AN	0.98	97.04	15.547	0.094	-4.561	0.0870	0.018	-0.027	-4.561	0.970
FF	0.97	93.11	24.388	0.148	-8.976	0.118	0.1536	-0.400	-8.976	0.932
GA	0.98	97.51	14.380	0.087	-4.569	0.1013	0.033	-0.421	-4.569	0.972
PS	0.98	97.18	14.596	0.088	1.6835	0.1013	0.014	0.003	1.6835	0.972

Table 3.4: Summary of model accuracy indicator test for training data set, (for the second combination), which was calculated in Excel; AN (ANFIS), FF(FFA), PS(PSO).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
AN	0.99	98.99	8.985	0.051	0.0002	0.0455	0.0003	0.0185	2.733	0.990
FF	0.98	97.93	12.802	0.073	0.0008	0.0819	0.00655	-0.015	0.0008	0.979
GA	0.99	98.32	11.656	0.066	0.403	0.072	0.0073	-0.011	0.403	0.983
PS	0.99	99.11	8.444	0.048	-0.040	0.042	0.0016	-0.001	-0.040	0.991

Table 3.5: Summary of model accuracy indicator test for test data set (for the second combination), which was calculated in Excel; AN (ANFIS), FF(FFA), PS(PSO).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
AN	0.99	98.42	11.943	0.067	3.730	0.0624	0.0068	0.0248	3.730	0.982
FF	0.98	97.45	15.030	0.085	4.248	0.0762	0.0183	0.0142	4.248	0.973
GA	0.98	97.52	14.631	0.083	3.503	0.0726	0.0244	0.0104	3.503	0.974
PS	0.98	97.50	15.079	0.085	4.611	0.081	0.0004	0.0241	4.611	0.972

Table 3.6: Summary of model accuracy indicator test during the testing period provided by 'MATLAB'.

Type of model	Training data				Test data			
	MSE	RMSE	MEAN	STD	MSE	RMSE	MEAN	STD
GA	146.92	12.12	-2.82	11.89	206.79	14.38	-4.57	13.89
ANFIS	58.23	7.63	-8.16	7.69	241.72	15.54	-4.56	15.14
PSO	58.75	7.66	0.11	7.73	213.05	14.59	1.68	14.77
FFA	507.20	22.52	-4.87	22.17	594.80	24.38	-8.97	23.10

Table 3.6 presents the results obtained by using MATLAB. It shows that MSE values for all the test models are very high (MSE for ANFIS 241.72, for FFA 594.80, for GA is 206.79 and for PSO is 213.05) for test data, and very high for training data. To ensure a rigorous comparison of the models, an extended analysis was performed using RMSE, R^2, MAE, MARE, RMSRE, SI, MRE, Bias, NASH and VAF as statistical indices for the estimated values. Table 3.2, 3.3, 3.4 and 3.5 present values of all statistical indices for training and testing data set of all models. According to all statistical indices, specially, R^2 , RMSE, VAF and NASH values, the second combination of data set presents better result than the first combination of data set, which is presented in table 3.3. Results of ANFIS and ANFIS-PSO models are almost identical in both combinations. RMSE is lower for ANFIS and ANFIS-GA. ANFIS-FFA poses worse result among all model in all the cases. Biasness is less for ANFIS model. According to the test results from table 3.3 and 3.5, R^2 for ANFIS, GA and PSO are almost identical, 0.99, where R^2 for FFA is 0.97. This result is similar to the training result. A commonly used correlation measure, i.e., (R^2) in the testing of statistical indices cannot always be accurate, or sometimes it could be misleading when used to compare predicted and observed models (Benzagtha, M. A., 2014). The two most widely used statistical indicators in evaluating the models are the Root Mean Square Error (RMSE) and the Bias Error.

The model performance is inversely proportional to RMSE value; lower RMSE value presents higher accuracy and vice versa. RMSE is minimum for PSO and GA, which are 14.59, 14.63 and 14.38, 15.07 respectively, whereas ANFIS is 15.54 and FFA presents the worst value: 24.38. Negative biasness has been noticed for all the models, where ANFIS and GA possess minimum biasness. Hence, MSE values are higher, relative statistical indices are compared to find better results. MARE and RMSRE results should also be minimal for the best fit model. Again, ANFIS shows the minimum MARE value (0.087), and PSO gives similar result to ANFIS. But, according to the RMSRE results, PSO shows the best result. For more clarity, NASH has been considered as another accuracy indicator and the value should be close to 1 for the best fit. Table presents the highest NASH value for ANFIS (0.97) and GA (0.97) and PSO (0.97). FFA is also close to 1 (0.93). To avoid confusion, VAF is calculated. Here, ANFIS, GA and PSO show higher results (all three results are close to 97.11) and FFA indicates 93.11.

After analyzing all the results, the FFA model is considered as the least acceptable model among the four. ANFIS with GA and PSO models were acceptable showing better fit in some situations. Although showing almost similar results to GA and PSO, ANFIS can be considered more acceptable because of its simplicity.

3.5 Conclusions

The comparison among Adaptive Neuro Fuzzy Inference System, ANFIS with Firefly Algorithm, ANFIS with Genetic Algorithm and ANFIS with PSO models for estimation of evaporation using climatic variables has been illustrated in this study. Sunshine, Relative Humidity, average temperature, maximum temperature, minimum temperature and wind speed have been considered as the climate variables for all models. Two combinations of data set were trained and tested to verify the similarities of the model results. The study illustrated the accuracy of all four models and various statistical measures (RMSE, RMSRE, MBE, VAF, NASH, Biasness, MBE, MARE, SI and R^2) were used to evaluate the performance of the models. According to the results, the second combination poses slightly better result than the first one. Overall, all four models are suitable for estimation of evaporation, but ANFIS and ANFIS with optimizer PSO were better with all accuracy indicator values. ANFIS with FFA took long time to run and presented low accuracy among four models. ANFIS with GA poses slightly better results than ANFIS-FFA. ANFIS-PSO and ANFIS-GA took similar time (15 to 30 minutes) to finish running. After analyzing all the statistical indices and comparing the observed and predicted data sets, ANFIS and ANFIS-PSO were found to be better among the four models. Relative and absolute accuracy tests have been performed to find the best model in this study. According to RMSE, MBE, VAF, NASH, Biasness, MBE, MARE, SI and R^2 value, ANFIS and ANFIS-PSO were almost identical. ANFIS is recommended due to its simple formulation and easy model development compared to the ANFIS-PSO model. The computational time of ANFIS model was less in comparison to the other models, those are optimized. The objective of the adoption of different optimizer techniques was to verify the accuracy of the outcome prediction by ANFIS model. Since the prediction was almost identical in all cases, the ANFIS model was chosen due to its simplicity. Therefore, the evaporation could be calculated from easily available data using the ANFIS model. Also, this model can be applied as a module for calculating evaporation data in hydrological modeling studies.

References

Benzagtha, M. A., "Estimation of Evaporation from Reservoir in Semi-arid Environment Using Artificial Neural Network and Climate Based Model", *British Journal of Applied Science and Technology,* 4(24), 3501-3518, (2014).

Chang, F.J., Chang, Y.T., "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir", *Advance in Water Resources* 29, 1-10 (2006).

Ferreira, C., "Gene Expression Programming; a new Adaptive Algorithm for Solving Problems", *Complex System* 13 (2), 87-129, (2001).

Friedman, J. H., "Multivariate Adaptive Regression splines", Ann stat, 1-67, (1991).

Goyal, M. K., Bharti, B., Quilty, J., Adamowsky, J., Pandey, A., "Modeling of Daily Pan Evaporation in Subtropical Climates Using ANN, LS-SVR, Fuzzy Logic, and ANFIS", *Expert Systems with Applications* 41, 5267-5276, (2014).

Ozgur Kişi, Ana Pour Ali Baba, and Jalal Shiri, "Generalized Neuro fuzzy Models for Estimating Daily Pan Evaporation Values from Weather Data", *Journal of Irrigation and Drainage Engineering* 138(4), 349-362, (2012).

Kisi, O., "Evolutionary Neural Networks for Monthly Pan Evaporation Modelling", *Journal of Hydrology*, (498), 36-45, (2013).

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., "Comparison of two Different Adaptive Neuro-Fuzzy Inference System in modelling Daily Reference Evapotranspiration", *Water Resource Management*, (28), 2655-2675, (2014).

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Niazi, F., "Long Term Monthly Evaporation Modeling by Sevaral Data-Driven Methods Without Climate Data", *Computers and Electronics in Agriculture* 115, 66-77, (2015).

