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

by Putu Doddy Heka Ardana

Submission date: 20-Feb-2022 04:11PM (UTC+0700)
Submission ID: 1766520781
File name: Manuscript_Putu_Doddy_FINAL.doc (1.23M)
Word count: 5941
Character count: 35171

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]Physics Department, Faculty of Mathematics and Science, Udayana University Denpasar, 80361, Bali, Indonesia
*Corresponding Author: doddyhekaardana@unr.ac.id

Abstract Accurate and reliable groundwater level prediction is critical in water resource management. This study aimed to develop two methods to predict 46 months of groundwater level fluctuation. The Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) methods were compared for predicting groundwater levels at the monitoring wells of Ubung and Ngurah Rai in Denpasar Region, Bali, Indonesia. The significant hydrometeorological time series data inputs were barometric pressure, evaporation, temperature, wind, bright sunshine, rainfall, and groundwater level. Moreover, the model performance was assessed statistically and graphically. The groundwater levels predicted by ANN were more consistent with the observed than the MLR-predicted levels at all sites. MLR had a mean square error (MSE) of 0.6325, root mean square error (RMSE) of 0.7953, and mean absolute error (MAE) of 0.6122 in the Ubung monitoring well, while ANN models obtained an MSE of 0.143, RMSE of 0.379, and MAE of 0.311. For the Ngurah Rai monitoring well, MLR models obtained an MSE of 1.3406, RMSE of 1.1579, and MAE of 0.9152. ANN models obtained an MSE value of 0.0483, RMSE of 0.2198, and MAE of 0.1266.

Keywords Groundwater level, prediction, hydrometeorological, multiple linear regression, artificial neural network

1. Introduction

Groundwater is a significant source of supplies for residential, industrial, and agricultural use. It is the only reliable source of supplies in certain areas and preferred due to its near-ubiquitous nature in other areas. However, increased urbanization and water consumption have caused over-exploitation of groundwater. This caused adverse environmental consequences, such as significant water level decrease, well desiccation, stream and lake shrinkage, and reduced well yields. This also caused water quality degradation, specifically in developing countries [1, 2], including Indonesia, namely in Denpasar City, an urban area in Bali Province. Groundwater level forecasting is critical for sustainable water management [3]. Although complex and nonlinear, the groundwater level is affected by various hydrometeorological elements such as barometric pressure, evaporation, temperature, wind, bright sunshine, and rainfall. It is critical to establish accurate models to estimate groundwater levels [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 management and protection, systematic water level observations offer crucial data needed to assess changes in groundwater resources, develop trend models and forecasts, design, implement, and monitor programs [7]. Groundwater fluctuations are the rise and fall of levels caused by natural and human-induced hydrological processes. Therefore, it is critical to understand these events because multiple mechanisms simultaneously require accurate observations. The factors inducing groundwater level variations are urbanization, seismicity, hydrometeorology, such as barometric pressure, evaporation, temperature, wind, bright sunshine, rainfall, tidal influences, and external stress [8].

Predicting groundwater level reactions is crucial for effective planning and management. These strategies have been developed to prevent groundwater mismanagement and over-exploitation. It is difficult to simulate groundwater level fluctuations due to the complexity and non-linearity. Therefore, 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]. Although there are tremendous efforts and resources, distributed numerical flow models' prediction accuracy has not improved enough for diverse water management challenges [9]. This necessitates a dynamic prediction model for handling persistent trends and time-variant behavior. Such instances favor empirical models such as regression and Artificial Neural Network (ANN) models that require less data and are less expensive. Although they cannot manage non-linearity between model inputs and outputs, Multiple Regression Linear (MLR) models are commonly used in hydrological studies [10, 11].

