Civil Engineering and Architecture 10(3): 784-799, 2022 http://www.hrpub.org
DOI: 10.13189/cea.2022.100304

Groundwater Level Forecasting Using Multiple Linear Regression and Artificial Neural Network Approaches

Putu Doddy Heka Ardana[1,*], I Wayan Redana[2], Mawiti Infantri Yekti[2], I Nengah Simpen[3]

[1]Doctor of Engineering Study Program, Udayana University, Denpasar, 80361, Bali, Indonesia
[2]Department of Civil Engineering, Faculty of Engineering, Udayana University, Denpasar, 80361, Bali, Indonesia
[3]Department of Physics, Faculty of Mathematics and Science, Udayana University Denpasar, 80361, Bali, Indonesia

Received November 23, 2021; Revised February 25, 2022; Accepted March 15, 2022

Cite This Paper in the following Citation Styles

***(a):** [1] Putu Doddy Heka Ardana, I Wayan Redana, Mawiti Infantri Yekti, I Nengah Simpen, "Groundwater Level Forecasting Using Multiple Linear Regression and Artificial Neural Network Approaches," Civil Engineering and Architecture, Vol. 10, No. 3, pp. 784-799, 2022. DOI: 10.13189/cea.2022.100304.*

***(b):** Putu Doddy Heka Ardana, I Wayan Redana, Mawiti Infantri Yekti, I Nengah Simpen (2022). Groundwater Level Forecasting Using Multiple Linear Regression and Artificial Neural Network Approaches. Civil Engineering and Architecture, 10(3), 784-799. DOI: 10.13189/cea.2022.100304.*

Abstract Accurate and reliable groundwater level prediction is a critical component in water resources management. This paper developed two methods to predict forty-six months of groundwater level fluctuation. The approaches of Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) were compared for predicting groundwater levels. MLR and ANN approaches were performed at two monitoring wells, Ubung and Ngurah Rai, in the Denpasar region of Bali, Indonesia, considering all significant inputs of hydrometeorological time series data: barometric pressure, evaporation, temperature, wind, bright sunshine, rainfall, and groundwater level. The model's performance was assessed statistically and graphically. The ANN-predicted groundwater levels agreed better with the observed groundwater levels than the MLR-predicted groundwater levels at all sites. The results show the ANN performs better than MLR in terms of statistical errors, notably mean square error (MSE) value of 0.6325; root mean square error (RMSE) value of 0.7953; mean absolute error (MAE) value of 0.6122 based on the MLR in the Ubung monitoring well, while ANN models got an MSE value of 0.143; RMSE value of 0.379, and MAE value of 0.311. For the Ngurah Rai monitoring well, the MSE value is of 1.3406, RMSE value of 1.1579, and MAE value of 0.9152 for MLR, while ANN models obtained MSE value of 0.0483, RMSE value of 0.2198, and MAE value of 0.1266.

Keywords Groundwater Level, Prediction, Hydrometeorological, Multiple Regression Linear, Artificial Neural Network

1. Introduction

Groundwater is a significant source of supplies for residential, industrial, and agricultural use. In certain areas, groundwater is the only reliable source of supplies. While in others, it is preferred due to its near-ubiquitous nature. However, with increased urbanization and water consumption, groundwater has been overexploited, causing adverse environmental consequences for instance significant water level decreases, well desiccating, stream and lake shrinkage, reduced well yields, and water quality degradation, especially in developing countries [1, 2] including in Denpasar City as an urban area in Bali Province, Indonesia. Groundwater level forecasting is critical for sustainable groundwater management [3]. Although complex and nonlinear, the groundwater level is affected by various elements of hydrometeorology such as barometric pressure, evaporation, temperature, wind, bright sunshine, and rainfall. So establishing accurate models to estimate groundwater levels is critical [4]. Modeling groundwater fluctuations is intricate because groundwater is concealed and has considerable temporal and spatial variability. Groundwater flow modeling methods exist for several hydrogeological conditions. Data for process-based models that impersonate groundwater changes are immense, complicated, or costly to gather, combined with limited field data [5, 6].

The principal source of information on hydrological pressures acting on the aquifer is groundwater level readings from observation wells. For long-term groundwater management and protection, systematic water level observations will offer crucial data needed to assess changes in groundwater resources, develop groundwater trend models and forecasts, design, implement, and monitor programs [7]. Groundwater fluctuations are the rise and decrease of groundwater levels caused by natural and human-induced hydrological processes. Understanding these events is critical because multiple mechanisms can function simultaneously, requiring accurate observations. Urbanization, seismicity, hydrometeorology (such as barometric pressure, evaporation, temperature, wind, bright sunshine, and rainfall), tidal influences, and external stress are all factors that induce groundwater level variations [8].

Predicting groundwater level reactions is crucial for effective groundwater planning and management. These strategies have been developed to prevent groundwater mismanagement and overexploitation. Simulating groundwater level fluctuations is difficult due to the complexity and non-linearity. Conceptual and process-based methods exist for modeling groundwater flow in various hydrogeological settings. The data requirements for process-based models used to simulate groundwater changes are vast and generally difficult or costly to collect [5][6]. Despite tremendous efforts and resources, distributed numerical flow models' prediction accuracy has not improved enough for diverse water management challenges [9]. It is desirable to have a dynamic prediction model for handling persistent trends and time-variant behavior. Such instances favor empirical models like regression and artificial neural network (ANN) models that require fewer data and are thus less expensive. Despite its inability to manage non-linearity between model inputs and outputs, multiple regression linear (MLR) models are commonly used in hydrological research. [10][11].

Some hydrologists or hydrogeologists have been interested in using Artificial Neural Network (ANN) tools and statistical techniques such as Multiple Linear Regression (MLR) for predicting or forecasting water resources systems over the last decade because of their simplicity and profitability [12]. MLR models can show how the correlation between observation and response variables works by adjusting a linear equation to the data that has been collected. [13] and can produce valuable findings with less data, less work, and cost-effectiveness [12]. It also allows for unlimited independent variables. Despite their incapacity to handle non-linearity between model inputs and outputs, MLR models have been widely used in hydrological research due to their ease of use and parameter interpretation [11]. However, the ANN approach is well adapted to modeling non-linear and dynamic systems like water resources. The fundamental advantage of ANN over previous techniques is that it does not necessitate a detailed mathematical description of underlying processes. After adequate training, ANN models can successfully anticipate various hydrological problems.

Limited studies about MLR application in groundwater level forecasting were reported. Hodgson [14] used MLR to predict water table responses in the South African Vryburg aquifer using precipitation and pumping as input factors. Shao and Campbell [15] utilized regression to model groundwater trends in Western Australia.

The ASCE Task Committee findings contain an in-depth examination of the application of ANN to hydrology [16, 17]. ANN has effectively predicted groundwater levels in unconfined aquifers [18–24]. The networks were provided monthly water depth, precipitation, temperature, river water level, and evapotranspiration. Uddameri [25] employed regression and artificial neural network (ANN) approaches to predict piezometric levels in a deep well in South Texas and Sahoo and Jha [26] compared MLR and ANN for simulating transient groundwater levels in an unconfined aquifer system.

According to our knowledge and research, no previous research has compared the predictive ability of the MLR and ANN techniques in simulating groundwater levels using limited hydrometeorological time series data (barometric pressure, evaporation, temperature, wind, bright sunshine, and rainfall) with data screening tests. A trend absence test, stationary, persistence, outlier, and data consistency test are all examples of filtering tests. As a result, the goal of this research is to see how well two data-driven techniques, such as MLR and ANN, can forecast the spatio-temporal distribution of water levels in groundwater basins utilizing restricted hydrometeorological time-series data that have already been filtered. The usual MLR and ANN modeling techniques were closely followed in this study, and selected hydrometeorological data were used as model inputs. As a result, this research presents a rigorous scientific technique for comparing two data-driven methodologies (modeling tools) for simulating groundwater levels using filtered hydrometeorological data.

2. Materials and Methods

2.1. Study Area

The research area is 31,42 km^2 in size and lies between 08°35'31" and 08°44'49" south latitude and 115°12'09" and 115°04'39" east longitude in Denpasar, Bali, Indonesia [27]. Denpasar's aquifer, particularly in the north, is an unconfined aquifer with a shallow groundwater level that runs through fissures and crevices between grains and is a highly productive aquifer [28]. The Denpasar-Tabanan groundwater basin includes this aquifer [29]. Denpasar is a volcanic-sediment-covered terrain, and alluvium and young volcanic sediments are often extremely permeable, whereas lower quaternary and tertiary sediments have a wide range of permeability according to the formation. Denpasar is made up of Miocene to Pliocene volcanic products and marine sediment as basement rock, which is overlain by a thick pyroclastic flow, volcanic products, and volcanic mudflow that resulted from intense volcanic activity during the Pleistocene to Holocene periods of the Quaternary period [30]. Figure 1 shows the location of the study area.

2.2. Data Collecting

General groundwater data on geography, geology, topography, and hydrogeology was provided by the Bali Province Department of Manpower, Energy, and Mineral Resources. Hydrometeorological data was also provided by the Bali-Penida River Basin Department and the Meteorological, Climatological, and Geophysical Agency III Bali Province. The Polygon Thiessen method was used to convert point precipitation data to area precipitation.