Lu, J., Sun, G., McNulty, S. G., Amatya. D. M., "A Comparison of Six Potential Evapotranspiration Methods for Regional Use in the South-Eastern United States", *Journal of the American Water Resources Association* 41(3), 621-633, (2005).

Moghaddamnia, A., Gousheh, M. G., Pirli, J., Amin, S., Han, D., "Evaporation Estimation Using Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System Techniques", *Advances in Water Resources* 32, 88-97, (2009).

Shirsath, P. B., Singh, A. K., "A Comparative Study of Daily Pan Evaporation Estimation Using ANN, Regression and Climate Based Models", *Water Resource Management* 24, 1571-1581, (2010).

Sanikhani, H., Kisi, O., Nikpour, M. R., and Dinpashoh, Y., "Estimation of Daily Pan Evaporation Using Two Different Adaptive Neuro-Fuzzy Computing Techniques", *Water Resource Management* 26, 4347–4365 (2012).

Stephens, J.C., Stewart, E.H., "A comparison of procedures for computing evaporation and evapotranspiration", International Association of Scientific Hydrology. International Union of Geodynamics and Geophysics, Berkeley, CA, pp. 123- 133 (1963).

Wang, L., Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Li, H., "Pan Evaporation Modelling Six Different Heuristic Computing Methods in Different Climates in China", *Journal of Hydrology* 544, 407-427, (2017).

Yaseen, Z.M., Ghareb, M.I., Ebtehaj, I., "Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA", *Water Resource Management* 32, 105-122 (2018).

Yager R. R., Filev D.P., "Approximate clustering via the mountain method", IEEE Transaction on Systems, Man and Cybernetics 24(8), 1279–1284 (1994).

Zaher Mundher Yaseen, Isa Ebtehaj, Hossein Bonakdari, Ravinesh C. Deo, Ali Danandeh Mehr, Wan Hanna Melini Wan Mohtar, Lamine Diop, Ahmed El-shafie, Vijay P. Singh, "Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model", *Journal of Hydrology* 554, 263-276 (2017).

M. Tamer Ayvaz, Alper Elci, "Identification of the optimum groundwater quality monitoring network using a genetic algorithm-based optimization approach", *Journal of Hydrology* 563, 1078-1091 (2018).

Kennedy, J. and Eberhart, R. C., "Particle swarm optimization." Proceedings of IEEE international Conference on Neural Networks, Perth, Australia, 1942-1948 (1995).

Halil Karahan, "Determining Rainfall-Intensity-Duration-Frequency Relationship Using Particle Swarm Optimization", *Journal of Civil Engineering* 16(4), 667-675 (2012).

Emrah Dogana, Mahnaz Gumrukcuoglu, Mehmet Sandalci, Mucahit Opan, "Modelling of evaporation from the reservoir of Yuvacik dam using adaptive neuro-fuzzy inference systems", *Engineering Applications of Artificial Intelligence* 23, 961-967 (2010).

Mohammad Ali Ghorbani, Reza Kazempour, Kwok-Wing Chaub, Shahaboddin Shamshirband and Pezhman Taherei Ghazvinei, "Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran", Engineering applications of computational fluid mechanics 12(1), 724-737 (2018).

Chapter 4

An integrated approach base on Multigene Genetics Programming for Prediction of Evaporation in Arid Climate.

4.1. Introduction

Evaporation is a significant element of the hydrologic cycle. It is important because it influences the volume of river basins, the volume of reservoirs, the wasteful utilization of water by crops and the development of underground supplies. Estimation of evaporation is important in many countries of the world where the availability of natural water is limited. Evaporation estimation is necessary in the planning and management of irrigation practices and can play an important role in water budgets for lakes or reservoirs. Many methods have been employed by researchers in order to calculate evaporation, such as, empirical, semi-empirical and climate-based models. Due to the limited availability of data, a better modeling approach such as Artificial Intelligence is required (Dogana et al., 2010). Artificial intelligence models are becoming very popular for predicting data instead of traditional models. The ANFIS model is one of them, which is also a data-driven model that can be used for various measurements, such as rainfall, streamflow, evaporation, water quality and many others. The ANFIS is a very useful model based on fuzzy logic. Several applications of this model can be found to control automatic trains, nuclear reactors, chemical reactors, as well as different purposes related to engineering, business, etc. For example, Chang, et al. (2006) applied the ANFIS model to predict reservoir water level. Identifying the relation between the input and output parameters without direct physical consideration is a characteristic of this model. Again, Kisi et al., (2014) showed the application of the ANFIS model in evaporation estimation on a daily basis. The ANFIS model was successfully applied in this study. For the modeling of reservoir performance, and to resolve the problem regarding data uncertainty or inexactness, fuzzy logic is a highly recommended system. It can work well during training sets carry noise and/or measurement errors, and can also adapt to situations over time, even in changing environments. Information-processing quality is another characteristic of this model. This can also be described as a “feed forward neural network with back propagation training algorithm”, which is used for developing the ANN modeling approach because it is a commonly used and reliable approach in hydrological modeling. ANFIS can be employed for modeling numerous processes, such as motor fault detection and diagnosis, power system dynamic load, wind speed and forecasting systems, demonstrating its ability to create and extending and identifying the best fit data set or model. ANFIS allows the difficult conversion of human intelligence to fuzzy systems, and the extraction of fuzzy rules to numerical data (Chang. et al., 2006).

Single gene genetic programme and multiple gene genetic programme (MGGP) are concepts inspired by a heuristic algorithm and have been used in various areas of researches. For example, Raj and Rajendran (2009) proposed a simple heuristic algorithm to resolve a single-stage Fixed-charge transportation problem (FCTP) and compared the result with the traditional famous method by using benchmark problem instances. Yan, et al. (in 2019) used this optimizer (Multigene Genetic-Programming-Based Models) to predict initial dilution of vertical buoyant jets. This study showed a comparison of the results of Single gene genetic programme and multiple gene genetic programme and demonstrated the superiority of MGGP.

Another optimization system that has become popular in different areas of researches, is Ant Colony Optimization (ACO). Many researchers work on this model for different purposes. For example, Silva, et al. (in 2009) applied ACO first to solve different operational activities for supply chain management. A meta-heuristic algorithm, which can also be called ACO model, has been used to solve problems. Mausavi, et al. (2017) also utilized this Ant Colony Optimizer in a statistics-based study on divorce rate reduction. This optimization system follows the way an ant finds a path to reach its food or goal by avoiding all obstacles. Later on, Hong, et al. used ACO on a two-stage supply chain problem and achieved a satisfactory result in 2018.

The objectives of this study are to evaluate the performance of the ANFIS with Ant Colony Optimization (ACO) and ANFIS with Multigene Genetic-Programming for evaporation estimation from the available atmospheric data. Then, all the results were compared using statistical analysis and the best model was determined. This study explored the ability of the MGGP model to improve the accuracy of daily evaporation estimation for arid environments in the United States.

4.2. Data Description

Arizona is considered as the sixth biggest state of USA which is located near the state of California. The area of this state is around 113,000 square miles and weather condition is quite caustic with tropical summers and muggy winters.

Figure 4.1: Study area, Phonix, Arizona (Source: maps.ie/coordinates).

Phoenix is the capital of Arizona territory which is located in the Northeastern part of Sonoran Desert; therefore, it has a hot desert climate condition. This city has an agricultural neighborhood which is closed to the confluence of the Salt and Gila river. Figure 1 shows the map of study area, which is 355.7m higher from sea level, with 33.4258 latitude and -111.9217 longitude. After collecting data, two third of them were used for the training purpose and one third of the data was used for testing. Table 4.1 summarizes the statistical indices of test, training and all data used in this study.

Table 4.1: Statistical indices of evaporation data with three subsections; total, train, test.

Statistics	N	Min	1st Q	X50	3^{rd} Q	Max	Avg	SD	CV (%)	Skewness	Kurtosis
All	85	44	82.5	158	254.5	331	172.30	89.48	51.93	0.0657	-1.446
Train	59	44	83	183	273	331	178.28	91.79	51.48	0.0145	-1.483
Test	26	49	74.75	154	247.25	298	158.73	82.40	51.91	0.1169	-1.500

N is number of data, Min and Max is Minimum and Maximum of data, 1st Q and 3dr Q is first and third Quarters, Avg is average, SD is Standard Deviation, CV is Coefficient of Variation. Standard deviation shows the distribution nature of data set. For example, standard deviation of test data set is 82.40 and average value of data is 158.73. That means, most of the test data lies between 78.33 (158.73-82.40=78.33) to 241.13 (158.73+82.40=241.13). Coefficient of Variation shows the precision of data set in this table. That parameter presents the ratio of standard deviation and mean value in percentages. Two more statistical indices; Skewness and Kurtosis were calculated in order to show the nature of data set.

