Modelling Unconfined Groundwater Recharge Using Adaptive Neuro-Fuzzy Inference System

Authors:

Khaled Mohamed Nabil I. Elsayed, Rabee Rustum, Adebayo J. Adeloye

Abstract:

Estimating groundwater recharge using mathematical models such as water budget or soil water balance method has been proved to be very difficult due to the complex, uncertain multidimensional nature of the process, despite the simplicity of the concept. Artificial Intelligence (AI) techniques have been proposed to deal with this complexity and uncertainty in a similar way to human thinking and reasoning. This study proposed the use of the Adaptive Neuro-Fuzzy Inference System (ANFIS) to model unconfined groundwater recharge using a set of data records from Kaharoa monitoring site in the North Island of New Zealand. Fifty-three data points, comprising a set of input parameters such as rainfall, temperature, sunshine hours, and radiation, for a period of approximately four and a half years, have been used to estimate ground water recharge. The results suggest that the ANFIS model is overall a reliable estimator for groundwater recharge, the correlation coefficient of the model reached 93% using independent data set. The method is easy, flexible and reliable; hence, it is recommended to be used for similar applications.

Article

Modelling Unconfined Groundwater Recharge Using Adaptive Neuro-Fuzzy Inference System

Khaled Mohamed Nabil I. Elsayed [1], Rabee Rustum [1],* and Adebayo J. Adeloye [2]

[1]   Institute for Infrastructure and Environment, School of Energy, Geosciences, Infrastructure and Society, Heriot Watt University, Dubai Campus, Dubai Knowledge Park, Blocks 5 & 14, Dubai 38103, UAE; kme6@hw.ac.uk

[2]   Institute for Infrastructure and Environment, School of Energy, Geosciences, Infrastructure and Society, Heriot Watt University, Edinburgh EH14 4AS, UK; a.j.adeloye@hw.ac.uk

*   Correspondence: r.rustum@hw.ac.uk

check for updates

Abstract: Estimating groundwater recharge using mathematical models such as water budget or soil water balance method has been proved to be very difficult due to the complex, uncertain multidimensional nature of the process, despite the simplicity of the concept. Artificial Intelligence (AI) techniques have been proposed to deal with this complexity and uncertainty in a similar way to human thinking and reasoning. This study proposed the use of the Adaptive Neuro-Fuzzy Inference System (ANFIS) to model unconfined groundwater recharge using a set of data records from Kaharoa monitoring site in the North Island of New Zealand. Fifty-three data points, comprising a set of input parameters such as rainfall, temperature, sunshine hours, and radiation, for a period of approximately four and a half years, have been used to estimate ground water recharge. The results suggest that the ANFIS model is overall a reliable estimator for groundwater recharge, the correlation coefficient of the model reached 93% using independent data set. The method is easy, flexible and reliable; hence, it is recommended to be used for similar applications.

Keywords: groundwater recharge; fuzzy logic; adaptive neuro-fuzzy inference system; water budget; soil water balance; lysimeter

1. Introduction

Groundwater is simply the water found underneath the surface land, which plays a role in the hydrological cycle [1]. It has been the most important source of fresh water used in many aspects of life such as in agriculture, irrigation, industrial processes and domestic supplies. It is usually recharged naturally due to infiltration from rains and snow melt and, to some extent, by surface water such as rivers and lakes. According to the environmental protection agency in the United States, over the last 50 years, rainfall levels have been decreasing vigorously in many areas, therefore, leading to a decrease of natural recharge of groundwater [2].

The process of natural groundwater recharge occurs when precipitation on land surface percolates into soil and reaches the water table. Several factors must be considered when choosing the location for this process such as the characteristics of the soil and its texture, the required space, percolation rates, aquifer type, geological formations overlying the aquifer and silt control. All the factors contribute to how successful the recharge levels will be in terms of the time taken for the water to infiltrate to the aquifer, whereas heterogeneous soils would increase the time and distance for water to reach the aquifer. Additionally, temperature has a direct effect on infiltration rates where areas which encounter big differences between winter and summer have very low infiltration rates compared to the summer

season [3]. However, in summer, clogging usually occurs due to biological factors, leading to low infiltration rates.

The estimation of groundwater recharge has been a very difficult task in hydrology. Water moves within the hydrologic cycle along many complex pathways over a wide variety of time scales. The challenge for humans is to monitor the hydrologic cycle. Water budget states that the difference between the rates of water flowing in and out of an accounting unit is balanced by a change in water storage: Flow In − Flow Out = Change In Storage. The Water Budget Equation is simple, universal, and adaptable because it relies on few assumptions on mechanisms of water movement and storage. A basic water budget for a small watershed can be expressed as:

$$P + Q_{in} = ET + \Delta S + Q_{out} \tag{1}$$

where P is precipitation, Q_{in} is water flow into the watershed, ET is evapotranspiration (the sum of evaporation from soils, surface-water bodies, and plants), ΔS is change in water storage, and Q_{out} is water flow out of the watershed.

Precipitation can be written as the sum of rain, snow, hail, rime, hoarfrost, fog drip, and irrigation. Water flow into or out of the site could be surface or subsurface flow resulting from both natural and human-related causes. Evapotranspiration could be differentiated into evaporation and plant transpiration. Evaporation can occur from open water, bare soil, or snowpack (sublimation); plants can extract ground water or water from the unsaturated zone [4]. All of this makes the complexity and uncertainty part of the equation.

Soil water balance method is an application of the water budget equation for estimating recharge often under most condition. The recharge is estimated as the residual of all the other fluxes in the hydrological cycle, the rationale being that all these fluxes (i.e., precipitation, evapotranspiration, runoff, irrigation and soil moisture changes) are usually relatively more easily measured or estimated than the recharge. Therefore, considering a volume of balance water in the root zone, the recharge R_e is estimated from:

$$R_e = P + I - RO - ET + \Delta S \tag{2}$$

where I is irrigation, RO is runoff [5] and all other symbols have been defined previously. Yet, it is clear that the results from this equation have limited significance without calibration and validation and that is because of the substantial uncertainty in its inputting data. The equation parameters do not have a direct physical representation, which can be properly measured, in the field. At the same time, the equation parameters are indirect relation with other natural phenomena that can be easily collected from the field as rainfall, temperature, humidity, sunshine hours, and wind speed.

The fuzzy logic approach, also known as the grey box modelling, has been introduced as a solution and has been used in real-world systems where the relationship between variables is very complex or difficult to identify [6]. This is performed by using fuzzy IF THEN rules where any object can be given a condition between '0' and '1' rather than the classical or 'crisp' set, which would either have an integer of '0' or '1'. Adaptive Neuro-Fuzzy Inference System (ANFIS) is an approach used to model the recharge process in a similar way of human thinking and reasoning. It is a combination of fuzzy logic and neural networks introduced by Jang [7]. This approach proved to be flexible and accurate in modelling complex, nonlinear systems since it has the ability to deal with large sets of data inputs and can provide more accurate and reliable results in comparison to other simulation and numerical-based models. More details can be found in [8–10].

This paper will work on the approach of using fuzzy logic to reduce the uncertainties involved in the measurement of groundwater recharge by handling input data in a more simple and efficient manner. Since the water budget method involves data that is already available, it has been decided to use its parameters to attempt obtaining more accurate and reliable results. It aims to test the performance of ANFIS on a set of variables in modelling unconfined groundwater recharge using a set of data records, as well as investigating the effect of different membership function types on results.