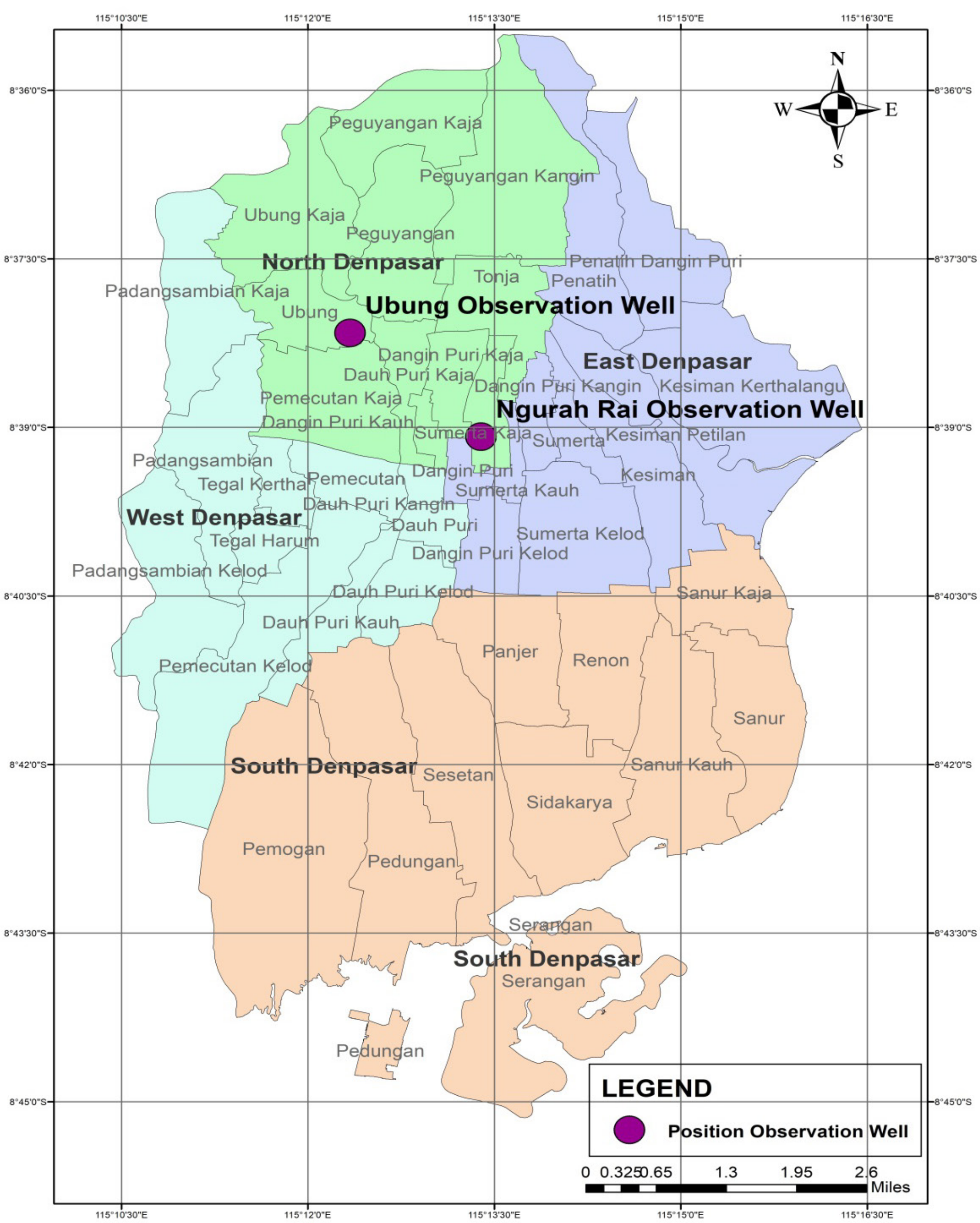

Figure 1. Location of study

The hydrometeorological data are adjusted to account for the fact that groundwater level data is only available for forty-six months. The study used hydrometeorological such as barometric pressure, evaporation, temperature, wind, bright sunshine, rainfall, and water table data from January 2017 to December 2019 and January to October 2015. Monthly hydrometeorological and groundwater level data for the 3 years period 2017–2019 were used for training (calibration) of the two observation wells for MLR and ANN models. The ten months period from January to October 2015 is used for testing (verification). Table 1 shows the location of each well.

Table 1. Position of observation well

	Well Number	Coordinate
Ubung	SP No. 04/DP/Distam	08° 39’ 9,5” LS; 115° 12’ 20,4” BT
Ngurah Rai	SP No. 02/DP/Distam	08° 39’ 05,0” LS; 115° 13’ 23,6” BT

2.3. Hydrometeorology Data Testing

Hydrometeorological data is a sequence of time sequence data that must be tested before being used in the analysis. Hydrometeorological data in this study are data on rainfall, evaporation, humidity, bright sunshine, temperature, wind speed, and air pressure coupled with data on groundwater level fluctuations. This testing phase can also be called data screening, to examine and sort or group data to obtain hydrometeorological data reliable enough for analysis so that the conclusions obtained are good enough [31]. The hydrometeorological data test is a consistency test, trend absence test, outlier test, stationary, and persistence test.

2.3.1. Consistency Test

A data consistency test is required to see the principality in a series of data obtained. The method used in this test is the Rescaled Adjusted Partial Sums (RAPS) method. This method could be used to investigate the variance of time series data trends and locate trend inflection points, shifts, data clustering, irregular fluctuations, and periodicities [32–34]. The method's benefit is that it eliminates the effects of different data units and random errors on analysis [34].

2.3.2. Trend Absence Test

This test is done to determine the randomness or absence of trends from periodic series data. To test the lack of trends used the Spearman Method's Statistical Correlation Ranking Method. This approach is a correlation between time and variant from a table of hydrological variables [31].

2.3.3. Outlier Test

An abnormality test or outlier test is used to determine whether the maximum and minimum data from an existing data set is worth using or not [35]. This test is based on data deviating from two thresholds, namely the lower and upper threshold, to be eliminated or adjusted to the threshold value.

2.3.4. Stationary Test

The stability of variant values and averages of a time series in this study is hydrometeorological data that can be seen with stationary tests. This study will conduct a stationary test with variant stability test (F-Test) and average stability test (t-Test). Suppose the calculated value is greater than the critical value. In that case, the data tested does not come from the same population or is not stationary at a certain significance level. Variant values are unstable and un homogeneous if test results show the null hypothesis is rejected [31].

2.3.5. Persistence Test

Persistence tests in this study are used as a requirement in frequency analysis by testing the absence or absence of dependence on each data. If there is no dependency on each value, then the data can be used in frequency analysis. The magnitude of the correlation coefficient needs to be taken into account [31].

2.4. Modelling

Based on the hydrometeorological data testing, new variables were identified. These variables were used in modeling using multiple linear regression (MLR) and Artificial Neural Network (ANN).

2.4.1. Multiple Linear Regression (MLR)

Multiple linear regression (MLR) expresses the linear connection between a dependent variable and several independent variables [14][36]. MLR uses least squares to fit the model, minimizing the sum of squares of observed and predicted values. MLR can be expressed as (1)

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + e \quad (1)$$

where Y means the dependent variable, Xi means independent variables, βi means predicted parameters, and ε is the error term.

2.4.2. Classic Assumption Test

Determining the value of Y (independent variable) to estimate the value associated with X as a dependent variable is necessary for estimating cause-and-effect relationships. So, the essence of regression analysis is to explain and test the relationship between one or more independent variables into one dependent variable. Multiple regression methods can be an unbiased estimation tool if it meets Best Linear Unbiased Estimation (BLUE) requirements. The first classic assumption test is done before the hypothetical test meet BLUE needs. Classical assumptions will comprise multicollinearity test, normality test, autocorrelation test and, heteroscedasticity test [37].

The normality test checks whether dependent and independent variables in regression have a normal distribution. A good regression model is data that is normally distributed or close to normal. Multicollinearity arises when all or some of the independent variables in a regression model are perfectly linear. The multicollinearity test checks whether two or more independent variables are highly correlated in a regression model. This means that an independent variable can be predicted from another independent variable in a regression model. A decent regression model does not correlate with the variables. Durbin Watson statistical tests to find a serial correlation (autocorrelation) in time series data. Serial correlation is a situation where there is a relationship or correlation between two observations for a variable. The heteroscedasticity test checks whether the residual variance of one observation differs from another in the gradient model. The variance of one residual observation remains the same, while the variance is heteroscedastic. A decent regression model has homoscedasticity [37].

2.4.3. Artificial Neural Network (ANN)

An ANN is a massively parallel distributed information processing system like biological neural networks [38]. The architecture of a neural network represents the pattern of connections between nodes and the activation function [39]. An ANN comprises simple, highly interconnected processing components, like neurons. ANN model is a black box of equations that calculate output based on input values [40]. According to Haykin [38], we may mathematically describe a neuron k as follows (2) and (3)

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{2}$$

$$Y_k = \varphi(u_k + b_k) \tag{3}$$

Bias, meant by b_k, has the effect of increasing or lowering the net input of the activation function. $x_1, x_2, ..., x_m$ are the inputs; $w_{k1}, w_{k2}, ..., w_{km}$ are the weights of the neuron k; u_k is the linear combiner output due to input signals; φ is the activation function; y_k is the output signal of the neuron.

Backpropagation is a popular ANN learning algorithm in multilayered feedforward networks. Backpropagation networks process data from input to hidden layer to output layer. Finding optimal weights is the goal to get close to targets [38].