4.3. Methodology

This section presents a short description of ANFIS, ANFIS-ACO and MGGP models. Data were collected with the consideration of six climatic variables for all three models. The variables are, sunshine, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed. After collecting data, 70% of the total data was used for the training purpose and 30% of the data was used for testing.

4.3.1 Adaptive Neuro Fuzzy Inference System

To understand ANFIS type models, it is important to know basic ANFIS model and its structure first. The ANFIS model is a composition of Fuzzy Inference System (FIS) and Artificial Neural Network (ANN). This model is adopted mainly due to its good capacity of extraction of data from input to fuzzy values in a range of 0 to 1.

ANFIS is a multilayer feed-forward network, which is capable of arranging and converting data. Node and rule are two key components in this system, where rule determines the relationship between input and output and membership Functions (MF). ANFIS applied “IF-THEN” rules to the input data. Equation (1) and (2) show the rules for an ANFIS model for two inputs (x and y) and one output f.

Rule 1: IF x is P_1 and y is Q_1, then $f_1 = p_1 x + q_1 y + r1$ ……………………..………….(1)

and

Rule 2: IF x is P_2 and y is Q_2, then $f_2 = p_2 x + q_2 y + r2$ ……………………..………..(2)

To understand the ANFIS model better it can be described layer by layer.

First layer: the input nodes and each node generate membership grades by utilizing the membership function, which is one of the fuzzy sets (Chang et al., 2006).

Second layer: Rule and node are the basic of this layer. In this layer, the AND operator is implemented to connect output and therefore, the output of layer 1 turns into the input for layer 2. Basically, the result of the previous layer becomes the input of this layer.

Third layer: An average node is considered, which aims to calculate the proportion of the single rule's strength to the total of all rule firing strengths (Chang et al., 2006). Firing strength is the degree to which the fuzzy rule from the previous segment is satisfied, and it forms the output function for the rule.

Forth layer: Consequent node is the basic of this layer; which can calculate the working power of each rule with consideration of the total output.

Fifth layer: This layer consists of output nodes. The output nodes are generated by adding up all the incoming signals.

4.3.1.1 Fuzzy Inference System (FIS)

For better understanding of ANFIS model, it is necessary to know the structure and working principle of Fuzzy Inference System (FIS). Fuzzy Inference System (FIS) has become very popular since it was first proposed to describe systems (Chang, 2006). Membership function is an important term that indicates the degree to which a component belongs to a set of data (Goyal et al., 2014). It is a binary logic, which can convert the idea of partial truth-truth values to "completely true" or "completely false" statements (Goyal et al., 2014). It provides a substructure that creates a path from the input zone to output zone. In the FIS system, a cluster system generates a cluster center during the data collection period, which indicates the behaviour of this system. Cluster center has an influence on controlling behaviour. Programming language "MATLAB" has been chosen as the fuzzy logic toolbox which controls the creation, editing and execution of the FIS system. This toolbox is very effective for such modeling with respect to predicting data sets. The limitation of this system is that no methodical way has been found for the design of a fuzzy controller (Chang et al., 2006). The ANN system is able to generates input-output pairs, arrange the structure and convert it in a collaborative manner. Therefore, an Adaptive Neuro-Fuzzy Inference System (ANFIS) is recommended for its ability to self-order and transform FIS data for predictive purpose.

4.3.1.2 Artificial Neural Network

To study Artificial Neural Network (ANN) is very important to understand the development and working process of any ANFIS type model. ANN model has the potential to perform non-linear mapping between input and output. This model depends on a hidden layer where, nodes are the main elements. Nodes are interconnected with weights and every layer is interconnected to the next layer, but nodes are not interrelated in the same layer. The input layer acts as a distribution structure, the second layer is known as the hidden layer, and the final layer is called the evaporation layer. ANN can be called a multilayer perceptron (MLP). The following figure can provide a clear idea of a three-layered ANN structure, which was employed in this analysis.

Each node accumulates process weighted input from the previous layer and sends its output to another node in the very next layer via the connections. The earliest assigned weight values are gradually corrected, and a comparison is done between forecasted outputs and known outputs in each iteration while training the data. Correct weight adjustments are important to reduce errors and back propagation can solve this problem. Weighted summation of inputs to a node is converted to an output according to a transfer function. This transfer function is basically a sigmoid function and the model can be described as a “feed forward neural network with back propagation (BPNN). When the sigmoid activation function is used for the continuous and differential process for any meteorological variable, x, then the equation can be expressed as follows,

$$f\ (x) = \frac{e^x}{e^x+1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

This function is a graphical representation where the curve does not meet any finite distance.

4.3.2 Ant Colony Optimization (ACO)

Ant colony optimization system is inspired by the nature of ants. This optimization system follows the way an ant finds a path to reach its food or goal by avoiding all obstacles. ACO consists of ant behaviour and a pheromone matrix. Figure 4.2 can provide a clear idea on how this system runs based on the nature of an ant.

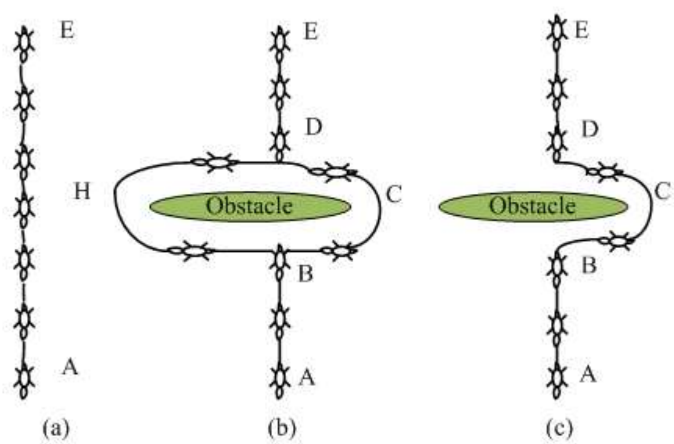


Figure 4.2:(a) Some ants are walking on a path between points A and E (b) Obstacle suddenly appears and the ants must get around it (c) At steady state the ants choose the shorter path. (Source: Colony, et al. 2014).

ACO algorithm builds a pheromone matrix which is an integrated record of optimization steps. This matrix is easily accessible during the optimization process. This approach functions in few steps:

1. Input data: data inserted with the consideration of input variables, sunshine hour, relative humidity, wind speed, maximum temperature, minimum temperature, average temperature.

2. Data initialization: This step starts with the initialization of parameters, such as,

   N is the number of ants.

   IT is the number of iterations.

   α is the pheromone matrix which represents ant intensity from i^{th} partner to j^{th} partner in the following stage.

   β is a parameter which represents the profitability from i^{th} partner to j^{th} partner in the following stage.

   ρ is the evaporation rate.

   Q is a constant value which governs pheromone increment.

3. Solution generation and evaluation:

   An ant creates a path in each stage by following a simulation-oriented process, which is shown step by step below;

   Step 1: In first stage an output data are selected randomly. If the data are not allocated, then go on to step 2 and redo the first step.

   Step 2: Selection of a uneven number from zero to one.

   Step 3: Attain cumulative probability for the specified output data from the likelihood matrix for stage 1.

   Step 4: Distribute selected output data from step 2, to the upstream unit where the cumulative probability meets the uneven number.

   Step 5: Again, create an uneven number from zero to one.

   Step 6: Attain cumulative probability from stage 2, for step 4.

Step 7: Depending on the generated uneven number in the 5th step, find that upstream unit where the cumulative probability value meets the uneven number. If the capacity of the upstream unit is zero, go to the 5th step. Otherwise proceed to step 8.

Step 8: Upgrade all the data. If the demand is satisfied totally, one set of distribution is finished; If not, repetition from the 5th step is required.

Step 9: Terminate when the demand is fulfilled Otherwise, , need repetition from 1st step for the next set of allocations.