2. Materials and Methods

2.1. Geology and Location of the Case Study

The data used in this study were from Kaharoa, it is in the Bay of Plenty region, North of Rotorua lake located in the North Island of New Zealand, it distances about 30 km from Rotorua city. The Kaharoa monitoring site has a Global Positioning System (GPS) of latitude $-38.012296°$ and longitude of $176.268337°$ on map and on an elevation of 420 m above sea level at a distance of about 2.3 km from North Rotorua, the water level in the lake is around 280 m above sea level [11]. Figure 1 illustrates the elevation section.
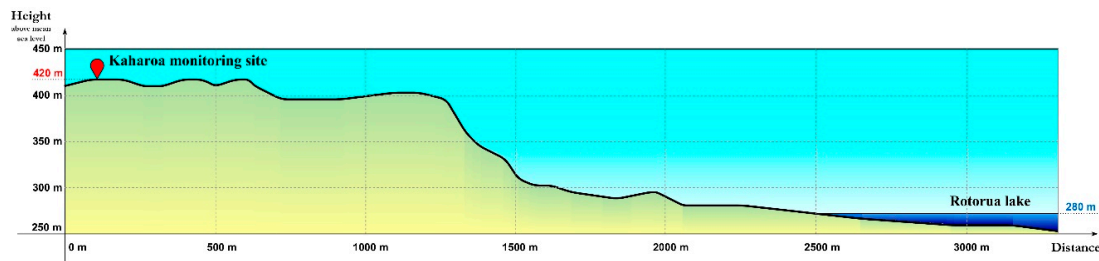

Figure 1. Topographic profile showing the elevation cross section at Kaharoa monitoring site.

The Kaharoa site overlays the Mamaku Ignimbrite, which is a type of rock unit that was caused from the eruption of a magma chamber in the Taupo volcanic zone now known as The Rotorua Caldera, having a formation of 145 km^3 in volume. The soil type of the site is classified as "Orthic Pumice Soils" and more specifically the "Oropi" soil series name [12]. Oropi soil series occur on flat, rolling and hilly uplands south of Te Puke and Tauranga and North of the Rotorua Caldera. This soil is characterized by low nutrient levels, strong leaching, sandy coarse texture which is well drained. Pumice soils in the Bay of Plenty Region, are dominated by pumice or pumice sands typically with a low clay content (<10%). Summer drought is a dominate characteristic. The soils are derived from the Kaharoa and Taupo Tephra overlaying rhyolitic tephra. They occur under an annual rainfall of 1900 to 2400 mm annually [12].

The specific physical properties of the pumice soils in the region are [12]:

Texture: Sand, Topsoil clay content: 5–10%;

Drainage class: Well-drained;

Permeability: Rapid;

Profile total available water (0–100 cm): Moderate to high (139 mm);

Profile readily available water (0–100 cm): High (87 mm);

Topsoil bulk density: 1.18 g/cm^3;

Subsoil bulk density: 0.84 g/cm^3.

The precipitation rates used in this study were collected at Kaharoa monitoring site by GNS Science under the supervision of the Bay of Plenty Regional Council. The site measures and records rainfall and rainfall percolation, collects soil water samples, besides analyzing the quality of the soil water that recharges the groundwater in Kaharoa.

2.2. Kaharoa Monitoring Site Installation

Kaharoa monitoring site was established directly beside an old rainfall recording site which has been the primary rain gauge under the number of [860205] (conventional elevated rain gauge), the site is shown in Figure 2 showing a closer look of the lysimeters. The site consists of another two non-weighing cylindrical soil monolith lysimeters, also known as percolation lysimeters, of a 500 mm diameter and a 700 mm depth and are constructed of galvanized steel connected to pipework made of polyethylene. The lysimeters are named under the number [2597] and [2598], as well as a ground-level rain gauge numbered [2599] in which a tipping bucket is attached to it to measure percolation and for

real-time data collecting as demonstrated in Figure 3. The data captured were reported to be reviewed monthly, seasonally, and annually concerning a level of rainfall and rainfall recharge at Kaharoa [13].
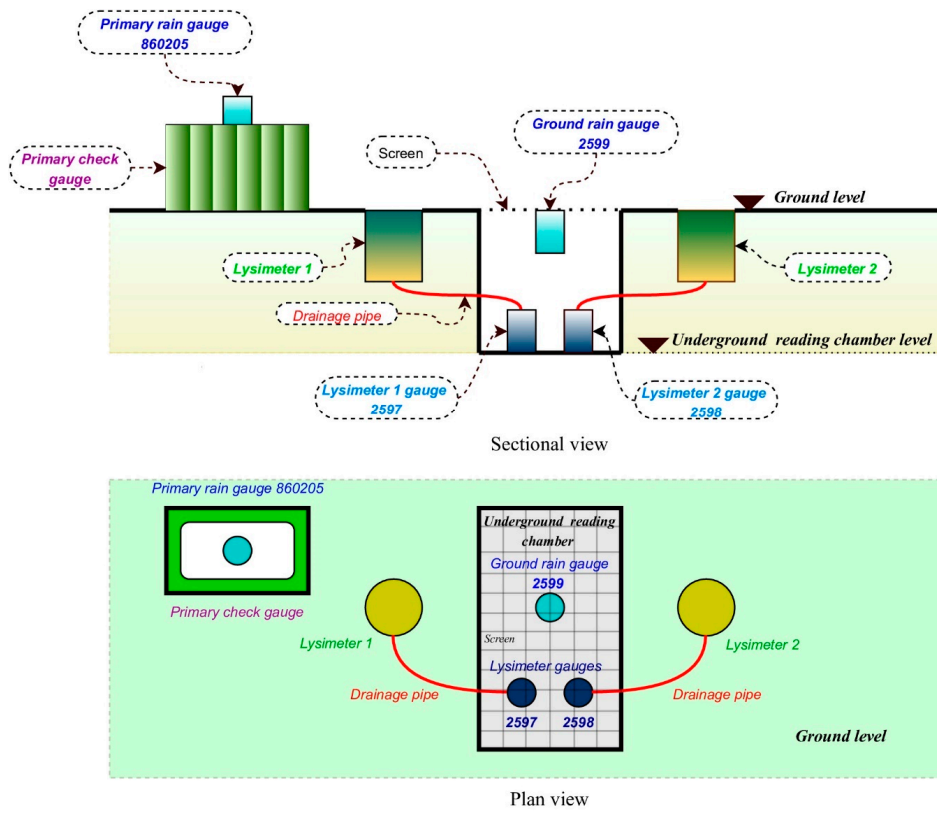

Figure 2. Kaharoa Lysimeter Instruments, upper diagram shows the position of gauges through sectional view, while lower diagram shows the location of gauges in plan view.
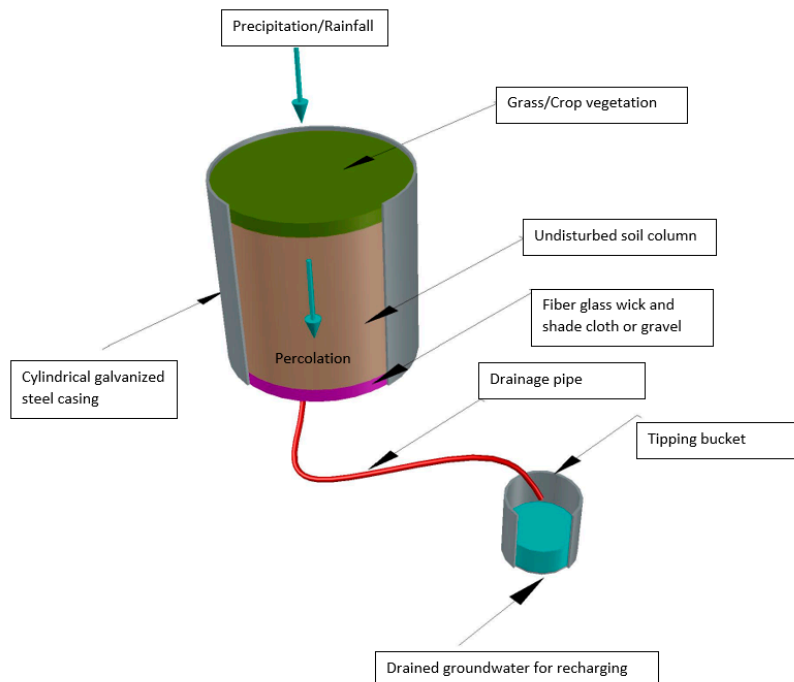

Figure 3. Non-weighing (free drainage) lysimeter.