Feedforward backpropagation neural network (FFBPNN) architecture was used in this research, along with gradient descent with momentum and adaptive learning rate backpropagation (traingdx) for training algorithms, to find the best algorithm for predicting groundwater levels over the study field. In the hidden layer, logistic sigmoid non-linear function (logsig) and output layer, linear transfer function (purelin) was used as an activation function.

2.4.4. Model Performance

The quantitative performance of MLR and ANN models was judged by using four statistical metrics (goodness-of-fit criteria): coefficient determination (R^2), root mean squared error (RMSE), mean squared error (MSE), and mean absolute error (MAE).

3. Results and Discussion

3.1. Data Quality Test

Before the analysis to get a model of groundwater level fluctuations, available meteorological data need to be tested statistically hydrologically [31]. In this study, data quality testing used outlier tests or abnormality tests, trend absence tests, persistence tests, stationary tests, and consistency tests. The test results can be seen in Table 2. Based on the data quality test, three data are eliminated: humidity data, bright sunshine, and wind speed data that are not used in subsequent analysis.

Generation or prediction of groundwater level fluctuation data extend the data of groundwater level fluctuations using data on groundwater level fluctuations at Ubung and Ngurah Rai monitoring wells. The use of the two monitoring wells is based on the position of the appropriate well to represent fluctuations in groundwater levels in the area of groundwater addition in the Denpasar city aquifer. The best model for predicting groundwater level fluctuations used in this study is comparing Multiple Linear Regression and Artificial Neural Networks.

Table 2. Hydrometeorology Data Quality Testing Recapitulation

Data	Data Quality Test					
	Outlier	**Consistency**	**Trend Absence**	**Persistence**	**Stationary**	**Information**
Evaporation (E)	No outlier	Consistent	Independent	Independent	Stable	Ok
Barometric pressure (BP)	No outlier	Consistent	Independent	Independent	Stable	Ok
Temperatures (T)	No outlier	Consistent	Independent	Independent	Stable	Ok
Humidity (H)	No outlier	Consistent	**Dependent**	Independent	**Unstable**	**Not Ok**
Wind speed (WS)	No outlier	Consistent	Independent	Independent	**Unstable**	**Not Ok**
Bright sunshine (BS)	No outlier	Consistent	**Dependent**	Independent	Stable	**Not Ok**
Rainfall (rain gauge Ngurah Rai)	No outlier	Consistent	Independent	Independent	Stable	Ok
Rainfall (rain gauge Sanglah)	No outlier	Consistent	Independent	Independent	Stable	Ok
Rainfall (rain gauge Sumerta)	No outlier	Consistent	Independent	Independent	Stable	Ok
Rainfall (rain gauge Kapal)	No outlier	Consistent	Independent	Independent	Stable	Ok
Rainfall (rain gauge Buagan)	No outlier	Consistent	Independent	Independent	Stable	Ok
Rainfall (rain gauge Sading)	No outlier	Consistent	Independent	Independent	Stable	Ok
Rainfall (rain gauge Penatih)	No outlier	Consistent	Independent	Independent	Stable	Ok

3.2. Multiple Linier Regression

The prediction model with the Multiple Linear Regression approach produces the following equations:

Ubung monitoring well equation model:

$$\begin{aligned} \text{GWL} = {} & -592318,829 + 587,063\text{BP} \\ & +33,579\text{E} + 467,869\text{T} + 5,406\text{P} + \varepsilon \end{aligned} \tag{4}$$

Model of equation of monitoring well Ngurah Rai:

$$\begin{aligned} \text{GWL} = {} & -263447,741 + 268,001\text{xBP} \\ & -3,916\text{xE} - 4,953\text{xT} + 9,958\text{xP} + \varepsilon \end{aligned} \tag{5}$$

Where, GWL = groundwater level, BP = barometric pressure, E = evaporation, T = temperature, P = precipitation.

Both models were obtained through several stages of analysis, namely having qualified data normality for three years (all normally distributed parameters) based on the Kolmogorov-Smirnov test and the Shapiro-Wilk test where all data has a p-value > 0.05 value. The accuracy of the regression function in estimating the actual value can be measured from its goodness of fit. Statistically, at least, this can be calculated from the value of the coefficient of determination (R^2), the statistical value F, and the statistical value t.

Regression models at Ubung monitor wells produce a coefficient of determination (R^2) value of 0.606 which means that 60.6% of groundwater level can be explained by parameters of barometric pressure (BP), evaporation (E), temperature (T), and precipitation (P). In contrast, the rest is explained by other variables, which are estimated to be due to the exploitation of groundwater by the community. Based on simultaneous tests, obtained value $F_{count} = 11.905$ with probability (Sig.) 0.000 (< 0.05) which indicates that there is an influence or contribution between variables of barometric pressure, evaporation, temperature, and precipitation simultaneously and significantly to fluctuations in groundwater levels so that regression models on Ubung monitoring wells can be used to predict changes in groundwater levels. Based on partial or individual tests, the p-value of precipitation is worth 0.361 (0.361 > 0.05), which means no significant relationship between precipitation parameters and fluctuating groundwater levels. In contrast, other parameters (barometric pressure, evaporation, and temperature) have a partially meaningful relationship to fluctuating groundwater levels. Linear regression models are called good models when they meet some assumptions, better known as classical assumptions. The classic assumptions that must be met are normally distributed residuals, no symptoms of multicollinearity, no heteroskedasticity, and no signs of autocorrelation. The classical assumption test analysis results showed that the regression model for the Ubung monitor well met the entire classical assumption test. In general, the groundwater level fluctuation model could be used with the unbiased estimation. To test the regression model that has been built, a calibration test based on data for 36 months (2017-2019) and a verification test for ten months data (January - October 2015) with statistical parameters Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE). An MSE value of 0.6325 is obtained, RMSE of 0.7953, and MAE of 0.6122. The smaller MSE, RMSE, and MAE value shows a good predictive value in the calibration process. Unlike the verification process where the MSE value is 1.6415; RMSE of 1.2812; MAE value of

0.8384 which means between the observation GWL value and the prediction GWL has a relatively high error rate. Figure 2 shows the comparison between the GWL observation and prediction at the calibration stage and Figure 3 at the verification stage in the Ubung observation well.

The regression model at Ngurah Rai monitoring well produces a coefficient of determination (R^2) value of 0.257. Only 25.7% of groundwater level can be explained by barometric pressure, evaporation, temperature, and precipitation. In comparison, the rest (74.3%) is defined by other variables. In this case, it is estimated that the exploitation of groundwater by the community where the position of aquifers in the Ngurah Rai area is relatively shallow, namely 15-20 meters below the face soil. Based on simultaneous tests, obtained value F_{count} as much as 2.685 with a more negligible probability equal to 0.05, which indicates that there is an influence or contribution between variables of barometric pressure, evaporation, temperature, and precipitation simultaneously and significantly to fluctuations in groundwater levels so that regression models on Ngurah Rai monitoring wells can be used to predict changes in groundwater levels. Based on the partial test or individual test (t-test), the p-value of barometric pressure and precipitation is more significant than 0.05 (p-value. > 0.05), which means there is a significant partial relationship between barometric pressure and precipitation variables with fluctuating groundwater levels. In contrast, other variables (evaporation and temperature) do not have a partially meaningful relationship to fluctuating groundwater levels. The analysis of classical assumption tests on the Ngurah Rai monitoring well showed that the model met all classical assumption testings. In general, the GWL model could be used with the unbiased estimation. To test the regression model that has been built, a calibration test is conducted based on data for 36 months (2017-2019) and a verification test for ten months data (January - October 2015) with statistical parameters RMSE, MSE, and MAE. In the calibration process, an MSE value of 0.3740 is obtained, RMSE of 0.6116, and MAE of 0.4717. The smaller the value of MSE, RMSE, and MAE shows a good predictive value in calibration process. Unlike the calibration process where the MSE value is 1.3406; RMSE of 1.1579; MAE value of 0.9152 which means between the observation GWL value and the prediction, GWL has a relatively high error rate. Figure 4 shows the comparison between the GWL observation and prediction at the calibration stage and Figure 5 at the verification stage in the Ngurah Rai observation well.