4. Probability and visibility function:

The probability function for ant l is predicted output k, selected to be distributed to the next data set j in stage l of iteration t, is shown by the following equation,

$$PM_{kj}^{l}(t) = \begin{Bmatrix} \frac{[\tau_{kj}(t)]^{\alpha}[\eta_{kj}(t)]^{\beta}}{\sum_{j=1}^{d}[\tau_{kj}(t)]^{\alpha}[\eta_{kj}(t)]^{\beta}} \\ 0, \quad otherwise \end{Bmatrix}, \quad \text{if j} \in N_{k}^{l} \quad \ldots\ldots(4)$$

Where, τ_{kj} denotes the pheromone concentration in side (k,j). N_{k}^{l} denotes the feasible neighbourhood of ant l , η_{kj} is the visibility that can be identified by a heuristic rule that includes the search with few important details about the problem, and α and β are the elements to direct the search which defines the necessity of pheromone trail and heuristic details.

5. Pheromone matrix updating:

Pheromone updating consist of two terminologies; evaporation and pheromone deposition. Pheromone evaporation is not difficult to be installed in a fine-grained parallel manner, such as a single thread that can independently lower each entry of the pheromone matrix by using a constant number. But pheromone deposition is difficult because ants might try to deposit pheromone on the same side at the same time. The solution to overcome this problem is to use atomic instructions to block racing during entering the pheromone matrix. The updated pheromone intensity for each iteration can be given by the following equation,

$$\tau_{kj}(t+1) = \rho\tau_{kj}(t) + \Delta\tau^{l} \quad \text{......(5)}$$

$$\tau_{ji}(t+1) = \rho\tau_{ji}(t) + \Delta\tau^{l} \quad \text{......(6)}$$

where ρ denotes the rate of evaporation of the pheromone matrix, t is iteration. The first term denotes evaporation and the second one denotes deposition. The term $\Delta\tau^{l}$ is the accretion in the pheromone on the links where the distribution is completed by ant l.

$$\Delta\tau^{l} = \begin{Bmatrix} \frac{Q}{iteration-best(t)} & If\ allocation\ is\ done\ by\ l \\ 0 & Otherwise \end{Bmatrix} \quad \text{......(7)}$$

Where Q is a constant.

6. Termination and output:

   When the stopping conditions (iterations) are fulfilled, the algorithm is terminated. Otherwise, the number of iterations is increased, and the entire process is redone. At the end, the best data set can be obtained by the completion of a heuristic run.

4.3.3 Multigene Genetic-Programming (MGGP)

Multigene Genetic-Programming (MGGP) is one of the modern alternatives to Genetic Programming (GP), which has an admired evolutionary skill that can be successfully applied to data-driven nonlinear models. This new approach is based on the multigene genetic-programming (MGGP) technique, which is an upgrade of single gene-genetic programming (SGGP). By using training data, MGGP can spontaneously develop a distinct model and it does not need to define the model structure ahead of time. This helps in developing a mathematical model and also ignore errors, such as errors regarding judgments related to the model structure. MGGP has two primary benefits: (1) it is able to create multiple genes and each gene of the MGGP model is a conventional

GP gene. Therefore, the accuracy of MGGP is higher compared with the traditional GP approach. (2) The order of the nonlinear term of a single gene is less because, each gene of MGGP obtains only a few tree depths, and therefore, the MGGP model becomes more concise. Basically, this approach builds an imperial model in this process. The first generation is called parent genes, and more generations are required to obtain the best set of equations, because the fitness of the first generation is always lower. Generations are developed by following three steps: reproduction, crossover and mutation (Yan and Mohammadian, 2019). The second generation consists of child genes which are formed by switching the sub-trees of the first generation. Mutation starts, when the crossover part comes to an end. In this process, the sub-trees are substituted with a new component. To build an entire MGGP model, numerous generations are required. After repeating all the steps, a few equations were found in this study, and the best one was selected. The best form of the equation is shown below,

$$y = (19.5x_1 + 19.5x_3 + 9.77\cos(x_3) + (5.55e^{-17} + (4.88e^{15}x_1 + 4.88e^{15}\cos(x_3) + 4.88e^{15}\cos(x_4)))/\sin(\sin(\log(x_2))) + (1.42e^{-14}(7.01e^{15}x_1 + 7.01e^{15}cosx_3 + 7.01e^{15}\log(x_2)))/x_2 - (5.55e^{-17}(6.74e^{15}x_1 + 6.74e^{15}\cos(x_1 + x_3) + 6.74e^{15}\cos(x_5)))/\sin(\log(x_2)) + (0.0931x_5^2x_6^3)/(x_1x_3x_4(2.0x_1 + x_3)) - 204.0) \quad \text{..............................(8)}$$

Where, $x_{1,}x_{2,}x_3, x_4, x_{5,}x_6$ are the weather variables, wind speed, relative humidity, sunshine, average temperature, maximum temperature and minimum temperature respectively for the final best fit equation.

4.4. Results and Discussion

All models are separately calibrated with a total of 85 data points for an eight-year period of 2010-2017 at each selected station within the United States of America, with a one-month lead time. Data are collected from the government database of the US. Two combinations of data set were studied to check the results and verify if they are similar in pattern or not. The data set is initially divided into two parts: training part and test part. Almost two-third, 59, of the total data set has been selected as a training data set, and one third, 26, of the total data set has been taken as a testing data set. First two third of data set was considered for training and rest one third was considered as testing for the first combination. Middle one third was taken for testing for second combination, where rest of the data set was considered for training. Observed data has been run through both models (ANFIS-ACO, MGGP) using the same climate variables. The ANFIS-ACO model produces the predicted results when the run is finished. “MATLAB” was chosen as a programming language for coding and it presents results in graphical form for both training data and test data. This presentation helps to understand the comparison better, but, the MGGP model is ended up by providing many equations. This model is not a part of ANFIS model but a numerical model.

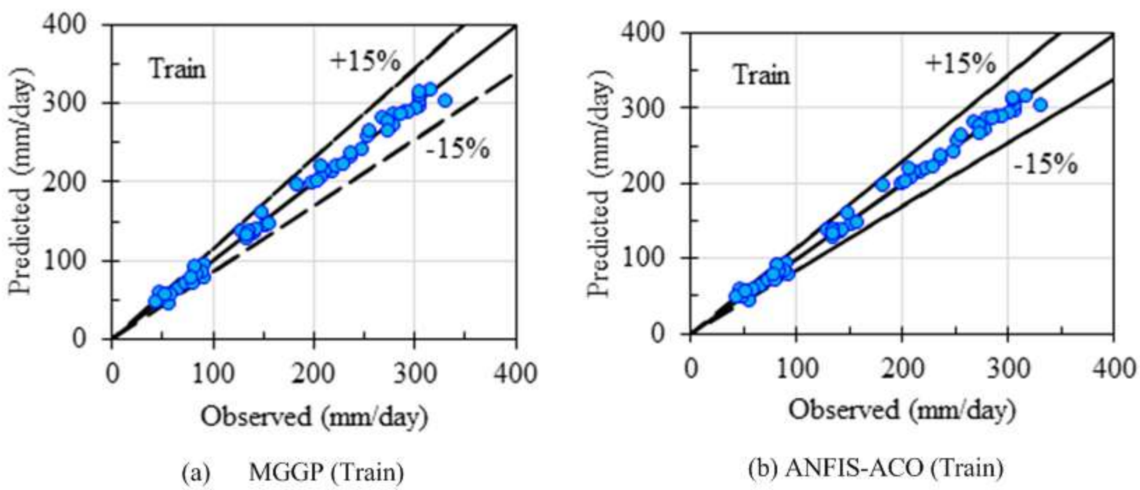
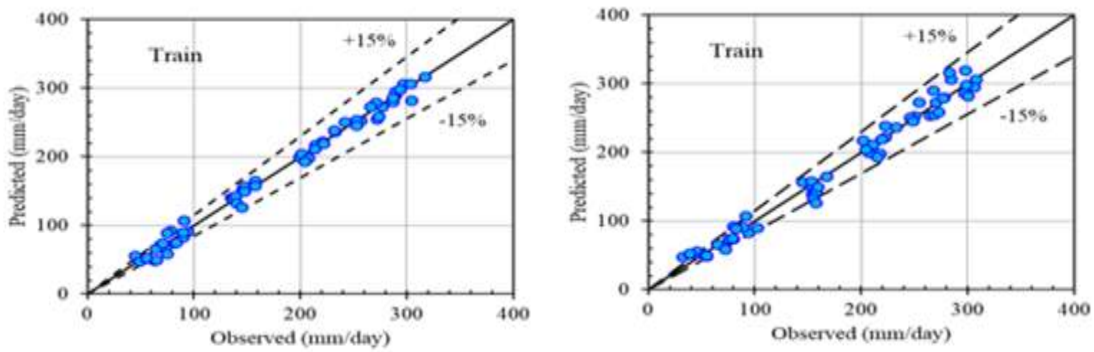