2.3. Lysimeters in Groundwater Recharge

In general, a lysimeter is a device used for measuring the amount of actual evapotranspiration released by plants, usually crops. More specifically, it is a device that isolates a volume of soil between the soil surface and a given depth and includes a percolating water sampling system at its base [13].

The soil–water balance is the simplest method when dealing with lysimeters and has been used to represent the relationship between input and output. An ANFIS model has been prepared showing the relationship between input and output using a set of existing training data packages, which will then be used as a future prediction tool. Refer to Section 2.6 regarding ANFIS for more detail. The main soil–water balance method in determining the evapotranspiration is shown in Equation (3) [14].

$$ET = I + P - RO - R_e + CR \, (\pm\Delta \, SF \, \pm \, \Delta \, SW) \tag{3}$$

where *ET* is Evapotranspiration; *I* is Irrigation; *P* is Precipitation; *RO* is Runoff; *CR* is Capillary rise (upward the root zone in case of shallow water table existence); ΔSF is subsurface flow (horizontally inflow to or outflow of the root zone); ΔSW is soil water content (moisture content, soil–water storage).

However, the use of lysimeters in this study has eliminated some of the parameters in Equation (3), like (*I*) irrigation, as the area under study is a rural pastoral grassland which depends on the rainfall only in vegetation process; (ΔSF) subsurface flow for in and out flow are stopped because of the shape of the lysimeter structural design using bounded cylinder casing; additionally, the non-existing of the shallow groundwater table makes the upward (*CR*) capillarity rising are not found and the rarely availability of (*RO*) runoff.

Thus, the previous equation can be reduced as shown in Equation (4):

$$ET = P - R_e - \Delta SW \tag{4}$$

The Kaharoa monitoring site records the precipitation (rainfall) and the drainage discharge (recharge) for groundwater daily. However, the full records are presented monthly as changes in soil water content (water storage) are not seen to be significant. The differences between the precipitation and drainage explain the losses in the water due to the evapotranspiration and changes in moisture content combined by using the free drainage lysimeter system, whereas the weighted lysimeter makes the assurance of the evapotranspiration values and moisture content hard to obtain. Depending on the knowledge of the precipitation and drainage or, in other words, the rainfall and recharge groundwater, and in the non-presence of other parameters, it was recommended to use the fuzzy logic techniques for forecasting the recharging water based on the available lysimeter records.

2.4. Data

The rainfall recharge site recorded monthly readings collected over 5 years, starting from December 2005 till May 2010. The site collects rainfall and recharge with two rain gauges [2599860205] and two recharge gauges [2597 and 2598]. The recharge gauge [2598] was eliminated due to its inaccuracy where a leakage was observed. The Kaharoa site is classified as a rain gauge data collector that can take records on an interval of 15 min, which can easily monitor events even for short-term rainfall. Therefore, it was preferred to use the ground rainfall gauge [2599] over [860205] as it is relatively very near to lysimeter's location and it imitates the same amount received; thus, the record readings in [2599] for rainfall is taken in consideration with lysimeter [2597] for recharge in this study.

The data record had to be divided into three packages: a package for 'Training data', 'Validation or checking data' and finally for 'Testing data'. These data were not enough for generating a proper fuzzy logic model as more variables are needed to improve the performance of the model in predicting the groundwater recharge. Temperature in Celsius degree, relative humidity and sunshine in hours, published on the certified environmental New Zealand official site, 2019, was used to support the modelling exercise. These data readings were noted to be mean monthly values ranging from the year

1981 till 2010 for the location under study having a minimum of five full-year data and scheduled yearly by monthly sequence starting from January till December.

A 53 data readings were divided into three sets. The first 29 readings starting from December 2005 till April 2008 were used for training, 12 readings starting from May 2008 till April 2009 were used for validation (checking) and last 12 readings which starts from May 2009 till April 2010 were used for testing, as shown in Table 1.

Table 1. Data readings from 2005 to 2010.

DATA Sets	Reading	Month-Year	Input				Output	
			Rainfall (mm)	Temperature (Co)	Relative Humidity (%)	Sunshine Hours	Recharge mm	Rainfall Recharge Ratio
Training Set	1	Dec-05	237	16.2	78.9	215.8	82	35%
	2	Jan-06	288	17.7	78.8	242.9	150.56	52%
	3	Feb-06	198	17.9	81.4	205.9	113.68	57%
	4	Mar-06	192	16	81.5	199.7	67.52	35%
	5	Apr-06	338	13.3	83.4	170.5	211.28	63%
	6	May-06	253	10.7	87.1	145.1	170.4	67%
	7	Jun-06	110	8.5	87.5	119.1	67.92	62%
	8	Jul-06	119	7.8	87.3	130.7	63.44	53%
	9	Aug-06	202	8.4	85.9	152.1	145.76	72%
	10	Sep-06	44	10.2	81.6	155.1	11.52	26%
	11	Oct-06	128	12	79.7	190.8	43.12	34%
	12	Nov-06	104	13.9	77.2	200.1	22.4	22%
	13	Dec-06	70	16.2	78.9	215.8	2.16	3%
	14	Jan-07	219	17.7	78.8	242.9	122	56%
	15	Feb-07	54	17.9	81.4	205.9	0	0%
	16	Mar-07	286	16	81.5	199.7	175.04	61%
	17	Apr-07	89	13.3	83.4	170.5	26.48	30%
	18	May-07	66	10.7	87.1	145.1	58.64	89%
	19	Jun-07	169	8.5	87.5	119.1	106.96	63%
	20	Jul-07	201	7.8	87.3	130.7	208.56	104%
	21	Aug-07	189	8.4	85.9	152.1	174.64	92%
	22	Sep-07	100	10.2	81.6	155.1	48.88	49%
	23	Oct-07	115	12	79.7	190.8	65.44	57%
	24	Nov-07	54	13.9	77.2	200.1	0	0%
	25	Dec-07	145	16.2	78.9	215.8	30.96	21%
	26	Jan-08	76	17.7	78.8	242.9	0	0%
	27	Feb-08	92	17.9	81.4	205.9	0	0%
	28	Mar-08	70	16	81.5	199.7	0	0%
	29	Apr-08	437	13.3	83.4	170.5	220.88	51%
Validity Set	30	May-08	114	10.7	87.1	145.1	60.4	53%
	31	Jun-08	252	8.5	87.5	119.1	158.48	63%
	32	Jul-08	235	7.8	87.3	130.7	202.32	86%
	33	Aug-08	152	8.4	85.9	152.1	184	121%
	34	Sep-08	180	10.2	81.6	155.1	106	59%
	35	Oct-08	148	12	79.7	190.8	62.64	42%
	36	Nov-08	85	13.9	77.2	200.1	5.44	6%
	37	Dec-08	150	16.2	78.9	215.8	18	12%
	38	Jan-09	166	17.7	78.8	242.9	3.28	2%
	39	Feb-09	118	17.9	81.4	205.9	56.16	48%
	40	Mar-09	175	16	81.5	199.7	0	0%
	41	Apr-09	94	13.3	83.4	170.5	9.12	10%

Table 1. *Cont.*

| DATA Sets | Reading | Month-Year | Input | | | | Output | |
			Rainfall (mm)	Temperature (Co)	Relative Humidity (%)	Sunshine Hours	Recharge mm	Rainfall Recharge Ratio
Testing Set	42	May-09	227	10.7	87.1	145.1	164.24	72%
	43	Jun-09	169	8.5	87.5	119.1	115.04	68%
	44	Jul-09	256	7.8	87.3	130.7	191.2	75%
	45	Aug-09	157	8.4	85.9	152.1	83.36	53%
	46	Sep-09	158	10.2	81.6	155.1	71.12	45%
	47	Oct-09	27	12	79.7	190.8	0	0%
	48	Nov-09	80	13.9	77.2	200.1	0	0%
	49	Dec-09	146	16.2	78.9	215.8	0	0%
	50	Jan-10	93	17.7	78.8	242.9	14.16	15%
	51	Feb-10	30	17.9	81.4	205.9	0	0%
	52	Mar-10	69	16	81.5	199.7	0	0%
	53	Apr-10	429	13.3	83.4	170.5	199.92	47%

The average recharge to rainfall was around 40% per month, reaching a maximum of 121% in August 2008 (more research to be conducted for the reason behind this as this out the scope of this study, but they are kept as the fuzzy model could eliminate their effects) and minimum of 0% which happened several times as counted 11 times for 53 records. Firstly, in February 2007 and lastly in March 2010, twice in 2007, three times in 2008, four times in 2009 and, lastly, two times in the uncompleted year 2010. That might indicate the reason behind the temperature rising and less rainfall.