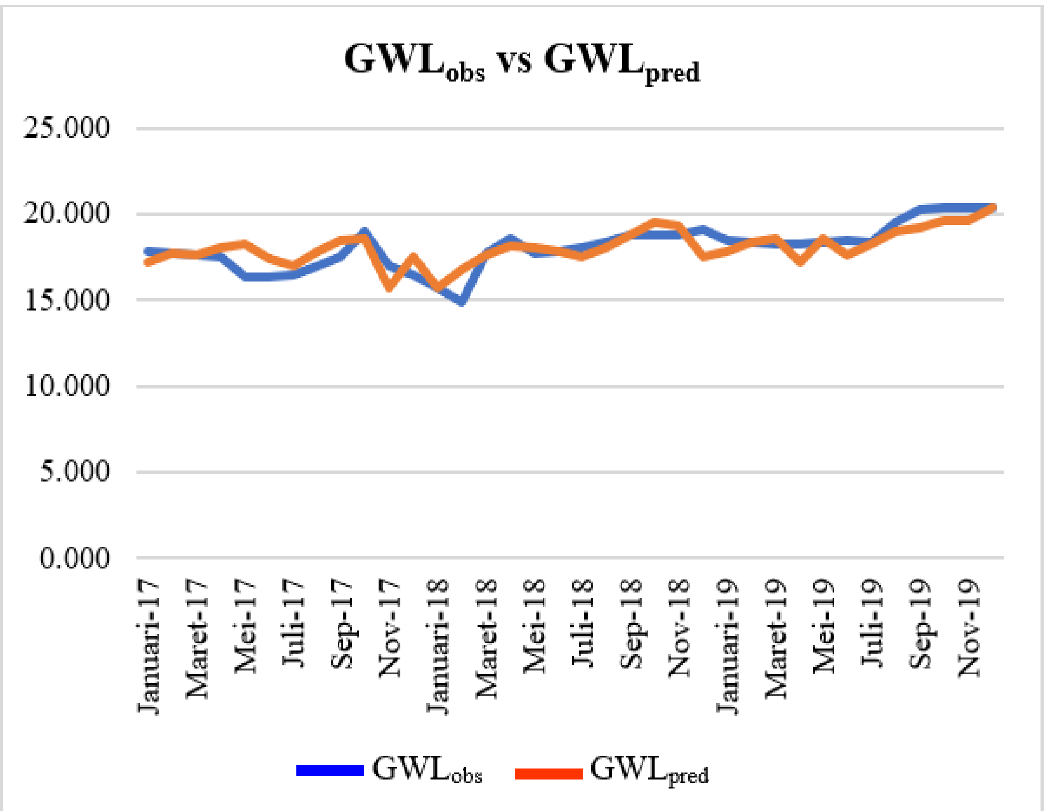

Figure 2. GWL_{obs} and GWL_{pred} in calibration stage (MLR; Ubung)

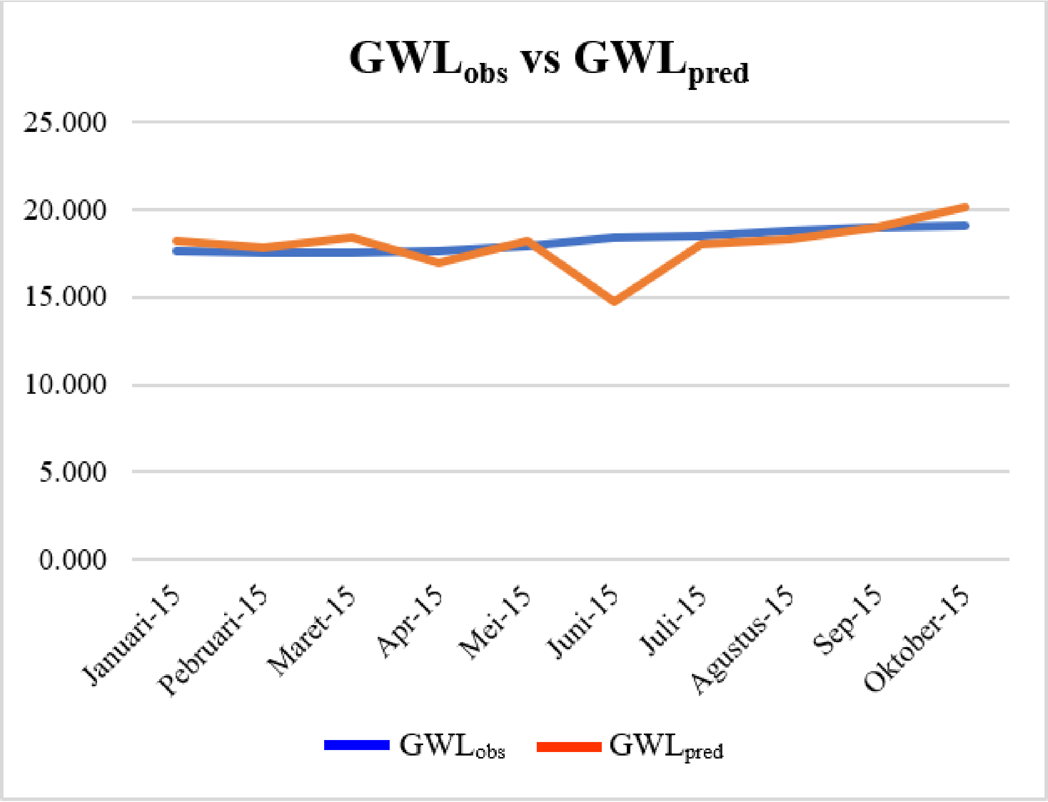

Figure 3. GWLobs and GWLpred in verification stage (MLR; Ubung)

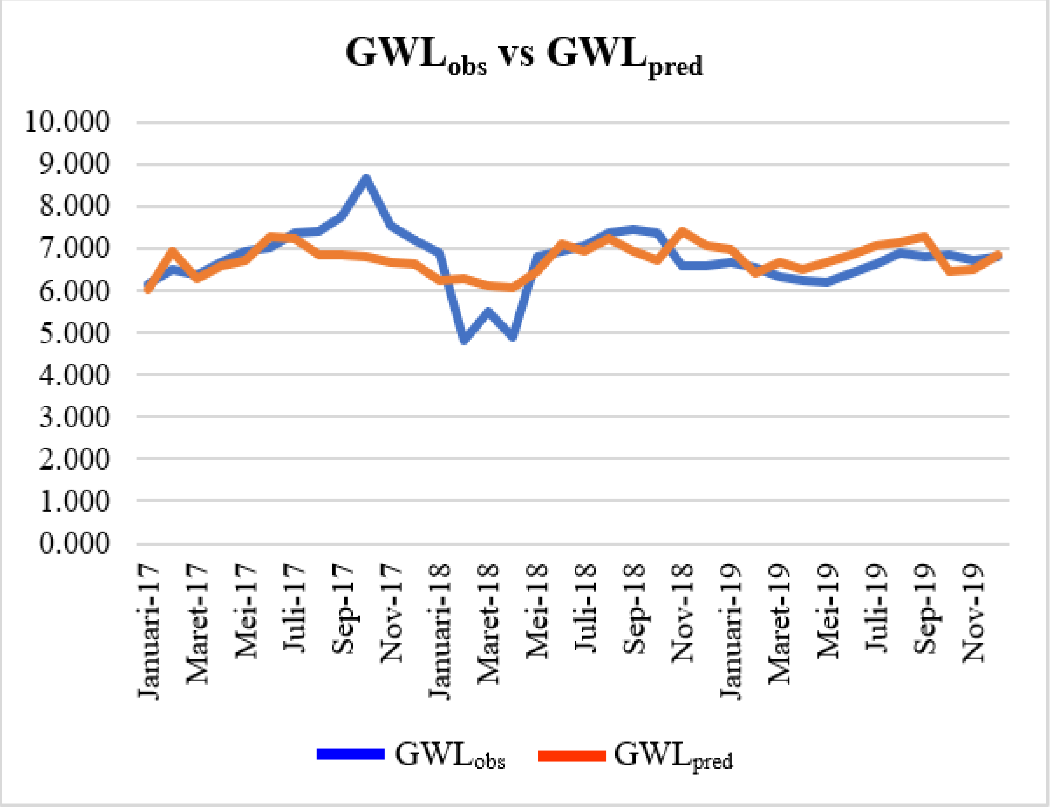

Figure 4. GWL_{obs} and GWL_{pred} in calibration stage (MLR; Ngurah Rai)

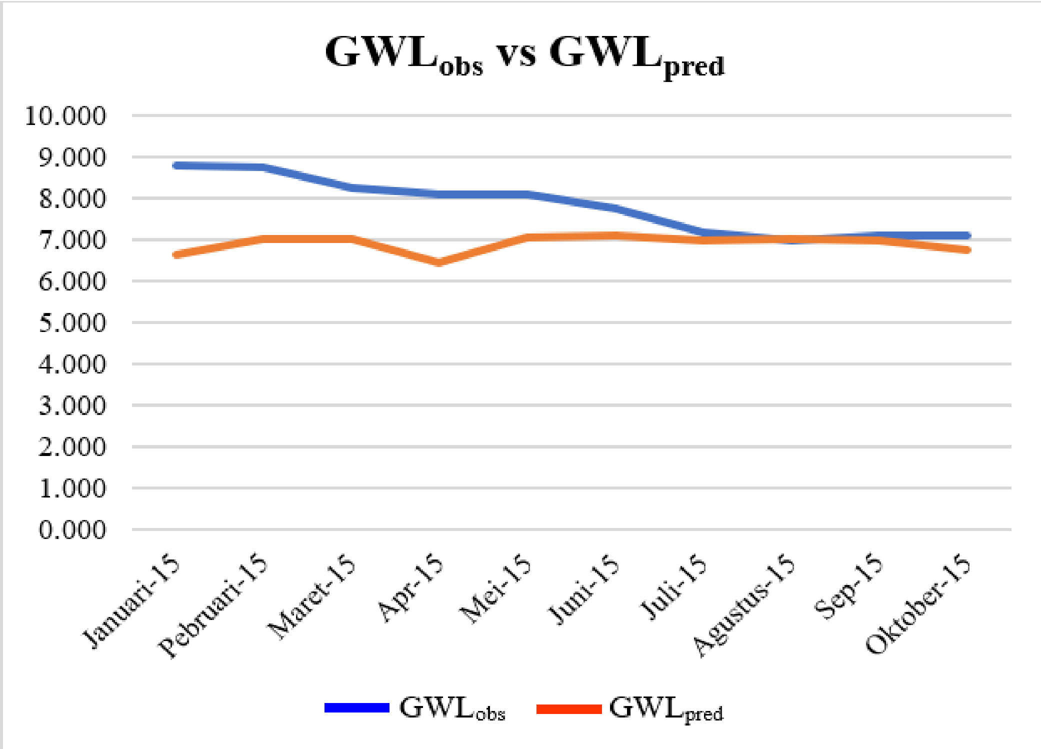

Figure 5. GWL_{obs} and GWL_{pred} in verification stage (MLR; Ngurah Rai)

3.3. Artificial Neural Network

Modeling of groundwater level fluctuations in the aquifer area in Denpasar region using four variations of network architecture, namely 4-4-1 architecture (4 input variables – 4 neurons hidden layer – 1 output variable); architecture 4-8-1 (4 input variables – 8 hidden layer neurons – 1 output variable); architecture 7-7-1 (7 input variables - 7 hidden layer neurons) - 1 variable output); and 7-14-1 (7 input variables – 14 hidden layer neurons – 1 output variable). Variations of network architecture can be seen in Figure 6 and Figure 7.