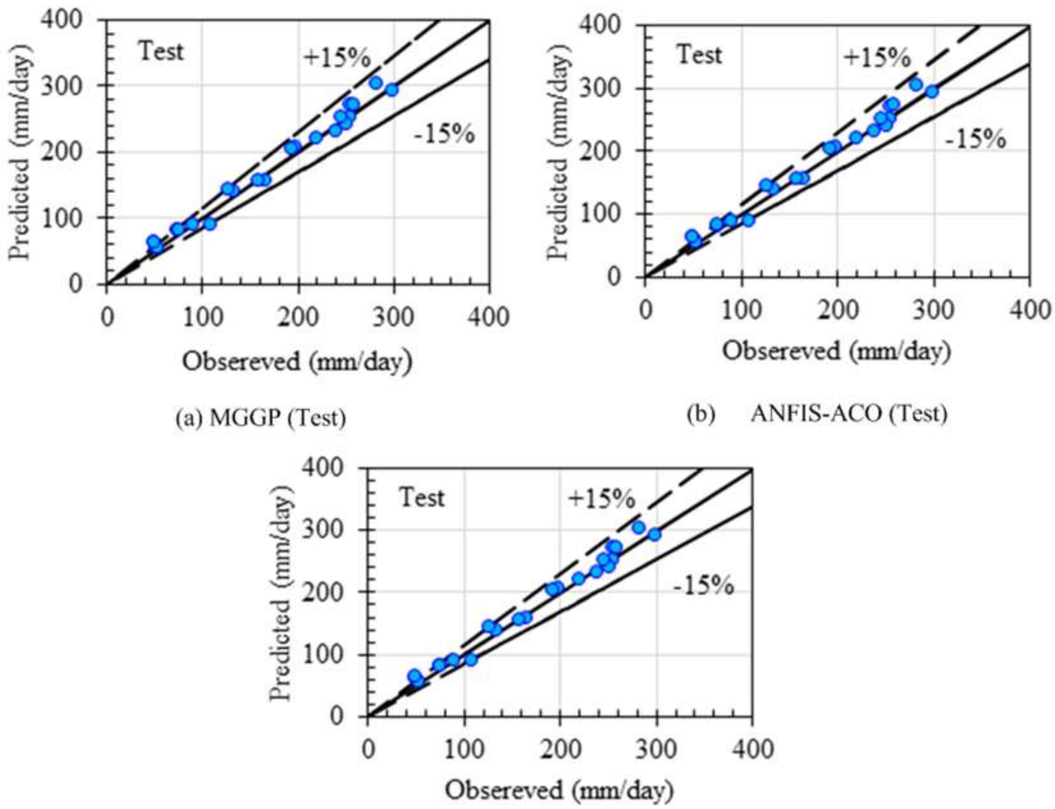
Figure 4.3: Comparison of target and output sample index of training data for (a) MGGP and (b) ANFIS-ACO respectively (for 1st combination data set).



(a) MGGP (Train)

(b) ANFIS-ACO (Train)

Figure 4.4: Comparison of target and output sample index of training data for (a) MGGP and (b) ANFIS-ACO respectively (for 2nd combination data set).



(a) MGGP (Test)

(b) ANFIS-ACO (Test)

(c) ANFIS (Test)

Figure 4.5: Comparison of target and output sample index of test data for (a) MGGP (b) ANFIS-ACO and (c) ANFIS respectively (for 1st combination data set).

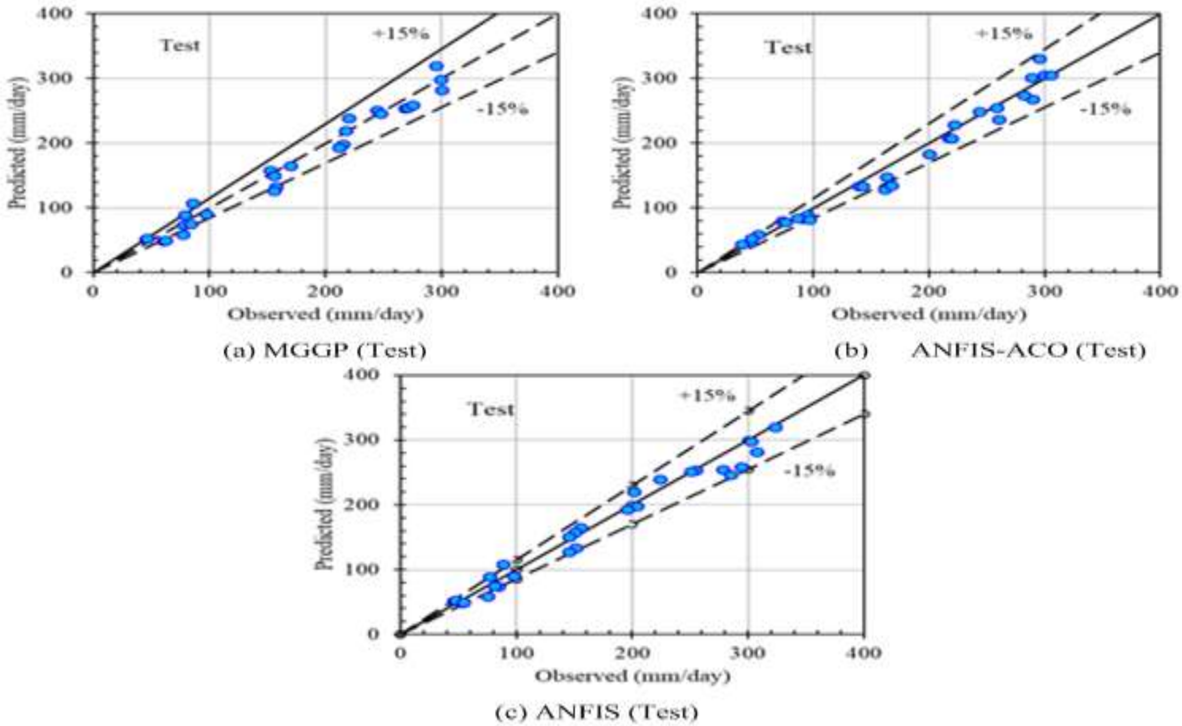


Figure 4.6: Comparison of target and output sample index of test data for (a) MGGP (b) ANFIS-ACO and (c) ANFIS respectively (for 2nd combination data set).

After analyzing all equations, the best one was verified and selected. This equation was applied and solved by using all variable data in excel. Observed and predicted evaporation values were plotted together to check model performance. Observed and predicted evaporation values were plotted together to check model performance. Figure 4.3 and 4.4 present the nature of observed data and predicted data for MGGP and ANFIS-ACO models. The pattern for Figure 4.3(a) MGGP (first combination) and 4.5 (a) MGGP (second combination) were the best fit and the pattern of ANFIS-ACO is very identical to them. Both MGGP model and ANFIS-ACO are very similar to basic ANFIS model. The graphical presentation shows the results lies in between -15% to +15% of the line of coefficient of determination (R^2) for both models. That means, the examined results can be varied from +15% to -15%, which is 0.85 to 1.15, as the value of coefficient of determination is considered 1 in order to get best performance. Results from basic ANFIS model are added to observe the difference of the results of other two models. A few accuracy tests were performed in order to obtain a better understanding for both training and test data set. Some statistical indices tests have been performed and summarized in Tables 4.2, 4.3, 4.4 and 4.5 in order to achieve a better understanding.

Table 4.2: Summary of model accuracy indicator test results for training data set AN (ANFIS), MP (MGGP), and AC (ANFIS-ACO), which were calculated in Excel (combination 1).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
MP	0.996	99.347	7.456	0.041	0.774	0.043	0.0003	0.002	0.774	0.993
AC	0.99	97.908	13.27	0.074	0.00	0.069	0.007	-0.006	0.00	0.978

Table 4.3: Summary of model accuracy indicator test results for training data set of AN (ANFIS), MP (MGGP), and AC (ANFIS-ACO), (combination 2).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
MP	0.985	98.87	9.57	0.048	0.77	0.048	0.0003	0.028	0.87	0.97
AC	0.989	97.72	12.80	0.073	0.00	0.082	0.006	-0.015	0.00	0.99

Table 4.4: Summary of model accuracy indicator test results for test data set AN (ANFIS), MP (MGGP), and AC (ANFIS-ACO), which were calculated in Excel (combination 1).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
AN	0.986	96.62	15.83	0.099	-4.61	0.09	0.018	-0.028	-4.61	0.967
MP	0.992	98.47	10.98	0.07	-4.12	0.014	0.002	-0.046	-4.12	0.982
AC	0.988	97.42	13.68	0.086	-3.55	0.015	0.033	-0.011	-3.55	0.975

Table 4.5: Summary of model accuracy indicator test results for test data set AN (ANFIS), MP (MGGP), and AC (ANFIS-ACO), which were calculated in Excel (combination 2).