Figure 4 illustrates the regression comparison between the input parameters and the output parameter (recharge) as in Figure 4a, shows that the rainfall to recharge $R^2 = 0.699$ that is a fairly linearly; however, the regression for the other input parameters temperature, relative humidity and sunshine hours represents ($R^2 = 0.24$, 0.35 and 0.25) a non-linearity nature.
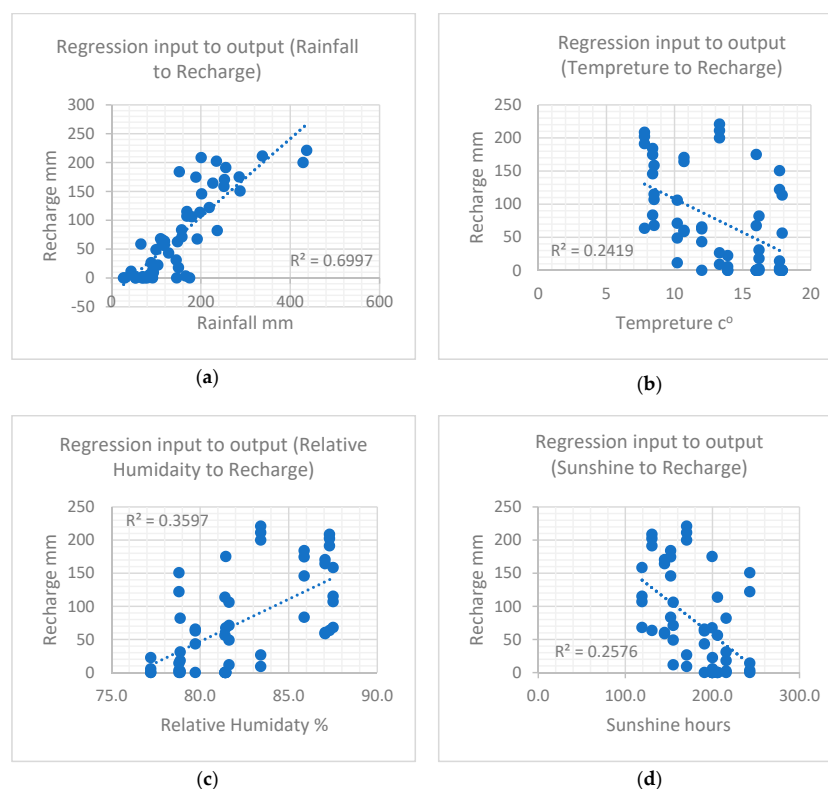

Figure 4. Graphs showing regression of the four parameters between input and output (**a**) rainfall; (**b**) tempreture; (**c**) relative humidaty; (**d**) sunshine hours.

2.5. Fuzzy Logic Inference Systems

Fuzzy logic systems are non-linear mapping of the input space to the output space, where the input is first converted from a crisp integer to a fuzzy value, known as 'fuzzification' through the use of membership functions [15]. It is then processed through a set of fuzzy rules. The output for each rule is then converted from fuzzy to a crisp value, this process is known as 'defuzzification'. The main components used in this system are fuzzy-sets, membership functions, fuzzy IF-THEN rules and fuzzy-logic operators.

There are two main types of fuzzy inference systems that could be used, the first inference system is called Mamdani's fuzzy system [15]. The second one is known as Takagi-Sugeno-Kang inference system (TSK) [16]. Mamdani's inference system relies that the consequent membership functions are fuzzy in nature and all the system is totally fuzzy starting from input, fuzzy rules and output variables. The Sugeno method is another approach for generating a set of fuzzy rules for a given input–output data. It is the same as the Mamdani method, however, the difference comes in the consequent part, which consists of linear or constant membership functions rather than fuzzy sets. Overall, Sugeno inference has been recognised to be flexible, more effective than Mamdani inference due to its efficiency in data processing and simplicity in the defuzzification.

The Adaptive Neuro Fuzzy Inference system (ANFIS), which is chosen in this paper, is a graphical representation of the Sugeno inference system embedded with neural network learning capabilities, where each output per rule can be linear or constant functions [17]. It provides an optimization method for dealing with the parameters in the fuzzy system that best fit with the input data. The diagram in Figure 5 illustrate the ANFIS model.
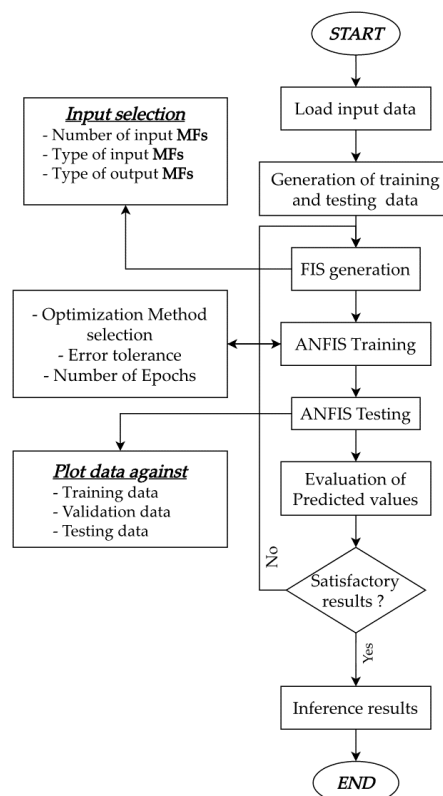

Figure 5. Step-by-step structure of the Adaptive Neuro-Fuzzy Inference System (ANFIS) model stages.

2.6. Modeling Framework

The fuzzy logic toolbox in MATLAB (Version 9.5 (R2018b), MathWorks, Natick, Massachusetts, United States, 12 September 2018) [18] was chosen for this study. Sugeno type was chosen using the

ANFIS. Initially, four main input parameters were chosen, rainfall, temperature, relative humidity percentage and sunshine hours. Relative humidity percentage was used to express the effect of evaporation and evapotranspiration, Sunshine hours to express the effect of solar radiations, whilst recharge was the output value. Gaussian membership function was used for the input space while linear membership function was chosen for the output space. For parameter optimization, hybrid method backpropagation was used, and the error tolerance was sat as Zero and the number of epochs was set to avoid over fitting of the model. The best structure of the model, the number of inputs and the number of memberships function, was found by trial and error in order to get the best model performance using the performance evaluation criteria.

The study is divided into two stages, the first stage aims to get the best ANFIS model from a limited number of input parameters by changing the number of membership functions and keeping the type of the membership function constant, while the second stage aims to get the best membership function type that increases the performance of the best ANFIS from the first stage by examining the different types. Available data points were divided into 3 parts, 29 data points are used for training, 12 data points are used for validation and checking, and 12 data points are used for testing.