Some hydrologists or hydrogeologists have used ANN tools and statistical techniques such as MLR to predict or forecast water resources systems over the last decade because they are simple and profitable [12]. MLR models show how the correlation between observation and response variables works by adjusting a linear equation to the data collected [13]. Moreover, it produces valuable findings with less data, work, and cost-effectiveness [12]. The models allow for unlimited independent variables. Although MLR models cannot handle non-linearity between model inputs and outputs, they have been widely used in hydrological studies due to their ease of use and parameter interpretation [11]. However, the ANN approach is adapted to modeling nonlinear and dynamic systems such as water resources. It is advantageous over previous techniques because it does not necessitate a detailed mathematical description of underlying processes. Additionally, ANN models anticipate various hydrological problems after adequate training.

Studies on MLR application in groundwater level forecasting are limited. Hodgson [14] used MLR to predict water table responses in the South African Vryburg aquifer using precipitation and pumping as input factors. Similarly, 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]. In line with this, 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 ANN approaches to predict piezometric levels in a deep well in South Texas. Moreover, Sahoo and Jha [26] compared MLR and ANN for simulating transient groundwater levels in an unconfined aquifer system.

No previous study compared the predictive ability of the MLR and ANN techniques in simulating groundwater levels using limited hydrometeorological time series data on barometric pressure, evaporation, temperature, wind, bright sunshine, and rainfall with data screening tests. Examples of filtering tests are trend absence, stationary, persistence, outlier, and data consistency tests. Therefore, this study aimed to examine how two data-driven techniques, such as MLR and ANN, could forecast the Spatio-temporal distribution of water levels in groundwater basins utilizing restricted hydrometeorological time-series data. The MLR and ANN modeling techniques were used, while hydrometeorological data were selected as as model inputs. Therefore, it presents a rigorous scientific technique for comparing two data-driven methodologies for simulating groundwater levels using filtered hydrometeorological data.

2. Materials and Methods

2.1. Study Area

The study area covers 31,42 km^2 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]. The northern Denpasar's aquifer is unconfined and highly productive, with a shallow groundwater level running through fissures and crevices between grains [28]. The Denpasar-Tabanan groundwater basin includes this aquifer [29]. It is a volcanic-sediment-covered terrain with permeable alluvium and young volcanic sediments. In contrast, lower quaternary and tertiary sediments have a wide range of permeability according to the formation. Denpasar is composed of Miocene to Pliocene volcanic products and marine sediment as basement rock, overlain by a thick pyroclastic flow. Also, it has volcanic products and mudflow that resulted from intense volcanic activity during the Pleistocene to Holocene periods of the Quaternary period [30]. Figure 1 shows the study area.