	R2	VAF	RMSE	SI	MAE	MARE	RMSRE	MRE	BIAS	NASH
AN	0.99	98.42	11.94	0.067	3.73	0.062	0.006	0.024	3.7	0.98
MP	0.99	99.00	9.13	0.052	1.36	0.05	0.0003	0.013	1.36	0.99
AC	0.987	97.45	15.03	0.085	4.24	0.076	0.018	-0.014	4.24	0.97

Table 4.2 and 4.3 demonstrate the statistical indices test results for training data set for MGGP and ANFIS-ACO models for both data combination. Those were calculated in Microsoft excel. Table 4.3 (second combination) presents slightly better result than table 4.2 (first combination).To ensure a rigorous comparison of the models, an extended analysis was performed using RMSE, R^2, MAE, VAF, Bias, NASH, VAF, MRE, RMSRE and SI as statistical indices for the estimated values. These tabular presentations advocate the perfection of training data set of two models. To verify the results of test data of two models, a simple statistical analysis test was also performed, and a comparison was shown with the basic ANFIS model in table 4.4 and 4.5.

Table 4.4 and 4.5 present values of all statistical indices for all models. R^2 for ANFIS and ANFIS-ACO are almost identical, 0.98, where R^2 for MGGP is 0.99. A commonly used correlation measure, i.e., (R^2) in the testing of statistical indices cannot always be accurate, or sometimes it may mislead when used to compare predicted and observed models (Benzagtha, M. A., 2014). Root Mean Square Error (RMSE) and the Bias Error are the two most widely used statistical indicators in evaluating the models. The minimal RMSE value indicates the better model. A significant different has been noticed on the RMSE results. RMSE is minimal for MGGP which is 9.13, whereas ANFIS is 11.94 and ANFIS-ACO is 15.03 (second combination), therefore, a close observation is needed before concluding. From Table 4.4, negative biasness has been noticed for all the models, where, table 4.5 (second combination) poses all positive biasness with minimal value. Comparison among three models for the second combination shows minimum biasness for MGGP model. Relative statistical indices are compared to verify better results. MARE and RMSRE results should also be minimal for the best fit model. Again, MGGP shows the minimum MARE value (0.014), and ANFIS-ACO is also very close to MGGP (0,015). But a big difference is shown in RMSRE results. According to the RMSRE results, MGGP shows the best result (0.002), but ANFIS shows the worst (0.018) in table 4 (first combination). For more clarity, NASH has been considered as another accuracy indicator and the value should be close to 1 for the best fit. Table 4 presents the highest NASH value for MGGP (0.99), then ANFIS (0.97) and lastly ANFIS (0.98) for second combination of data set. To avoid confusion, VAF is calculated. Here, MGGP shows a higher result (99), ANFIS holds the second position (98.42) and again, ANFIS-ACO shows the worst value (97.45). According to the table, the values from three models are almost similar except RMSE. After analyzing all the results, the three models are acceptable for this estimation, but MGGP and ANFIS-ACO models showed better fit in some situations. Although showing very similar results, MGGP is considered as a little ahead from ANFIS-ACO model. Overall, MGGP model showed the best result in all instances. Therefore, MGGP can be declared more acceptable according to statistical analysis.

5. Conclusions and Final Remarks

The comparison between ANFIS with Ant Colony Optimization and MGGP models for estimation of evaporation by making use of climatic variables has been illustrated in this paper. Sunshine, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed have been considered as the climate variables for all models. To check the test and train results two combination of data set were considered. This study illustrated the accuracy of both models and various statistical measures (RMSE, MRE, VAF, NASH, Biasness, MBE, MARE, SI and R2) were used to evaluate the performance of the models. Results from both models were also compared with the traditional ANFIS model to observe the differences. According to the results, second combination of data set provides slightly better result than first combination of data set. All three models are suitable for estimation of evaporation, but MGGP showed better results based on all accuracy indicator values. Although ANFIS-ACO presented slightly better results in some cases, after analyzing all the statistical indices and comparing the observed and predicted data sets, MGGP was found to be better among the three models. Relative and absolute accuracy tests have been performed to find the best model in this study. RMSE, MBE, VAF, NASH, Biasness, MBE, MARE, SI and R2 of ANFIS, ANFIS-ACO and MGGP were calculated. According to the statistical analysis, MGGP was considered as a best fitted model. It should be noted that, the ANFIS model also took less time to run than any other model using an optimizer, whereas the MGGP model takes more time to find the best equation and to calculate the equations in excel. MGGP is somewhat complicated as it does not provide the results and graphs; rather it gives a set of equations. Despite this, MGGP provides more accurate results at the end compared to ANFIS-ACO. Therefore, this model can be applied as a module for calculating evaporation data in hydrological modeling studies.

References:

Benzagtha, M. A., “Estimation of Evaporation from Reservoir in Semi-arid Environment Using Artificial Neural Network and Climate Based Model”, *British Journal of Applied Science and Technology* 4(24), 3501-3518, (2014).

Xiaohui, Y. and Abdolmajid, M. “Multigene Genetic-Programming-Based Models for Initial Dilution of Laterally Confined Vertical Buoyant Jets”, *Marine Science and Engineering* 7(246), (2019).

Moghaddamnia, A., Gousheh, M. G., Pirli, J., Amin, S., Han, D., “Evaporation Estimation Using Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System Techniques”, *Advances in Water Resources* (32), 88-97, (2009).

Kennedy, J. and Eberhart, R. C., “Particle swarm optimization.” Proceedings of IEEE international Conference on Neural Networks, Perth, Australia, 1942-1948, (1995).

Kisi, O., “Evolutionary Neural Networks for Monthly Pan Evaporation Modelling”, *Journal of Hydrology* 498, 36-45, (2013).

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., “Comparison of two Different Adaptive Neuro-Fuzzy Inference System in modelling Daily Reference Evapotranspiration”, *Water Resource Management* 28, 2655-2675, (2014).

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Niazi, F., “Long Term Monthly Evaporation Modeling by Sevaral Data-Driven Methods Without Climate Data”, *Computers and Electronics in Agriculture* 115, 66-77, (2015).

Goyal, M. K., Bharti, B., Quilty, J., Adamowsky, J., Pandey, A., “Modeling of Daily Pan Evaporation in Subtropical Climates Using ANN, LS-SVR, Fuzzy Logic, and ANFIS”, *Expert Systems with Applications* 41, 5267-5276, (2014).

Chang, F.J., Chang, Y.T., “Adaptive neuro-fuzzy inference system for prediction of water level in reservoir”, *Advance in Water Resources* 29, 1-10 (2006).

Silva, C. A., Sousa, J.M.C., Runkler, T.A., Sa Da Costa, J.M.C., “Distributed Supply Chain Management Using Ant Colony Optimization”, *European Journal of Operational Research* 199, 349-358, (2009).

Hong, J., Diabat, A., Panicker, V. V., Rajagopalan, S., “A two-stage Supply Chain Problem with Fixed costs: An Ant Colony Optimization Approach”, *International Journal of Production Economics* 204, 214-226, (2018).

Yan, X., and Mohammadian, M., “Multigene Genetic-Programming-Based Models for Initial Dilution of Laterally Confined Vertical Buoyant Jets”, *Journal of Marine Science and Engineering* 7, 1-15 (2019).

Mousavi, S. M. H., MiriNezhad, S. Y., Layshenko, V., “An Evolutionary-Based Adaptive Neuro-Fuzzy Expert System as a Family Counselor before Marriage with the Aim of Divorce Rate Reduction”, 2nd International Conference on Research Knowledge Base in *Computer Engineering and Information Technology,* Tehran Iran (2017).

Colorni, A., Dorigo, M., Maniezzo, V., “Distributed Optimization by Ant several”, *Elsevier*, 134-142, (2014).