The outputs of the ANFIS model are compared with the observed readings from the field (the lysimeter gauges records) to see how far the outputs are related to the field and how much is the error weights, by using several statistical methods. These statistical methods elect the most adapted ANFIS model that could be as nearest as the recorded field reading with less error differences.

The number of input parameters to be estimated in the ANFIS can be very large. For example, depending on the number of input parameters and the number of membership functions for each input, and the shape of membership function chosen, the total number of parameters is estimated by Equations (5)–(7) [19].

$$N_{total} = \left(N_{input} \times N_{mf} \times N_{pp}\right) + \left(l \times N_{cp}\right) \tag{5}$$

$$l = \left(N_{mf}\right)^{N_{input}} \tag{6}$$

$$N_{cp} = N_{input} + 1 \tag{7}$$

where

N_{total} = Number of modified parameters, i.e., to be estimated;
N_{input} = Number of input parameters;
N_{mf} = Number of membership functions associated with each input;
N_{pp} = Number of modified parameters per membership function, i.e., 2 in the case of Gaussian membership function;
l = Number of rules;
N_{cp} = Number of modified parameters in the sequence part of each rule.

The first stage of analysis consists of two trial set models, the first trail set is conducted with ANFIS using MATLAB for three inputs (rainfall, temperature, and relative humidity) and recharge as an output having 3 to 8 Gaussian membership functions generating 27 rules as starting up with three membership functions to 512 rules with 8 membership functions, while the second trail set measures the effect of adding the sunshine hours as the 4th input variable parameter to the ANFIS model and keeping the previously used sequence, starting with 3 to 6 membership functions of the Gaussian type generating 81 rules as starting up to 1296 rules.

The second stage of analysis aims for more refining and perfection for the ANFIS model and for studying the effect of changing the type of the membership function on the outputs. The first stage assumption for the membership function type was Gaussian type; in the second stage, it was replaced by other types of membership functions in which it have to be checked versus all other types through ANFIS modelling; the other types used in this study are tri_{mf} (triangle type), $trap_{mf}$ (trapezoidal type), $gbell_{mf}$ (Gaussian bell type), $gauss2_{mf}$ (2nd Gaussian type), pi_{mf} (Pi shaped) and $dsig_{mf}$ (difference

between two sigmoidal functions). Figure 6 shows a detailed structure of the two-stage analysis for a better understanding of the overall model.
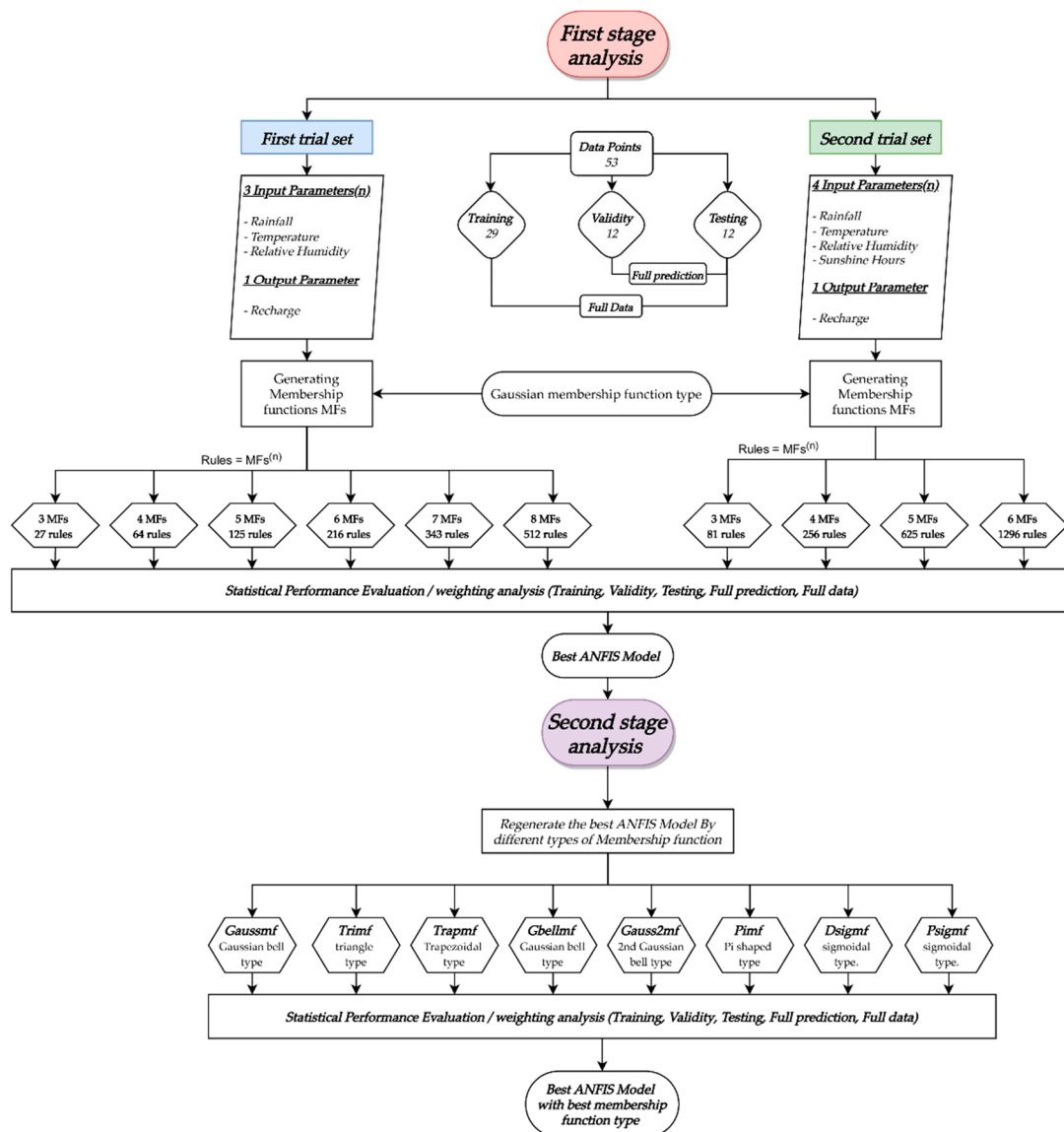

Figure 6. Structure of first- and second-stage analysis.

2.7. Performance Evaluation Criteria

The model has been evaluated using several evaluation criteria such as correlation coefficient (R), mean squared error (MSE), average absolute error (AAE), root mean square error (RMSE), and coefficient of efficiency (CE). More details on these criteria can be found in [19,20].

3. Results

In the first stage of analysis, two trial set models have been developed. The first trial set model used three input parameters, whereas the second trial set model used four input parameters, both used different numbers of Gaussian membership functions. The performance of these models in predicting recharge (the lysimeter gauges records) is presented in Table 2.

Table 2. First-stage analysis (two trial sets) ANFIS models performance using different input parameters and different Gaussian MFs.

	No. of Membership Functions	No. of Rules	Evaluation Criteria	Training (T)	Validity (V)	Testing (Ts)	Validity + Testing (VTs)	Total (TVTs)
	3	27	R	1.00	0.04	0.12	0.09	0.52
	4	64	R	1.00	0.73	0.81	0.77	0.83
	5	125	R	1.00	0.89	0.87	0.87	0.90
3 INPUTS	6	216	R	1.00	0.71	0.49	0.59	0.79
(Rainfall			R	1.00	0.77	0.93	0.86	0.93
Temperature			MSE	0.00	2237	846	1541	698
R-Humidity)	7	343	MAE	0.00	36.37	20.03	28.20	12.77
			RMSE	0.00	47.30	29.08	39.26	26.42
			CE	1.00	0.55	0.86	0.72	0.87
	8	512	R	1.00	0.74	0.66	0.69	0.86
	3	81	R	1.00	0.57	0.78	0.67	0.77
4 INPUTS	4	256	R	1.00	0.78	0.85	0.81	0.85
Rainfall			R	1.00	0.90	0.79	0.84	0.89
(Temperature			MSE	0.00	2509	4898	3704	1677
R-Humidity	5	625	MAE	0.00	38.68	40.35	39.52	17.90
Sunshine			RMSE	0.00	50.09	69.98	60.86	40.95
hours)			CE	1.00	0.50	0.16	0.32	0.68
	6	1296	R	1.00	0.74	0.66	0.69	0.86

The training set yield relatively the same values for the observed, however the differences are clearly noticed in the validation and testing sets in comparison to the observed. At the same time, it expresses some limitations that must be taken into consideration for future improvements, these improvements must include more readings and data entry as the more values for the training set the more output perfection.