2.2. Data Collecting

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

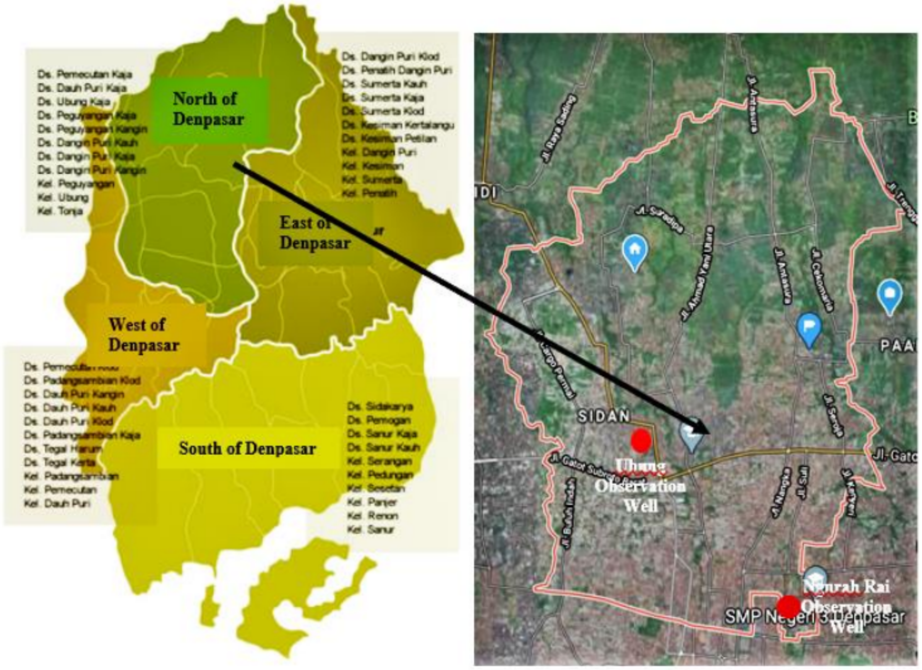

Figure 1. Study area

The hydrometeorological data were adjusted because groundwater level data is only available for 46 months. This 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 2017–2019 period were used to train or calibrate the two observation wells for MLR and ANN models. The ten months from January to October 2015 were used for testing or 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° 43' 56,3" LS; 115° 10' 36,4" 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 time sequence data that must be tested before being used in the analysis. This study collected hydrometeorological data on rainfall, evaporation, humidity, bright sunshine, temperature, wind speed, air pressure, and groundwater level fluctuations. The testing phase, also called data screening, examines and sorts data to obtain hydrometeorological data reliable for analysis to draw good conclusions [31]. The hydrometeorological data test comprises the consistency, trend absence, outlier, stationary, and persistence tests.

2.3.1. Consistency Test

A data consistency test examines the principality in the data obtained using 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]. It eliminates the effects of different data units and random analysis errors [34].

2.3.2. Trend Absence Test

This test determines the randomness or absence of trends from periodic series data using the Spearman Method's Statistical Correlation Ranking Method. This approach correlates time and variant from hydrological variables [31].

2.3.3. Outlier Test

An abnormality or outlier test determines the maximum and minimum data usability from an existing data set [35]. This test is based on data deviating from 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 hydrometeorological data was determined with stationary tests. This study conducted a stationary test with variant stability test (F-Test) and average stability test (t-Test). When the calculated value is greater than the critical value, the data tested does not come from the same population or is not stationary at a certain significance level. Variant values are unstable and nonhomogeneous when test results show that the null hypothesis is rejected [31].

2.3.5. Persistence Test

Persistence tests as a requirement in frequency analysis by testing the presence or absence of dependence on each data were used. When there is no dependency on each value, the data could be used in frequency analysis. The magnitude of the correlation coefficient should be considered [31].

2.4. Modelling

The hydrometeorological data testing resulted in new variables that were used to model MLR and ANN.

2.4.1. Multiple Linear Regression (MLR)

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

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

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

2.4.2. Classic Assumption Test

The value of Y or independent variable to estimate the value associated with X as a dependent variable should be determined to estimate cause-and-effect relationships. Regression analysis helps explain and test the relationship between independent variables into one dependent variable. Multiple regression methods could be an unbiased estimation tool to meet Best Linear Unbiased Estimation (BLUE) requirements. Therefore, the first classic assumption test is performed before the hypothetical test meets BLUE needs. Classical assumptions comprise multicollinearity, normality, autocorrelation, and heteroscedasticity tests [37].

The normality test checks whether dependent and independent variables in regression are normally distributed. A good regression model is normally distributed data or close to normal. Multicollinearity arises when all or some independent variables in a regression model are perfectly linear. Therefore, the multicollinearity test checks whether two or more independent variables are highly correlated in a regression model. This means that an independent variable could be predicted from another independent variable in a regression model. In this case, a decent regression model does not correlate with the variables. Durbin Watson statistical tests are used to find a serial correlation or autocorrelation in time series data. Serial correlation is a relationship between two observations for a variable. In contrast, 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 other is heteroscedastic. Therefore, a decent regression model has homoscedasticity [37].