The selection of input parameters for the ANN model is based on previous studies using barometric pressure, evaporation, temperature, humidity, wind speed, bright sunshine, and groundwater level fluctuation data (GWL), but the input parameters are used partially [42]. In this study, input parameters in the form of hydrometeorological variables, including barometric pressure, evaporation, temperature, humidity, wind speed, and bright sunshine data, will be used simultaneously as variations of network architecture. Another variation is to use hydrometeorological input parameters based on data quality test results where the selected parameters are barometric pressure, evaporation, temperature, and precipitation. The data used are hydrometeorology and GWL data for 36 months (2017 - 2019), calibration and hydrometeorology, and GWL data for ten months (January - October 2015) as verification data.

Modeling groundwater level fluctuations with the ANN approach is done with the help of Matlab R2015a software to facilitate and accelerate analysis to get a prediction model on the Ubung monitoring well and Ngurah Rai monitoring well as an observation well for the groundwater addition area in Denpasar city aquifer. The stages of analysis with the ANN approach are to perform the process of normalization or preprocess data or data transformation following the range of activation functions applied, the process of input data training and data testing into the network architecture (4-4-1; 4-8-1; 7-7-1; and 7-14-1), conducting the analysis process until it obtains the most optimal results based on the mean value of square error. (MSE) network and correlation coefficient value (R) stage training, testing, and validation of the overall variation of the ANN model. Obtaining a different ANN model for each monitoring well, the equation obtained is:

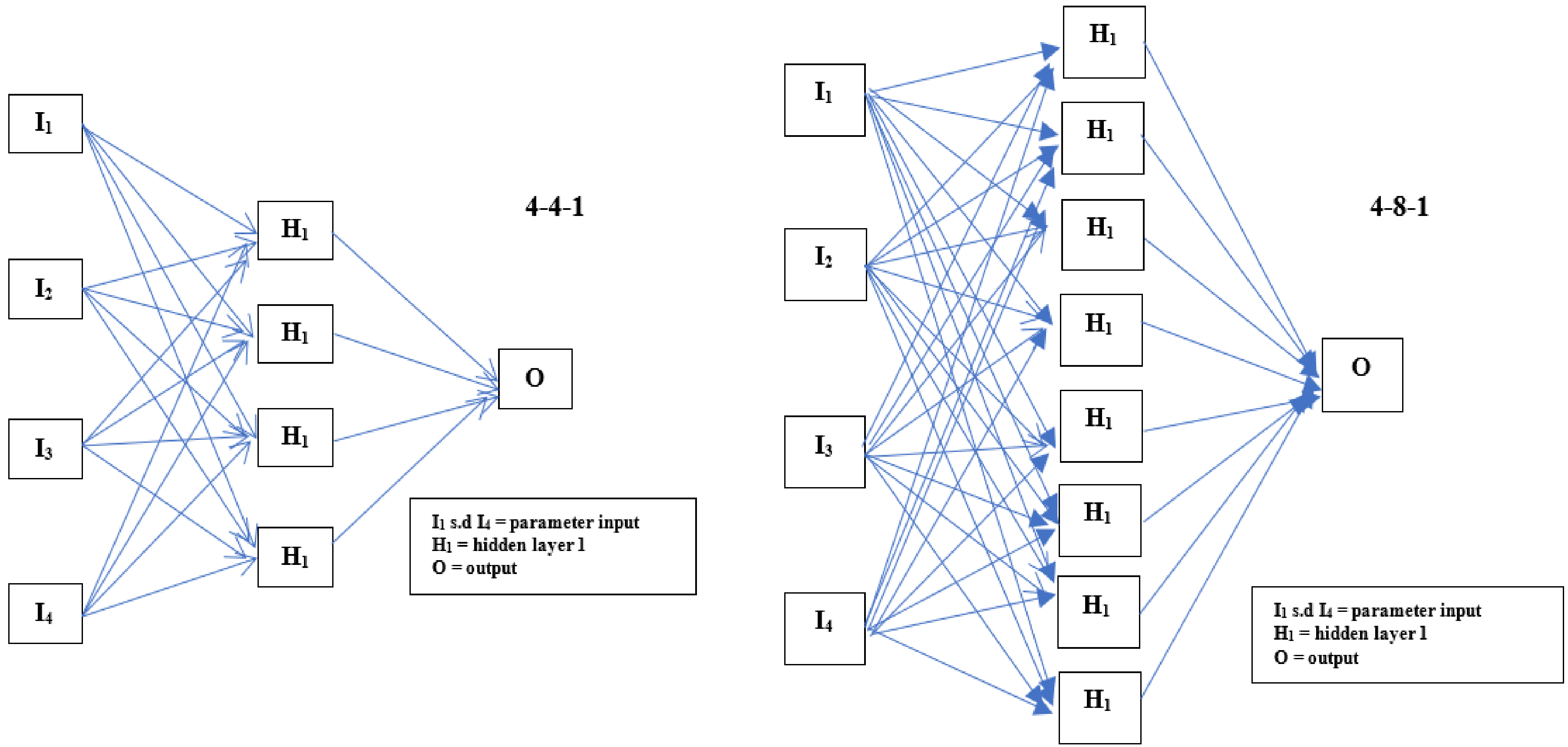

Figure 6. Variations of network architecture (4-4-1 and 4-8-1)

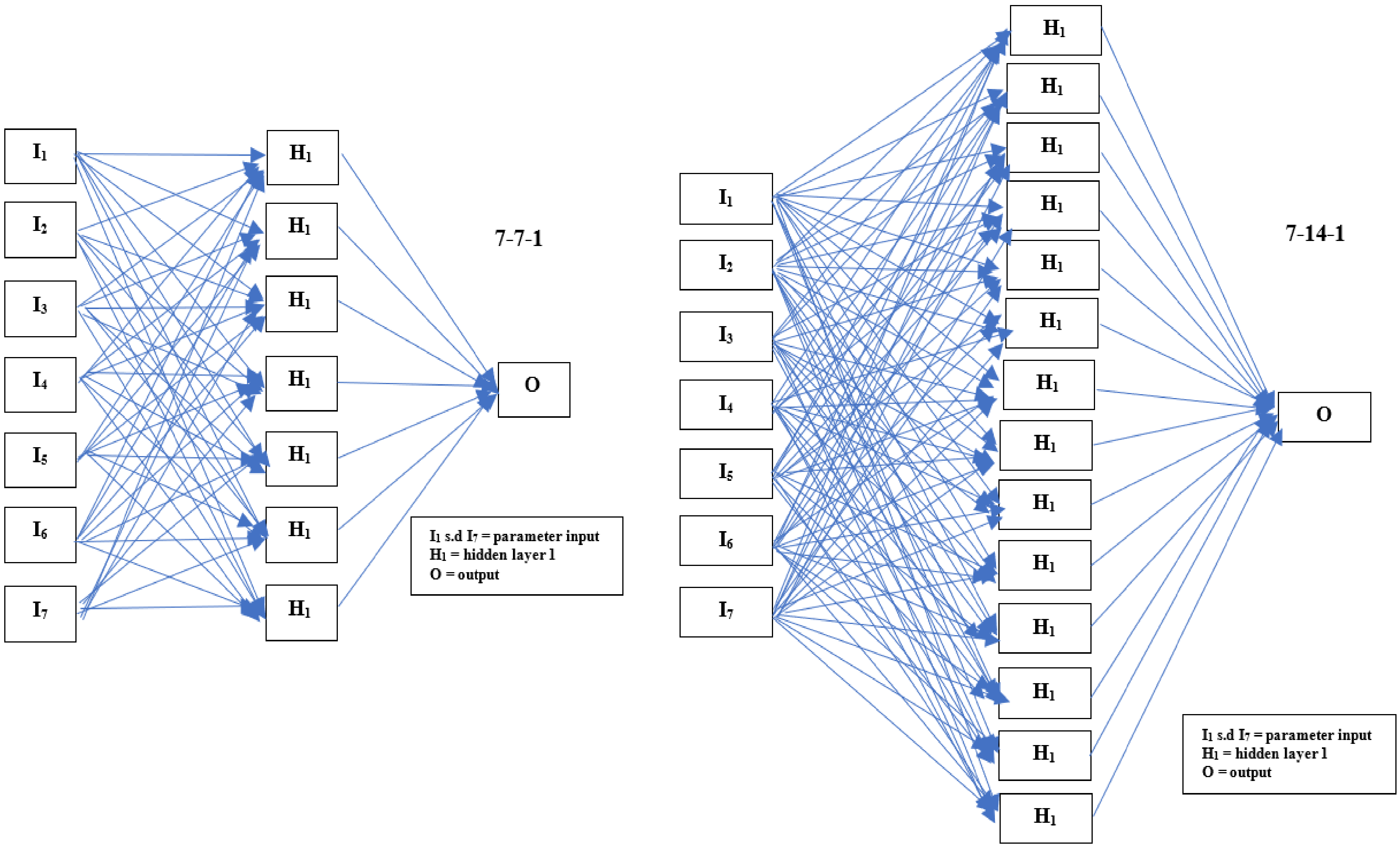

Figure 7. Variations of network architecture (7-7-1 and 7-14-1)