By comparing the developed models, it is found that three input parameters with seven membership functions generate the best performance. Figure 7 shows the comparison between predictions observed during training, validation, and testing.
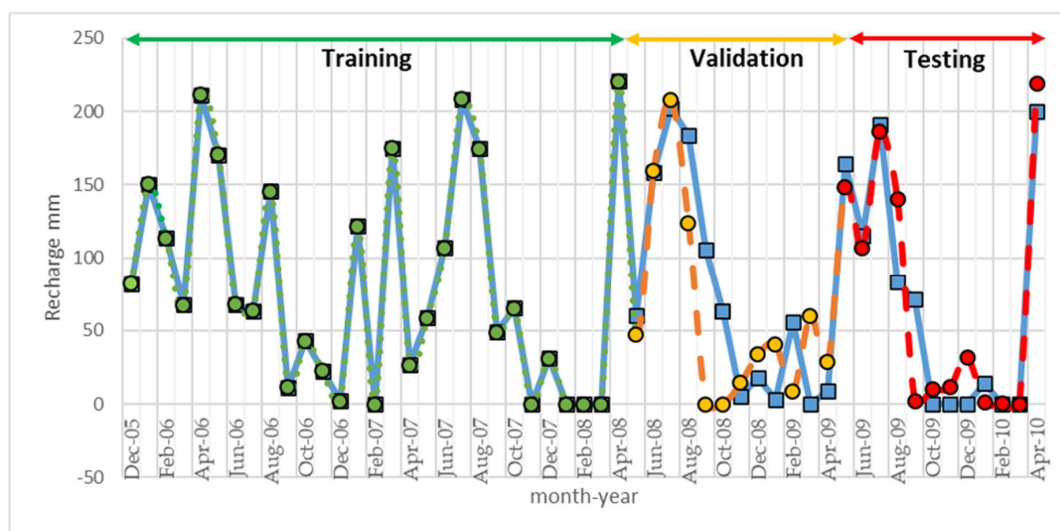

Figure 7. Time series plot of the performance of the ANFIS model (3×7Mfs) during training, validation and testing for the first-stage analysis (1st and 2nd trail set).

The number of parameters used is a solid part of the ANFIS modelling, more effective variables equals more accuracy and imitation of the real filed observation. The first trial set was concerning three constant numbers of parameters (rainfall, temperature and relative humidity), equally incrementing

the number of membership functions for each parameter, while the second trail set was concerning four parameters instead of three parameters by adding the parameter of sunshine hours. Despite increasing the number of parameters and that the ANFIS models generates a greater number of rules, the best model came in the first trail set, three parameters and seven MFs. This was due to the fact that the "sunshine hours" parameter was noticeably in the same range as the temperature parameter, therefore giving the same meaning with not much effect on the output.

For more refinement and perfection of the ANFIS model, and for studying the effect of changing the type of the function on the outputs, the previous assumption for the membership function type as Gaussian have to be checked versus all the other types that can be used through ANFIS modelling, as trimf (triangle type), trapmf (Trapezoidal type), gbellmf (Gaussian bell type), gauss2mf (2nd Gaussian type), pimf (Pi shaped) and dsigmf (Difference between two sigmoidal functions). This is implemented in stage 2 analysis. The performance of different models using correlation coefficient is shown in Table 3.

Table 3. The performance of the ANFIS model using different types of membership functions using correlation coefficient R.

Membership Function Type	Training Data	Validity Data	Testing Data	Full Prediction Data (Validity +Testing)	Full Data (Training + Validity + Testing)
$Gauss_{mf}$	1.00	0.72	0.93	0.83	0.92
Tri_{mf}	1.00	0.75	0.93	0.85	0.93
$Trap_{mf}$	1.00	0.78	0.93	0.85	0.93
$Gbell_{mf}$	1.00	0.66	0.92	0.79	0.89
$Gauss2_{mf}$	1.00	0.72	0.93	0.83	0.92
Pi_{mf}	1.00	0.79	0.93	0.86	0.93
$Dsig_{mf}$	1.00	0.56	0.92	0.72	0.85
$Psig_{mf}$	1.00	0.56	0.92	0.72	0.85

As seen from Table 4, the tri_{mf} (triangle) membership function type was the most statistically agreed on by 13/20 or 65% of all the comparisons; however, the differences between it and the Gaussian type is relatively small and the same for $trap_{mf}$, pi_{mf} and $gauss2_{mf}$, yet it was touchable in comparison to $dsig_{mf}$, $psig_{mf}$ and $gbell_{mf}$. Figure 8 illustrates the performance of the best model $3 \times 7Mfs$ type Trimf.

Table 4. Performance evaluation as statistics comparison for the second-stage analysis.

|	Results 2nd Stage Analysis	Training Data	Validity Data	Testing Data	Full Prediction Data (Validity + Testing)	Full Data (Training + Validity + Testing)
R correlation	Best Value	1.00	0.793	0.932	0.860	0.930
	Best Type	All	Pi_{mf}	$Gauss_{mf}$	Pi_{mf}	Tri_{mf}
MSE	Best Value	1.00	2432.85	823.03	1637.48	741.50
	Best Type	All	$Trap_{mf}$	Tri_{mf}	Tri_{mf}	Tri_{mf}
AAE	Best Value	1.00	37.88	20.03	29.58	13.39
	Best Type	All	Tri_{mf}	$Trap_{mf}$	Tri_{mf}	Tri_{mf}
RMSE	Best Value	1.00	49.32	28.69	40.47	27.23
	Best Type	All	$Trap_{mf}$	Tri_{mf}	Tri_{mf}	Tri_{mf}
CE	Best Value	1.00	0.51	0.86	0.70	0.86
	Best Type	All	$Trap_{mf}$	Tri_{mf}	Tri_{mf}	Tri_{mf}
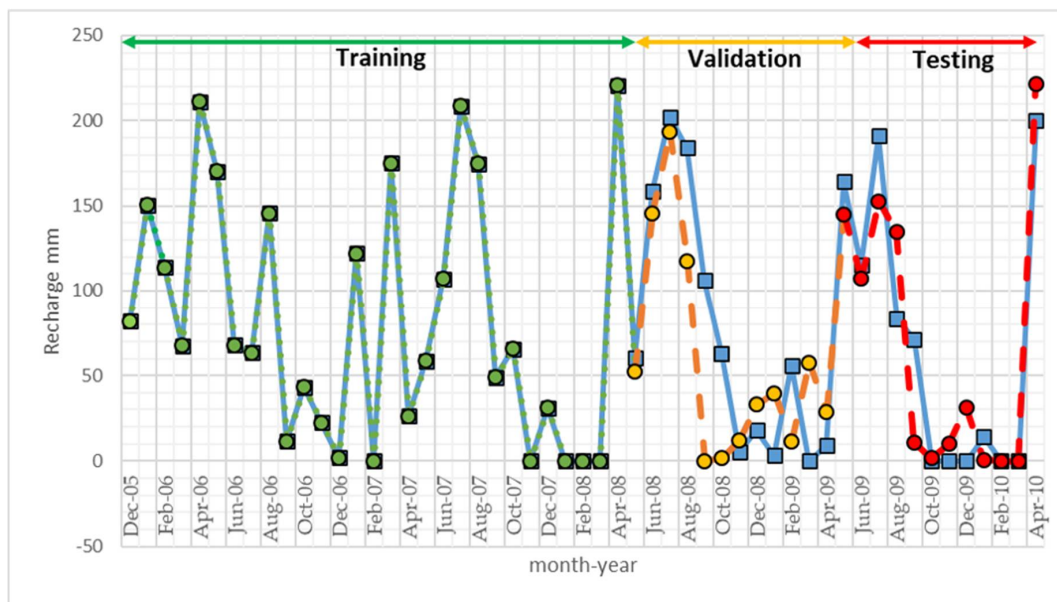
Figure 8. Time series plot of the performance of the ANFIS model (3 × 7Mfs) during training, validation and testing for the second-stage analysis.