2.4.3. Artificial Neural Network (ANN)

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

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (2)$$

$$Y_k = \varphi(u_k + b_k) \qquad (3)$$

The bias b_k increases or lowers the net input of the activation function. $x_1, x_2, \ldots, x_m$ are the inputs, $w_{k1}, w_{k2}, \ldots, w_{km}$ are the weights of the neuron k, u_k is the linear combiner output due to input signals, φ is the activation function, while y_k is the output signal of the neuron.

Back-propagation is a popular ANN learning algorithm in multilayered feed-forward networks. The back-propagation networks process data from the input to the output layer through the hidden layer. Finding optimal weights is the goal to get close to targets [38].

This study used feed-forward back-propagation neural network (FFBPNN) architecture and gradient descent with momentum and adaptive learning rate back-propagation (traingdx) for training algorithms. The aim was to find the best algorithm for predicting groundwater levels over the study field. In the hidden layer, logistic sigmoid nonlinear 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 using four statistical metrics or goodness-of-fit criteria, including 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 obtain a model of groundwater level fluctuations, available meteorological data should be tested statistically hydrologically [31]. In this study, data quality testing used outlier or abnormality, trend absence, persistence, stationary, and consistency tests. The test results are in Table 2. Based on the data quality test, the three data eliminated are humidity, bright sunshine, and wind speed data not used in subsequent analysis.

Groundwater level fluctuation data was generated or predicted to extend the data of fluctuations at Ubung and Ngurah Rai monitoring wells. The two monitoring wells were selected due to their appropriate position to represent fluctuations in groundwater levels in the Denpasar City aquifer. The groundwater level fluctuations were predicted by comparing MLR and ANN.

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 Linear Regression

The prediction model with MLR approach produced the following equations:
Ubung monitoring well equation model:

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

The model equation of Ngurah Rai monitoring well is

$$\begin{aligned} GWL = & -263447{,}741 + 268{,}001BP \\ & -3{,}916E - 4{,}953T + 9{,}958P + \varepsilon \end{aligned} \tag{5}$$

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

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

Regression models at Ubung monitoring wells produced a coefficient of determination (R^2) of 0.606. This means that 60.6% of groundwater level could be explained by parameters of barometric pressure (BP), evaporation (E), temperature (T), and precipitation (P). In contrast, the rest is explained by other variables estimated to be due to the exploitation of groundwater by the community. Simultaneous tests obtained F_{count} = 11.905, with probability p-value 0.000 (< 0.05). This indicates an influence or contribution between variables of barometric pressure, evaporation, temperature, and precipitation simultaneously and significantly to fluctuations in groundwater levels. Therefore, regression models on Ubung monitoring wells could be used to predict changes in groundwater levels. Partial or individual tests obtained a p-value of precipitation of 0.361 (0.361 > 0.05). It means no significant relationship between precipitation parameters and fluctuating groundwater levels. In contrast, parameters of barometric pressure, evaporation, and temperature have a partially meaningful relationship to fluctuating groundwater levels. Linear regression models are good when they meet classical assumptions of normally distributed residuals with no multicollinearity, heteroskedasticity, or autocorrelation. The analysis results showed that the regression model for the Ubung monitoring well met the entire classical assumption test. Therefore, the groundwater level fluctuation model could be used with unbiased estimation. The developed regression model was tested using a calibration test based on data for 36 months from 2017 to 2019 and a verification test for ten months data between January and October 2015 with RMSE, MSE, and MAE. The test obtained an MSE value of 0.6325, RMSE of 0.7953, and MAE of 0.6122. The smaller MSE, RMSE, and MAE values show a good predictive value in the calibration process. The verification process obtained an MSE of 1.6415, RMSE of 1.2812, and MAE of 0.8384. This implies a relatively high error rate between the observation and prediction GWL values. Figure 2 shows the comparison between the GWL observation and prediction at the calibration stage and Figure 3 at the verification stage in Ubung monitoring well.