Equation of ANN model of Ubung monitoring well:

$$\text{y\_GWL} = \sum_{c=1}^{C}\sum_{d=1}^{D} W_{2-o}\_cd.(1-\exp(\sum_{a=1}^{A}\sum_{b=1}^{B} W_{i-1}\_ab.X_{tu} + W_{i-1}\_ab.X_{p} + W_{i-1}\_ab.X_{s} + W_{i-1}\_ab.X_{ch} + B_{1}\_b))^{-1} + B_{o}\_d))^{-1} \quad (6)$$

Equation of ANN model Ngurah Rai monitoring well:

$$\text{y\_GWL} = \sum_{c=1}^{C}\sum_{d=1}^{D} W_{2-o}\_cd.(1-\exp(\sum_{a=1}^{A}\sum_{b=1}^{B} W_{i-1}\_ab.X_{tu} + W_{i-1}\_ab.X_{p} + W_{i-1}\_ab.X_{s} + W_{i-1}\_ab.X_{ku} + W_{i-1}\_ab.X_{ka} + W_{i-1}\_ab.X_{lpm} + W_{i-1}\_ab.X_{ch} + B_{1}\_b))^{-1} + B_{o}\_d))^{-1} \quad (7)$$

Where, GWL = groundwater level, X_{mn} = input variable value (barometric pressure, evaporation, temperature, wind, bright sunshine, rainfall, and groundwater level), W_{mn} = weight matrix layer-m to layer-n, B_n = bias layer-n.

The ANN model for Ubung monitoring well uses network architecture of 4-4-1 (4 input variables – 4 hidden layer neurons – 1 output variable). The MSE model value is of 0.0018388 in the 87th epoch with an overall R_{model} value of 0.95493. Based on the results of calibration tests (training) obtained a value of R = 0.955; R^2 = 0.912; MSE = 0.143; RMSE = 0.379; and MAE = 0.311 and based on the results of the verification test (testing) obtained the value R = 0.891; R^2 = 0.794; MSE = 0.129; RMSE = 0.359; and MAE = 0.319. The 7-14-1 network architecture also provides good value in modeling groundwater level fluctuations in Ubung monitor wells. It can be seen from the $R_{training}$ value of 0.9674 and the $R_{testing}$ value of 0.7635, which means that it has good reliability in modeling fluctuations in groundwater levels. These values show the reliability of the ANN model with architecture of 4-4-1 or 7-14-1 as a model for predicting fluctuations in groundwater levels in Ubung monitoring wells. Figure 8 shows the comparison between the GWL observation and prediction at the calibration stage and Figure 9 at the verification stage in the Ubung observation well.

The ANN model for Ngurah Rai monitoring wells uses network architecture of 7-14-1 (7 input variables – 14 hidden layer neurons – 1 output variable). The MSE model value is of 0.0010372 in the zero epoch with an overall model R-value of 0.95568. Based on the results of calibration tests (training) obtained a value of R = 0.9557; R^2 = 0.9133; MSE = 0.0483; RMSE = 0.2198; and MAE = 0.1266 and based on the results of the verification test (testing) obtained the value R = 0.2227; R^2 = 0.0496; MSE = 0.6621; RMSE = 0.8137; and MAE = 0.5985. These values show the ANN model with a 7-14-1 architecture has the ability as a model to predict fluctuations in groundwater levels at the Ngurah Rai monitoring well. Figure 10 compares the GWL observation and prediction at the calibration stage and Figure 11 at the verification stage in the Ngurah Rai observation well.

The relationship between the $GWL_{observation}$ and the $GWL_{prediction}$ the results of the modeling of ANN with the network architecture 4-4-1 at the Ubung monitoring well, resulting in a correlation coefficient (R) of 0.955 and a coefficient of determination (R^2) of 0.912 at the calibration stage (training) and 0.891 coefficient of determination (R^2) of 0.794 at the verification stage (testing), which means the robust correlation. While the relationship between GWL observation and the GWL prediction, the results of the modeling of ANN with the network architecture 7-14-1 at Ngurah Rai monitoring well produces a correlation coefficient (R) of 0.9557 and coefficient of determination (R^2) of 0.9113 at the calibration stage (training). In contrast, the results are different at the verification or testing stage, where a value of 0.2227 is interpreted with a low correlation and a coefficient of determination (R^2) of 0.0496. But the value is still better than other ANN models, so the 7-14-1 architecture is still used in the process of predicting MAT fluctuations in the Ngurah Rai monitoring well.

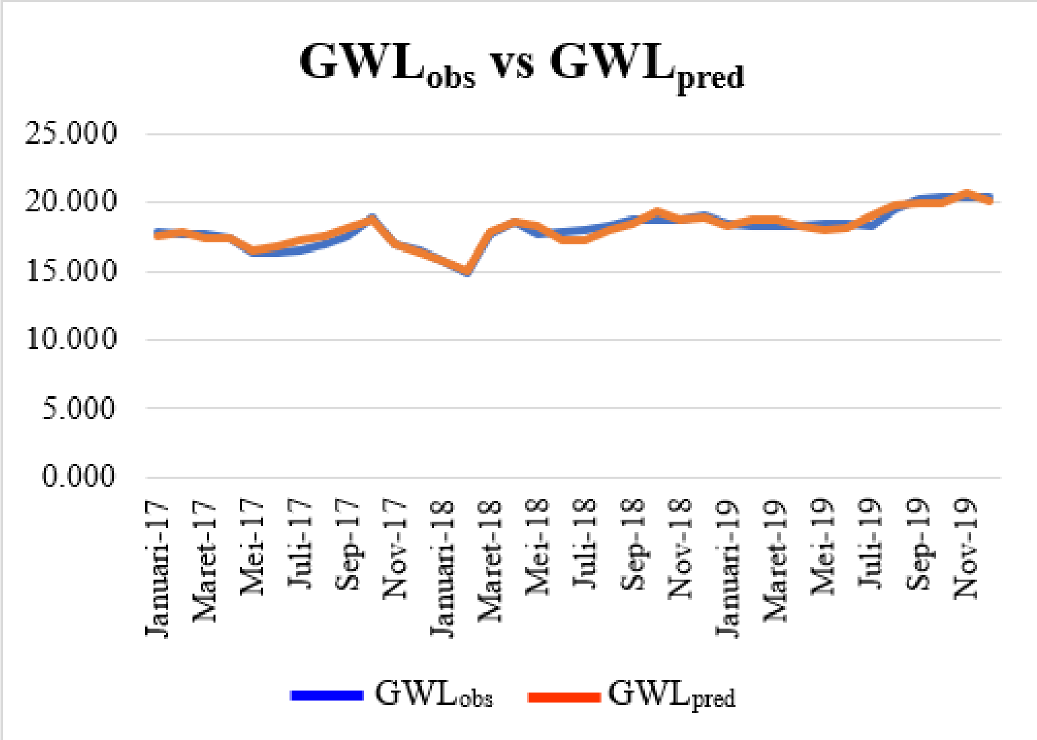

Figure 8. GWL_{obs} and GWL_{pred} in calibration stage (ANN 4-4-1; Ubung)

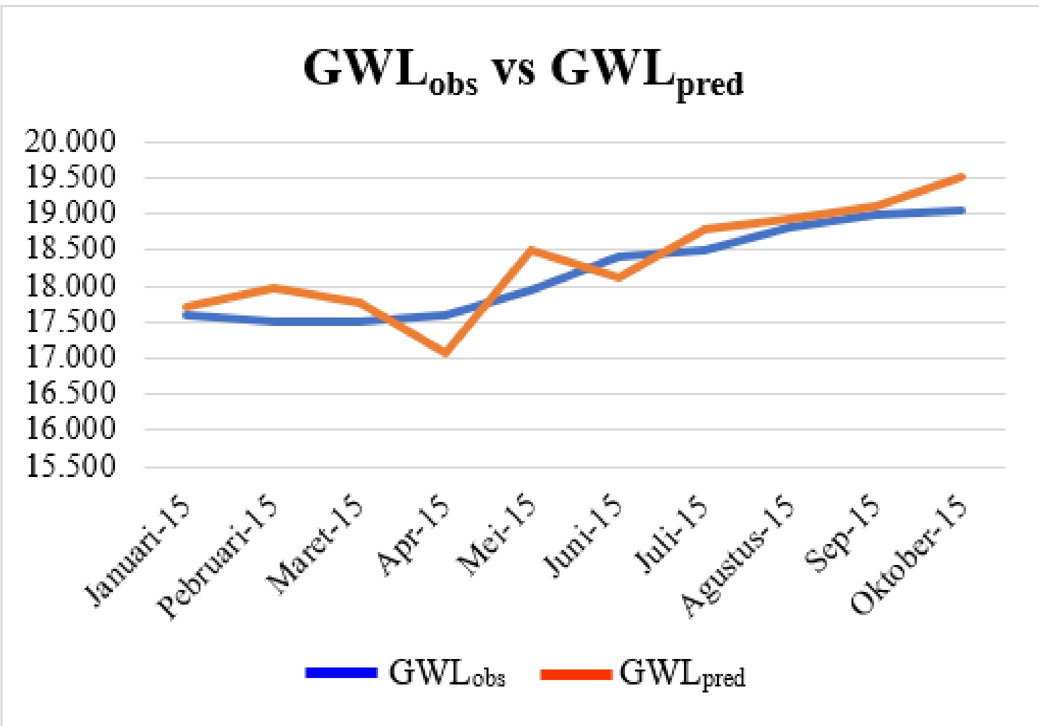

Figure 9. GWL_{obs} and GWL_{pred} in verification stage (ANN 4-4-1; Ubung)