4. Discussion

Predicting groundwater recharge in the past has been a very difficult process when using the traditional deterministic methods due to its complexity and uncertainty of data; therefore, the use of artificial intelligence techniques was seen to be the most successful in the estimation of groundwater recharge due to the fact that it can handle large amounts of data in a simple manner, with no prior knowledge of the structure of the system. However, these techniques depend greatly on the quality of the data provided.

This work attempted to test the hypothesis that AI techniques can significantly help in forecasting future readings by minimizing the maximum prediction error rates and reducing the uncertainties involved by handling input data in a more simple and efficient manner. The main aim was to use ANFIS to predict the groundwater recharge using some easily measured parameters such as temperature, rainfall, relative humidity and sunshine radiation. Other objectives included investigating whether or not the number of input parameters have a direct effect on the results. Additionally, to investigate the effect of different membership function types on the results. The benefits and merits of using ANFIS modelling for predicting groundwater recharge and the obtained results corresponding to our objectives are discussed in the following paragraphs.

The paradox of prediction of groundwater recharging theoretically requires measuring of many parameters which may be unmeasurable or difficult to ensure. The use of ANFIS model with a few numbers of easy measurable parameters provides us with significant results. For example, with the help of 29 readings of 3 parameters, 3 membership functions in training set, 52% is reached as a correlation coefficient (R) for a total set of data as a starting point. Moving to the best ANFIS model, three parameters and seven membership functions, the correlation coefficient increased to be 93% for the total set, 77% for the validity set and 93% for the testing set.

For more investigation, Figures 9–11 show the relation between the correlation coefficient (R) and ANFIS model number of generated rules. They reveal the fluctuation of the three-parameter models which start with a low range, and then increase with increasing number of membership functions in the mentioned three sets. In the four parameters, the curve rises gradually to a peak before then fallings down gradually. Here, the curve could either move upwards, which is less likely, or continue moving downwards. Hereby, the idea of the added fourth parameter may not be as effective as expected; thus,

the choice of parameters greatly depends on the quality of the data and how strong and effective it is in connection to the point of study rather than the number of parameters.
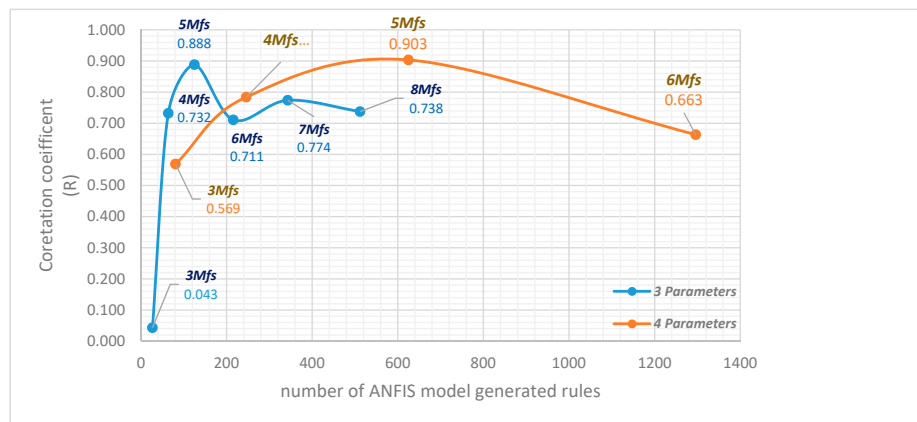

Figure 9. Correlation coefficient (R) vs ANFIS generated rules for validation data set (first-stage analysis).
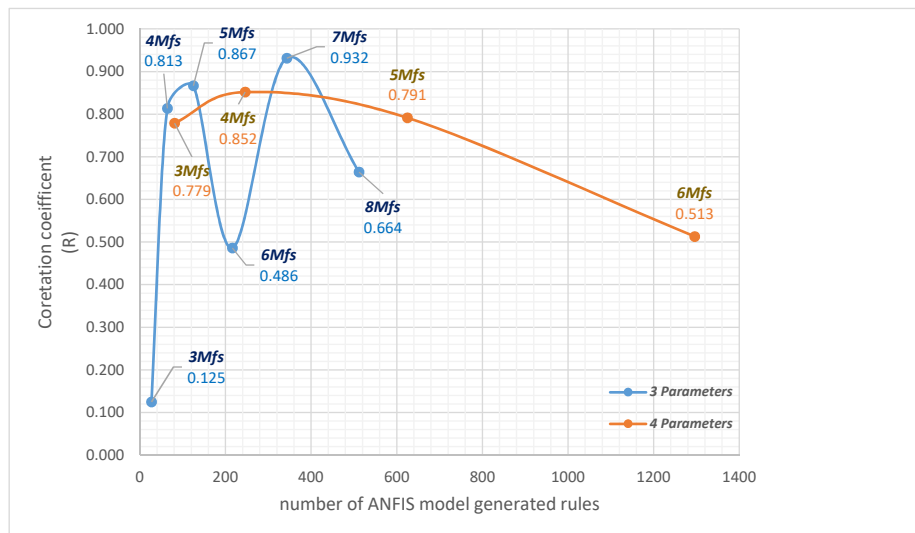

Figure 10. Correlation coefficient (R) vs ANFIS generated rules for testing data set (first-stage analysis).
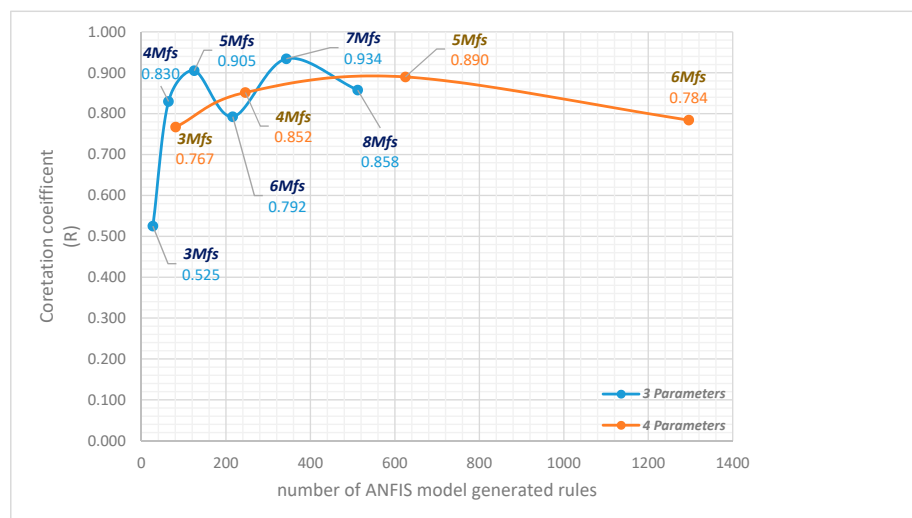

Figure 11. Correlation coefficient (R) vs ANFIS generated rules for total sets (first-stage analysis).