Raj, K.A.A.D., Rajendran, C., "Fast heuristic algorithms to solve a single-stage Fixed Charge Transportation Problem", International Journal of Operational Research 6 (3), 304–329 (2009).

Chapter 5: Conclusion

5. Conclusion

A comparison among ANFIS-type models; ANFIS, ANFIS-FFA, ANFIS-GA, ANFIS-PSO, ANFIS-ACO and an equation-based approach, MGGP for estimation of evaporation by making use of climatic variables was performed in this study. Sunshine, relative humidity, average temperature, maximum temperature, minimum temperature and wind speed have been considered as the climate variables for all models. This study also demonstrated the accuracy of all models with various statistical indices (RMSE, MRE, VAF, NASH, Biasness, MBE, MARE, SI and R2). Data were collected from the government database from the state of Arizona, USA. Two combinations of data set were studied to check the results and verify if they are similar in pattern or not. The data set is initially splited into two parts: training part and test part. Almost two-third, 59, of the total data set has been selected as a training data set, and one third, 26, of the total data set has been taken as a testing data set. First two third of data set was considered for training and rest one third was considered as testing for the first combination. Middle one third was taken for testing for second combination, where rest of the data set was considered for training. The data set was divided into two parts; training and testing. MATLAB was used as programming language and the results were shown in graphical form for training data and test data. The predicted results varied from model to model. This graphical presentation was very helpful to understand the comparison better, but, the "MATLAB" programming for MGGP model came up with a bunch of equations at the end. The best equation was finalized by analyzing all equations. This equation was applied and solved by using all data in Excel. To confirm the comparison, the observed and predicted evaporation values were plotted together for training data and testing data. Graphical presentation confirmed that all data set were trained properly. But graphical presentation for testing data set showed almost identical results, therefore, close investigations were needed. Therefore, some statistical tests were performed. To ensure a rigorous comparison of the models, an extended analysis was performed by using ten statistical indices; R^2, MAE, MRE, MARE, RMSE, RMSRE, SI, VAF, Bias, and NASH. After analyzing all the results, it was clear that all the models were suitable for the evaporation estimation, and the models were trained properly. Based on the statistical analysis, ANFIS-FFA showed less accuracy among all six models. Furthermore, ANFIS-FFA took long time to run than other models. According to R^2 values, almost all models showed identical results and took similar time to run. But, ANFIS, ANFIS-ACO and MGGP models were better among all models in all aspects. It was very difficult to declare the best model and conclude the study. ANFIS with ACO and MGGP presented even better results than traditional ANFIS model in some cases. According to RMSE and NASH value, ANFIS-ACO was more accurate, but, overall MGGP model was better than ANFIS-ACO model. Relative statistical indices were compared to verify better results.

According to the results, all six models are suitable for estimation of evaporation, but ANFIS, ANFIS-ACO and MGGP advocated better result with higher accuracy. The practical approach, MGGP showed the best results based on all accuracy indicator values. Though ANFIS-ACO and MGGP offered slightly better results than basic ANFIS model, ANFIS model was recommended for its simple formation. Also, the computational time was less for ANFIS model time than other models using with an optimizer. Therefore, ANFIS-ACO took more time to run as it worked as an optimizer to the basic ANFIS model, though ACO presented slightly better results in some cases. And for MGGP approach, all data were formed by calculating the equations provided by "MATLAB" programming. MGGP was somewhat complicated as it did not supply the results in graphical form; rather it came up with a set of equations. Therefore, MGGP model was also time consuming. Despite this, the practical approach, MGGP provides more accuracy at the end compared to ANFIS-ACO. After analyzing all the statistical indices, relative and absolute accuracy tests and, comparing the observed and predicted data sets, MGGP was considered as a best fitted model. Therefore, this module for calculating evaporation data in hydrological modeling studies. However, optimizers were observed and compared to check if the results are better than ANFIS or not. The objectives of the adoption of different optimizer technique and MGGP model was to verify the accuracy of the prediction and to provide accurate evaporation prediction which can contribute on agriculture and hydrological model in arid area. Both MGGP and ANFIS model can be used for calculating evaporation. As equation-based approach provided better results than ANFIS model, therefore, MGGP can be suggested as the best one. On the other hand, computational time of MGGP is higher and therefore, ANFIS can be suggested in the specific cases. ANFIS is also a very popular and well accepted artificial intelligence model in calculating evaporation for its simplicity.

Recommendations for Future studies

Estimation of evaporation is an important topic in hydrology. Research is going on in order to meet the higher accuracy on the evaporation forecasting. In order to do that, shortcomings of any model or approach can be an area of interest for the upcoming researches. The limitation of this study can be the target of future work. This study was performed for the hot climate conditions. Similar approaches can be applied for other climate conditions. Investigating the nature of forecasting evaporation in different weather conditions can be a research subject. Also, different weather variables combinations can be explored to observe which combination accounts for the best results with higher accuracy.

Appendix: Statistical indices and error measures

The relationships for statistical indices and error measures used in this paper are provided in the following.

R^2 :   Coefficient of determination, which can be expressed in the following form:

$$r = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}} \ldots\ldots (1)$$

RMSE : Root Mean Square Error, which can be formulated as follows:

$$RMSE = \left[\frac{\sum_{i=1}^{M}(Y_{i(model)}-Y_{i(actual)})}{M}\right]^{1/2} \ldots\ldots (2)$$

MARE : Absolute Relative Error. The formula is given below:

$$MARE = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{|Y_{i(model)}-Y_{i(actual)}|}{Y_{i(actual)}}\right) \ldots\ldots (3)$$

$$Bias = \frac{\sum_{i=1}^{M}(Y_{i(model)}-Y_{i(actual)})}{M} \ldots\ldots (4)$$

SI : Scatter Index which can be expressed as follows:

$$SI = \frac{RMSE}{\frac{1}{M}\sum_{i=1}^{M}(Y_{i(actual)})} \ldots\ldots (5)$$

RMSRE : Root Mean Square Relative Error. This error can be calculated from the following equation:

$$RMSRE = \frac{1}{N}\sqrt{\sum\left(\frac{y_t-\widehat{y_t}}{y_t}\right)^2} \ldots\ldots (6)$$

MAE : Mean Absolute Error. This error can be calculated from the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|T_{i.Actual} - T_{i.Predicted}| \ldots\ldots (7)$$

VAF : Variance Account For. This term can be presented by the following equation:

$$VAF = \left(\frac{1-var(T_{i.Actual}-T_{i.Predicted})}{var(T_{i.Actual})}\right) * 100 \ldots\ldots (8)$$

NSE : Nash-Sutcliffe coefficient. This coefficient can be formulated as follows:

$$E_{NSC} = 1 - \left(\frac{\sum(y_t-\widehat{y_t})^2}{\sum(y_t-\overline{y_t})^2}\right) \ldots\ldots (9)$$

Where,

$Y_{i(actual)}$: the output observational parameter

$Y_{i(model)}$: the y parameter predicted by the models

$Y_{i(model)}$: the mean predicted y parameter

M : the number of parameters

n : number of samples

E_{NSC} : the Nash-Sutcliffe test statistic

$T_{i.Actual}$: the ith value of actual data

$T_{i.Predicted}$: the ith value of predicted data

References

Chang, F.J., Chang, Y.T. et al., "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir", *Advance in Water Resources* 29,1-10 (2006).

Fi-John Chang, Li-Chiu Chang, Huey-Shan Kao, Gwo-Ru Wu, "Assessing the effort of meteorological variables for evaporation for evaporation estimation by self-organizing map neural network", *Journal of Hydrology* 384, 118-129 (2010).

Sheng-Feng Kuo, Feng-Wen Chen, Pei-Yu Liao, Chen-Wuing Liu, "A Comparative Study on the Estimation using backpropagation neural network: Penman-Monteith method versus pan Evaporation method," *Paddy Water Environment* 9, 413-424(2011).

Hossein Tabari, Safar Marofi, Ali-Akbar Sabziparvar, "Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression", *Springer* 28, 399-406 (2010).

Ali Rahimi Khoob, "Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment," *Springer* 27, 35-39 (2008).

Sungon Kim, Jalal Shiri, Vijay P. Singh, Ozgur Kisi and Gorka Landera, Predicting daily pan evaporation by soft computing models with limited climatic data, *"Hydrological Sciences Journal* 60, (2015).

Adebayo J. Adeloye, Rabee Rustum and Ibrahim D. Kariyama, "Neural Computing modeling of the reference crop evapotranspiration," *Elsevier* 29, 61-73 (2012).