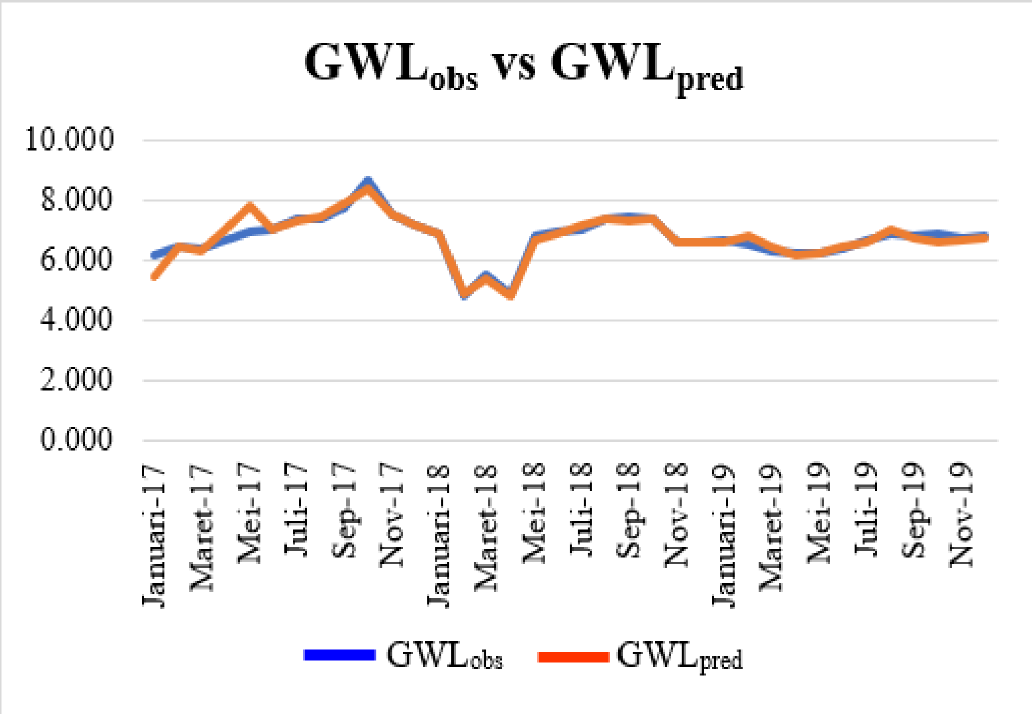

Figure 10. GWL_{obs} and GWL_{pred} in calibration stage (ANN 7-14-1; Ngurah Rai)

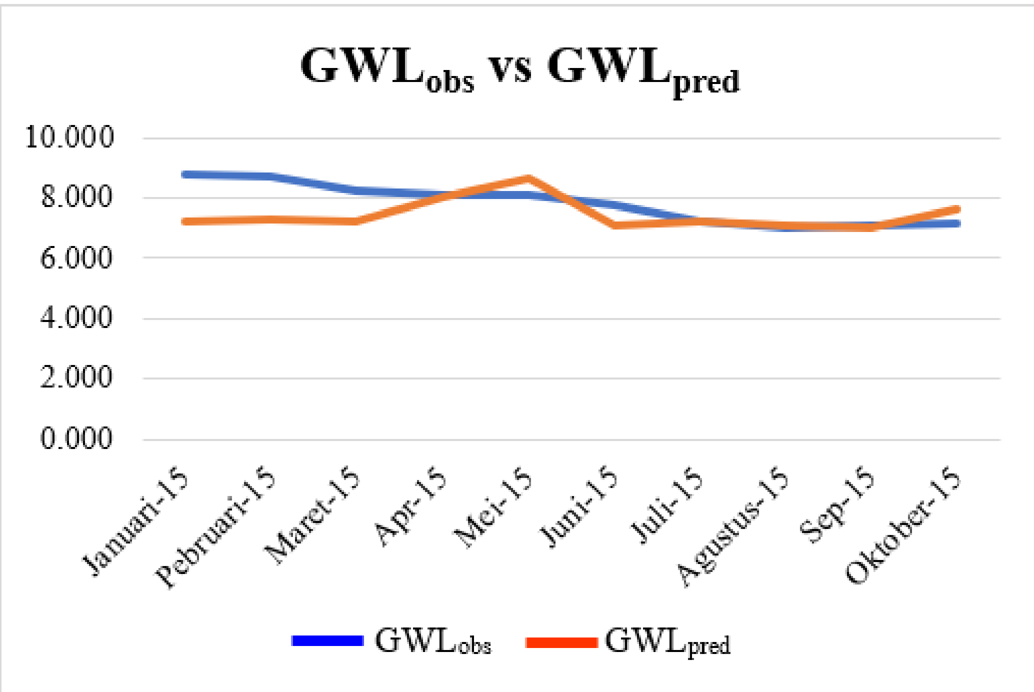

Figure 11. GWL_{obs} and GWL_{pred} in verification stage (ANN 7-14-1; Ngurah Rai)

4. Conclusions

The purpose of this study was to predict groundwater level fluctuation at two monitoring well in Denpasar (Ubung and Ngurah Rai) obtained from hydrometeorology data such as barometric pressure, evaporation, temperature, humidity, wind speed, bright sunshine, and precipitation by using two different modeling methods. The models compared in this study are multiple linear regression (MLR) and artificial neural network (ANN) models. The developed ANN model has three-layer structures: one input layer with four and seven neurons, one hidden layer with four, seven, eight, and fourteen neurons, and one output layer. Logistic sigmoid (logsig) and linear transfer function (purelin) were used in the hidden and output layers as activation functions for the ANN method. Gradient descent with momentum and adaptive learning rate (traingdx) was the training algorithm used.

Barometric pressure, evaporation, temperature, humidity, wind speed, bright sunshine, and precipitation were used as input parameters to predict groundwater level fluctuation. In the MLR model, the hydrometeorology parameter was filtered by quality data test: a consistency test, trend absence test, outlier test, stationary, and persistence test. MLR model results for Ubung wells obtained the R^2 value of 0.606 and for Ngurah Rai well got the R^2 value of 0.257. While based on the ANN model, the value of R^2 was obtained by 0.912 for the Ubung well and R^2 of 0.9133 for the Ngurah Rai well. The results showed that the ANN model has a high determination coefficient (R^2) between the predicted and the observed groundwater level. Based on the model performance, MSE, RMSE, and MAE values of the ANN model were lower than MLR. The Ubung monitoring wells obtained an MSE value of 0.6325, RMSE of 0.7953, MAE of 0.6122 based on the MLR model, while ANN models got an MSE value of 0.143; RMSE is 0.379, and MAE is 0.311. The Ngurah Rai monitoring, well-obtained MSE value of 1.3406, RMSE of 1.1579, and MAE value of 0.9152 for MLR model and ANN models obtained MSE value of 0.0483, RMSE of 0.2198, and MAE of 0.1266. When the results of MLR and ANN model performance were compared, it can be concluded that the ANN provides a more efficient prediction model than the MLR model.

Based on the results of modeling of MLR and ANN obtained, the ANN is superior to the MLR models. In general, the groundwater level prediction model with the ANN approaches provides an excellent correlation and determination coefficient value in Ubung and Ngurah Rai monitoring wells. It can also be concluded the ANN model to predict groundwater level fluctuation offers the advantage of being fast, more accurate, and reliable than MLR due to the account of non-linearities. There is also a significant simplicity in using ANN due to its power to deal with multivariate and complicated problems.

Acknowledgments

The authors wish to express thanks to our supervisor teams reviews for the constructive comments and suggestions for improvement of the manuscript.

REFERENCES

[1] Adamowski, J., Chan, H.F., "A wavelet neural network conjunction model for groundwater level forecasting," Journal of Hydrology, vol. 407, nos. 1–4, pp. 28–40, 2011. DOI: 10.1016/j.jhydrol.2011.06.013.

[2] Konikow, L.F., Kendy, E., "Groundwater depletion: A global problem," Hydrogeology Journal, vol. 13, no. 1, pp. 317–320, 2005. DOI: 10.1007/s10040-004-0411-8.

[3] Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., Lee, K.K., "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer," Journal of Hydrology, vol. 396, nos. 1–2, pp. 128–138, 2010. DOI: 10.1016/j.jhydrol.2010.11.002.

[4] Verma, A.K., Singh, T.N., "Prediction of water quality from simple field parameters," Environmental Earth Sciences, vol. 69, no. 3, pp. 821–829, 2012. DOI: 10.1007/s12665-012-1967-6.