ANFIS modelling contains eight different types of membership functions with different behaviors for the same number of parameters and number of membership functions, the second-stage analysis in the study highlights the importance of the type of the membership functions used in modelling. The type of the curve was established to be Gaussian type (gauss$_{mf}$) as an assumption, eventually a comparison is driven between different types to assure the type for the best ANFIS model. Changing of types can make a direct impact on the output values. In this case, the superiority of the tri$_{mf}$ type on the gauss$_{mf}$ is not that much remarked.

The findings appear to be consistent with other AI techniques applied to groundwater recharge estimations. For example, Umamaheswari and Kalamani [19], who tried to develop the best ANFIS model that could reflect the groundwater fluctuation based on four input parameters as duration, groundwater recharge and groundwater discharge, and monthly groundwater level as output, had results showing that R^2 gives 0.921, CE gives less than 0.1, CC gives more than 0.96, MSE gives less than 0 and RMSE gives 0.3. These results are very close to the results found from this paper where R^2 was 0.865, CE was 0.86, CC was 0.93, the rest of MSE and RMSE are not applicable for comparison.

In comparison to another work by Raghavendra and Deka [21], who also compared two methods for prediction of groundwater recharge using ANFIS and GPR (Gaussian Process Regression) modelling based on two wells, our results were clearly better than their findings when comparing the performance evaluation for ANFIS models. The work expressed good results as NSE was 0.86, while their research gave 0.82 as maximum and 0.71 as minimum value. For CC, these results gave 0.93, while their research gave 0.85 as maximum and 0.75 as minimum value. These differences could be due to different time periods being used as well as different types of parameters involved.

The ANFIS modelling is a highly talented artificial technique that can solve and predict with high accuracy with no need for extensive or deep data exploration. Few parameters and data can eliminate many of uncertainties and unmeasurable data parameters.

5. Conclusions

The specific conclusions of this study are:

- First-stage analysis results revealed that the ANFIS model is capable of getting the statistically best prediction outputs from a limited data resource, by the help of the weighting analysis, the best model was for three parameters with seven MFs for each.
- The use of more input parameters in the second trial model set did not really contribute to any significant differences in comparison to the first trial model set which had fewer parameters; therefore, a higher number of parameters does not necessarily mean better results.
- Second-stage analysis results reflect the importance of the type of the MF where triangle (trimf) was elected as best MF type; however, the differences between the type Triangle (trimf) and Gaussian (gaussmf) was actually very minor, whereas the rest of the results for the other types revealed a significantly high variety between their results.
- Our findings were remarkably successful, giving highly accurate results than other researcher's work conducted in the same field who used ANFIS as well as other predicting tools for groundwater recharge estimation.
- The use of Artificial Intelligence, especially fuzzy rule based, has proven to be an important tool specifically in the prediction of groundwater recharge, which can provide very accurate and reliable results in a short period of time where large and complex amounts of data are being used, providing minimal error rates.
- We can now conclude that the main aim of investigating and evaluating the performance of ANFIS as a highly recommended tool for estimation of groundwater recharge has been successfully accomplished.

References

1. Senthilkumar, P.; Nithya, J.; Babu, S.S. Assessment of groundwater vulnerability in Krishnagiri District, Tamil Nadu, India using DRASTIC approach. *Int. J. Innov. Res. Sci. Eng. Technol.* **2014**, *3*, 10544–10553.
2. EPA. What Are the Trends in the Extent and Condition of Ground Water and Their Effects on Human Health and the Environment? United States Environmental Protection Agency. Available online: https://www.epa.gov/report-environment/ground-water (accessed on 16 April 2020).
3. Al-Muttair, F.F.; Al-Turbak, A.S. Modeling of infiltration from an artificial recharge basin with a decreasing ponded depth. *J. King Saud Univ. Eng. Sci.* **1991**, *3*, 89–99. [CrossRef]
4. Valhondo, C.; Carrera, J. Water as a finite resource: From historical accomplishments to emerging challenges and artificial recharge. *Sustain. Water Wastewater Process.* **2019**, 1–17. [CrossRef]
5. De Silva, R.B. Limitations of Soil Water Budgeting Method in estimating Groundwater Recharge with particular reference to the dry zone in Sri Lanka. *J. Natl. Sci. Found. Sri Lanka* **2004**, *33*. [CrossRef]
6. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
7. Jang, J. ANFIS: Adaptive network-based fuzzy inference system. *IEEE Trans. Syst. Man. Cybern.* **1993**, *23*, 665–685. [CrossRef]
8. Rustum, R.; Adeloye, A.J. Replacing outliers and missing values from activated sludge data using Kohonen self-organizing map. *J. Environ. Eng.* **2007**, *133*, 909–916. [CrossRef]
9. Rustum, R. Modelling Activated Sludge Wastewater Treatment Plants Using Artificial Intelligence Techniques (Fuzzy Logic and Neural Networks). Ph.D. Thesis, Heriot-Watt University, Edinburgh, UK, 2009. Available online: https://www.ros.hw.ac.uk/handle/10399/2207 (accessed on 16 April 2020).
10. Adeloye, A.; Rustum, R. *Kohonen Self-Organizing Map as a Software Sensor Estimator of Reference Crop Evapotranspiration*; IAHS-AISH Press: Wallingford, UK, 2011; Volume 344, pp. 7–14. Available online: https://iahs.info/uploads/dms/16754.05-7-14-344-37-Adeloye_J-H02_Proof_annotated.pdf (accessed on 20 April 2020).
11. EBOF. Environment Bay of Plenty. 2015. Available online: https://www.boprc.govt.nz/ (accessed on 20 February 2020).
12. Rijkse, W.C.; Guinto, D.F. *Soils of the Bay of Plenty Volume 1: Western Bay of Plenty*; Environment Bay of Plenty Regional Council: Whakatane, New Zealand, 2010.
13. Freeman, J. *Groundwater Recharge at the Kaharoa Rainfall Recharge Site—Rotorua*; Environment Bay of Plenty Regional Council: Ratorua, New Zealand, 2010. Available online: https://www.boprc.govt.nz/media/78668/environmental_publication_201021_groundwater_recharge_at_the_kaharoa_rainfall_recharge_site_rotorua.pdf (accessed on 29 May 2020).
14. Meissner, R.; Rupp, H.; Haselow, L. Use of lysimeters for monitoring soil water balance parameters and nutrient leaching. In *Climate Change and Soil Interactions*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 171–205. [CrossRef]
15. Mamdani, E.H. Application of fuzzy algorithms for control of simple dynamic plant. In *Proceedings of the Institution of Electrical Engineers*; IEEE: New York, NY, USA, 1974; Volume 121, pp. 1585–1588. [CrossRef]
16. Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **1985**, 116–132. [CrossRef]
17. Walia, N.; Singh, H.; Sharma, A. ANFIS: Adaptive neuro-fuzzy inference system—A survey. *Int. J. Comp. Appl.* **2015**, *123*, 32–38. [CrossRef]
18. Azimi, H.; Shabanlou, S.; Ebtehaj, I.; Bonakdari, H.; Kardar, S. Combination of Computational Fluid Dynamics, Adaptive Neuro-Fuzzy Inference System and Genetic Algorithm for Predicting Discharge Coefficient of Rectangular Side Orifices. *J. Irrig. Drain. Eng.* **2017**. [CrossRef]

19. Umamaheswari, G.R.; Kalamani, D. Adaptive neuro fuzzy inference system for monthly groundwater level prediction in Amaravathi river minor basin. *J. Theor. Appl. Inf. Technol.* **2014**, *68*, 8.

20. Bouwer, H. Artificial recharge of groundwater: Hydrogeology and engineering. *Hydrogeol. J.* **2002**, *10*, 121–142. [CrossRef]

21. Raghavendra, N.S.; Deka, P.C. Multistep ahead groundwater level time-series forecasting using Gaussian process regression and ANFIS. In *Advanced Computing and Systems for Security*; Springer: New Delhi, India, 2016; Volume 396, pp. 289–302. [CrossRef]