Hai Tao, Lamine Diop, Ansoumana Bodian, Koffi Djaman, Papa Malick Ndiaye, Zaher Mundher Yaseen, "Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso," *Agricultural Water Management* 208,140-151(2018).

Lifeng Wu, Hanmi Zhou, Xin Ma, Junliang Fan, and Fucang Zhang, Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: Application in contrasting climates of China, *"Journal of Hydrology* 577, 123960, (2019).

Yu Feng, Yong Peng, Ningbo Cui, Daozhi Gong and Kuandi Zhang, "Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data", *"Computers and Electronics in Agriculture* 136, 71-78 (2017).

Sheng-Feng Kuo, Feng-Wen Chen, Pei-Yu Liao and Chen-Wuing Liu, "A comparative study on the estimation of evapotranspiration using backpropagation neural network: Penman–Monteith method versus pan evaporation method", *Paddy Water Environment* 9, 413-424 (2011).

Peng Xuanrong, Ye Tingdong and Wang Yuesheng, "Research and Design of Precision Irrigation System Based on Artificial Neural Network", *The 30th Chinese Control and Decision Conference",* (2011).

Yu Feng, Ningbo Cui, Lu Zhao, Xiatao Hu and Daozhi Gong, "Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southeast China", *Journal of Hydrology* 536, 376-383(2016).

Yang Yanga, Yuanlai Cui, Kaihua Bai, Tongyuan Luo, Junfeng Dai, Weiguang Wang and Yufeng Luo, “Short-term forecasting of daily reference evapotranspiration using the reduced-set Penman-Monteith model and public weather forecasts”, *Agricultural Water Management* 211, 70-80 (2019).

Mandeep Kaur Saggi and Sushma Jain, “Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning”, *Computers and Electronics in Agriculture* 156, 387-398 (2019).

Sheng Wanga, Jinjiao Lian, Yuzhong Peng, Baoqing Hu and Hongsong Chen, “Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China”, *Agricultural Water Management* 211, 220-230 (2019).

Lucas Borges Ferreira, Fernando França da Cunha, Rubens Alves de Oliveira and Elpídio Inácio Fernandes Filhob, “Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach”, *Journal of Hydrology* 572, 556-570 (2019).

Benzagtha, M. A., “Estimation of Evaporation from Reservoir in Semi-arid Environment Using Artificial Neural Network and Climate Based Model”, *British Journal of Applied Science and Technology 4*, 3501-3518, (2014).

Chang, F.J., Chang, Y.T., “Adaptive neuro-fuzzy inference system for prediction of water level in reservoir”, *Advance in Water Resources* 29, 1-10 (2006).

Ferreira, C., “Gene Expression Programming; a new Adaptive Algorithm for Solving Problems”, *Complex System* 13, 87-129, (2001).

Friedman, J. H., “Multivariate Adaptive Regression splines”, *Annals of Statistics*, 1-67, (1991).

Goyal, M. K., Bharti, B., Quilty, J., Adamowsky, J., Pandey, A., “Modeling of Daily Pan Evaporation in Subtropical Climates Using ANN, LS-SVR, Fuzzy Logic, and ANFIS”, *Expert Systems with Applications* 41, 5267-5276, (2014).

Ozgur Kişi, Ana Pour Ali Baba, and Jalal Shiri, “Generalized Neurofuzzy Models for Estimating DailyPan Evaporation Values from Weather Data”, *Journal of Irrigation and Drainage Engineering*, 138, 349-362, (2012).

Kisi, O., “Evolutionary Neural Networks for Monthly Pan Evaporation Modelling”, *Journal of Hydrology* 498, 36-45, (2013).

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., “Comparison of two Different Adaptive Neuro-Fuzzy Inference System in modelling Daily Reference Evapotranspiration”, *Water Resource Management* 28, 2655-2675, (2014).

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Niazi, F., “Long Term Monthly Evaporation Modeling by Several Data-Driven Methods Without Climate Data”, *Computers and Electronics in Agriculture* 115, 66-77, (2015).

Lu, J., Sun, G., McNulty, S. G., Amatya. D. M., “A Comparison of Six Potential Evapotranspiration Methods for Regional Use in the South-Eastern United States*”, Journal of the American Water Resources Association* 41, 621-633, (2005).

Moghaddamnia, A., Gousheh, M. G., Pirli, J., Amin, S., Han, D., “Evaporation Estimation Using Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System Techniques”, *Advances in Water Resources* 32, 88-97, (2009).

Shirsath, P. B., Singh, A. K., “A Comparative Study of Daily Pan Evaporation Estimation Using ANN, Regression and Climate Based Models”, *Water Resource Management* 24, 1571-1581, (2010).

Sanikhani, H., Kisi, O., Nikpour, M. R., and Dinpashoh, Y., “Estimation of Daily Pan Evaporation Using Two Different Adaptive Neuro-Fuzzy Computing Techniques”, *Water Resource Manage*ment 26, 4347–4365, (2012).

Stephens, J.C., Stewart, E.H., A comparison of procedures for computing evaporation and evapotranspiration, International Association of Scientific Hydrology. International Union of Geodynamics and Geophysics, Berkeley, CA, pp. 123- 133 (1963).

Wang, L., Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Li, H., “Pan Evaporation Modelling Six Different Heuristic Computing Methods in Different Climates in China”, *Journal of Hydrology* 544, 407-427, (2017).

Yaseen, Z.M., Ghareb, M.I., Ebtehaj, I., “Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA”, *Water Resource Management* 32, 105-122 (2018).

Yager R. R., Filev D.P., “Approximate clustering via the mountain method”, IEEE Transaction on Systems, Man and Cybernetics 24, 1279–1284, (1994).

Zaher Mundher Yaseen, Isa Ebtehaj, Hossein Bonakdari, Ravinesh C. Deo, Ali Danandeh Mehr, Wan Hanna Melini Wan Mohtar, Lamine Diop, Ahmed El-shafie, Vijay P. Singh, “Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model”, *Journal of Hydrology* 554, 263-276 (2017).

M. Tamer Ayvaz, Alper Elci, “Identification of the optimum groundwater quality monitoring network using a genetic algorithm-based optimization approach”, *Journal of Hydrology* 563, 1078-1091(2018).

Kennedy, J. and Eberhart, R. C., “Particle swarm optimization.” Proceedings of IEEE international Conference on Neural Networks, Perth, Australia, 1942-1948, (1995).

Halil Karahan, “Determining Rainfall-Intensity-Duration-Frequency Relationship Using Particle Swarm Optimization”, *Journal of Civil Engineering* 16, 667-675(2012).

Emrah Dogana, Mahnaz Gumrukcuoglu, Mehmet Sandalci, Mucahit Opan, “Modelling of evaporation from the reservoir of Yuvacik dam using adaptive neuro-fuzzy inference systems”, *Engineering Applications of Artificial Intelligence* 23, 961-967 (2010,).

Mohammad Ali Ghorbani, Reza Kazempour, Kwok-Wing Chaub, Shahaboddin Shamshirband and Pezhman Taherei Ghazvinei, “Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh,Northern Iran", *Engineering applications of computational fluid mechanics* 12, 724-737 (2018).

Silva, C. A., Sousa, J.M.C., Runkler, T.A., Sa Da Costa, J.M.C., "Distributed Supply Chain Management Using Ant Coloni Optimization", *European Journal of Operational Research* 199, 349-358, (2009).

Hong, J., Diabat, A., Panicker, V. V., Rajagopalan, S., "A two-stage Supply Chain Problem with Fixed costs: An Ant Coloni Optimization Approach", *International Journal of Production Economics* 204, 214-226, (2018).

Yan, X., and Mohammadian, M., "Multigene Genetic-Programming-Based Models for Initial Dilution of Laterally Confined Vertical Buoyant Jets", *Journal of Marine Science and Engineering* 7, 246, (2019).

Mousavi, S. M. H., MiriNezhad, S. Y., Layshenko, V., "An Evolutionary-Based Adaptive Neuro-Fuzzy Expert System as a Family Counselor before Marriage with the Aim of Divorce Rate Reduction", 2nd International Conference on Research Knowledge Base in *Computer Engineering and Information Technology,* Tehran Iran (2017).

Colorni, A., Dorigo, M., Maniezzo, V., "Distributed Optimization by Ant Colonies", *Elsevier*, 134-142, (2014).

Raj, K.A.A.D., Rajendran, C., Fast heuristic algorithms to solve a single-stage Fixed Charge Transportation Problem. *International Journal of Operational Research* 6, 304–329(2009).