[5] Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., "Artificial neural network modeling of water table depth fluctuations," Water Resources Research, vol. 37, no. 4, pp. 885–896, 2001. DOI: 10.1029/2000WR900368.

[6] Nikolos, I.K., Stergiadi, M., Papadopoulou, M.P., Karatzas, G.P., "Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design," Hydrological Processes, vol. 22, pp. 3337–3348, 2008. DOI: 10.1002/hyp.6916.

[7] Taylor, C.J., Alley, W.M., "Monitoring and the Importance of Long-Term Water-Level Data", U.S. Geological Survey Circular 1217, Colorado, 2001.

[8] Todd, D.K., Mays, L.W., "Groundwater Hydrology", 3rd ed., John Wiley & Sons, Inc, California, 2005.

[9] Coppola Jr, E.A., Rana, A.J., Poulton, M.M., Szidarovszky, F., Uhl, V.W., "A Neural Network Model for Predicting Aquifer Water Level Elevations," Ground Water, vol. 43, no. 2, pp. 231–241, 2005.

[10] Heuvelmans, G., Muys, B., Feyen, J., "Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets," Journal of Hydrology, vol. 319, nos. 1–4, pp. 245–265, 2005. DOI: 10.1016/j.jhydrol.2005.07.030.

[11] Adeloye, A.J., "Multiple Linear Regression and Artificial Neural Networks Models for Generalized Reservoir Storage–Yield–Reliability Function for Reservoir Planning," Journal of Hydrologic Engineering, vol. 14, no. 7, pp. 731–738, 2009. DOI: 10.1061/(asce)he.1943-5584.0000041.

[12] Sahoo, S., Jha, M.K., "On the statistical forecasting of groundwater levels in unconfined aquifer systems," Environmental Earth Sciences, vol. 73, no. 7, pp. 3119–3136, 2015. DOI: 10.1007/s12665-014-3608-8.

[13] Makridakis, S.G., Wheelwright, S.C., Hyndman, R.J., "Forecasting: Methods and Applications", Wiley, 1998.

[14] Hodgson, F.D.I., "The Use of Multiple Linear Regression in Simulating Ground‐Water Level Responses," Ground Water, vol. 16, no. 4, pp. 249–253, 1978. DOI: 10.1111/j.1745-6584.1978.tb03232.x.

[15] Shao, Q., Campbell, N.A., "Modelling trends in groundwater levels by segmented regression with constraints," Australian and New Zealand Journal of Statistics, vol. 44, no. 2, pp. 129–141, 2002. DOI: 10.1111/1467-842X.00216.

[16] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, "Artificial Neural Networks in Hydrology I: Preliminary Concepts," Journal of Hydrologic Engineering, vol. 5, no. April, pp. 115–123, 2000. DOI: 10.5121/ijsc.2012.3203.

[17] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, "Artificial Neural Networks in Hydrology. II: Hydrologic Application," Journal of Hydrologic Engineering, vol. 5, no. 2, pp. 124–137, 2000.

[18] Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., "Groundwater level forecasting using artificial neural networks," Journal of Hydrology, vol. 309, nos. 1–4, pp. 229–240, 2005. DOI: 10.1016/j.jhydrol.2004.12.001.

[19] Lallahem, S., Mania, J., Hani, A., Najjar, Y., "On the use of neural networks to evaluate groundwater levels in fractured media," Journal of Hydrology, vol. 307, nos. 1–4, pp. 92–111, 2005. DOI: 10.1016/j.jhydrol.2004.10.005.

[20] Nayak, P.C., Satyaji Rao, Y.R., Sudheer, K.P., "Groundwater level forecasting in a shallow aquifer using artificial neural network approach," Water Resources Management, vol. 20, no. 1, pp. 77–90, 2006. DOI: 10.1007/s11269-006-4007-z.

[21] Affandi, A.K., Watanabe, K., Tirtomihardjo, H., "Application of an artificial neural network to estimate groundwater level fluctuation," Journal of Spatial Hydrology, vol. 7, no. 2, pp. 17–32, 2007.

[22] Krishna, B., Rao Y.R.S, Vijaya, T., "Modelling groundwater levels in an urban coastal aquifer using artificial neural networks," Hydrological Processes, vol. 22, pp. 1180–1188, 2008. DOI: 10.1002/hyp.6686.

[23] Sethi, R.R., Kumar, A., Sharma, S.P., Verma, H.C., "Prediction of water table depth in a hard rock basin by using artificial neural network," International Journal of Water Resources and Environmental Engineering, vol. 2, no. 4, pp. 95–102, 2010.

[24] Ardana, P.D.H., Redana, I.W., Yekti, M.I., Simpen, I.N., "Prediction of Groundwater Level in the Shallow Aquifer Using Artificial Neural Network Approach," IOP Conference Series: Earth and Environmental Science, vol. 837, no. 1, 2021. DOI: 10.1088/1755-1315/837/1/012010.

[25] Uddameri, V., "Using statistical and artificial neural network models to forecast potentiometric levels at a deep well in South Texas," Environmental Geology, vol. 51, no. 6, pp. 885–895, 2007. DOI: 10.1007/s00254-006-0452-5.

[26] Sahoo, S., Jha, M.K., "Groundwater-level prediction using multiple linear regression and artificial neural network techniques: A comparative assessment," Hydrogeology Journal, vol. 21, no. 8, pp. 1865–1887, 2013. DOI: 10.1007/s10040-013-1029-5.

[27] BPS Kota Denpasar, "Kota Denpasar Dalam Angka 2020 [Denpasar City in 2020 Figures]," Badan Pusat Statistik Kota Denpasar, 2020.

[28] Sudadi, P., Setiadi, H., Denny, B.R., Arief, S., Ruchijat, S., Hadi, S., "Peta Hidrogeologi Lembar Pulau Bali [Hydrogeological Map of Bali Island Sheet]," Direktorat Tata Lingkungan Geologi dan Kawasan Pertambangan Sub. Direktorat Pendayagunaan Air Tanah, 1986.

[29] Anonim, "Laporan Akhir Pekerjaan Pembuatan Peta Zonasi Pemanfaatan Air Tanah Provinsi Bali Tahun Anggaran 2014 [Final Report on Zoning Work for Groundwater Utilization of Bali Province in Fiscal Year 2014]," Dinas Pekerjaan Umum Provinsi Bali, 2014.

[30] Purbo-Hadiwidjojo, M.M., Samodra, H., Amin, T.C., "Peta Geologi Lembar Bali, Nusa Tenggara [Geological Map of Bali, Nusa Tenggara]," Pusat Penelitian dan Pengembangan Geologi, 1998.

[31] Soewarno, "Analisis Data Hidrologi Menggunakan Metode Statistik Dan Stokastik [Analysis of Hydrological Data Using Statistical and Stochastic Methods]," Graha Ilmu, 2015.

[32] Basarin, B., Lukić, T., Pavić, D., Wilby, R.L., "Trends and multi-annual variability of water temperatures in the river Danube, Serbia," Hydrological Processes, vol. 30, no. 18, pp. 3315-3329, 2016. DOI: 10.1002/hyp.10863.

[33] Bonacci, O., Trninic, D., Roje-Bonnaci, T., "Advanced Bash-Scripting Guide An in-depth exploration of the art of shell scripting Table of Contents," Hydrological Processes, vol. 22, pp. 1014-1021, 2008. DOI: 10.1002/hyp.

[34] Zhu, S., Nyarko, E.K., Hadzima-Nyarko, M., Heddam, S., Wu, S., "Assessing the performance of a suite of machine learning models for daily river water temperature prediction," PeerJ, vol. 7, no. May, 2019. DOI: 10.7717/peerj.7065.

[35] Chow, V. Te, Maidment, D.R., Mays, L.W., "Applied Hydrology," McGraw-Hill Book Company, 1988.

[36] Giacomino, A., Abollino, O., Malandrino, M., Mentasti, E., "The role of chemometrics in single and sequential extraction assays: A Review. Part II. Cluster analysis, multiple linear regression, mixture resolution, experimental design and other techniques," Analytica Chimica Acta, vol. 688, no. 2, pp. 122-139, 2011. DOI: 10.1016/j.aca.2010.12.028.

[37] Ghozali, I., "Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23 [Multivariete Analysis Application with IBM SPSS 23 Program]," 8th ed, Badan Penerbit Universitas Diponegoro, 2016.

[38] Haykin, S., "Neural Networks - A Comprehensive Foundation", 2nd ed., Prentice Hall, Ontario, 2005.

[39] Fausett, L., "Fundamental of Neural Networks", Prentice Hall, New Jersey, 1994.

[40] Ozgoren, M., Bilgili, M., Sahin, B., "Estimation of global solar radiation using ANN over Turkey," Expert Systems with Applications, vol. 39, no. 5, pp. 5043-5051, 2012. DOI: 10.1016/j.eswa.2011.11.036.

[41] Sugiyono, "Metode Penelitian Kuantitatif, Kualitatif Dan R&D [Quantitative, Qualitative and R&D Research Methods]," Alfabeta, 2017.

[42] Ardana, P.D.H., Redana, I.W., Yekti, M.I., Simpen, I.N., "A Bibliometric Analysis of Input Parameter in Artificial Neural Network Approach for Groundwater Level Prediction," International Journal of Engineering and Emerging Technologychnology, vol. 5, no. 2, pp. 131-136, 2020.