The regression model at Ngurah Rai monitoring well produced a coefficient of determination (R^2) value of 0.257. This means that only 25.7% of groundwater level could be explained by barometric pressure, evaporation, temperature, and precipitation. The rest is defined by other variables, mostly groundwater exploitation by the community where the position of aquifers in Ngurah Rai area is shallow or 15-20 meters below the face soil. Simultaneous tests obtained a F_{count} of 2.685, with a more negligible probability of 0.05. This indicates an influence or contribution between variables of barometric pressure, evaporation, temperature, and precipitation simultaneously and significantly to fluctuations in groundwater levels. Therefore, regression models on Ngurah Rai monitoring wells could be used to predict changes in groundwater levels. The partial or individual test (t-test) obtained a p-value of barometric pressure and precipitation more significant than 0.05 (p-value. > 0.05). This implies a significant partial relationship with fluctuating groundwater levels between barometric pressure and precipitation variables. In contrast, the variables of evaporation and temperature do not have a partially meaningful relationship to fluctuating groundwater levels. The analysis of the Ngurah Rai monitoring well showed that the model met all classical assumption testing. Therefore, the GWL model could be used with unbiased estimation. The developed regression model was tested using a calibration test based on data for 36 months from 2017 to 2019 and a verification test for ten months data between January and October 2015, with RMSE, MSE, and MAE. The calibration process obtained an MSE value of 0.3740, RMSE of 0.6116, and MAE of 0.4717. The smaller MSE, RMSE, and MAE values show a good predictive value in the calibration process. In contrast, the verification process obtained an MSE value of 1.3406, RMSE of 1.1579, and MAE of 0.9152. This implies a relatively high error rate between the observation and prediction GWL values. Figure 4 shows the comparison between the GWL observation and prediction at the calibration stage and Figure 5 at the verification stage in Ngurah Rai monitoring well.

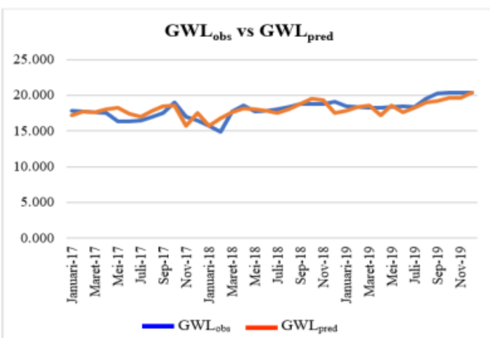

Figure 2. GWL_{obs} and GWL_{pred} in calibration stage (MLR; Ubung)

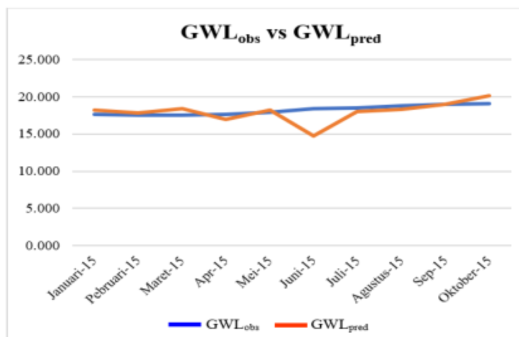

Figure 3. GWLobs and GWLpred in verification stage (MLR; Ubung)

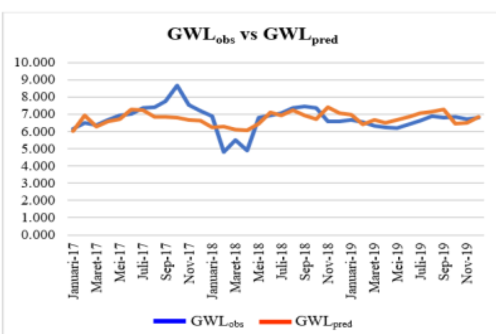

Figure 4. GWL_{obs} and GWL_{pred} in calibration stage (MLR; Ngurah Rai)

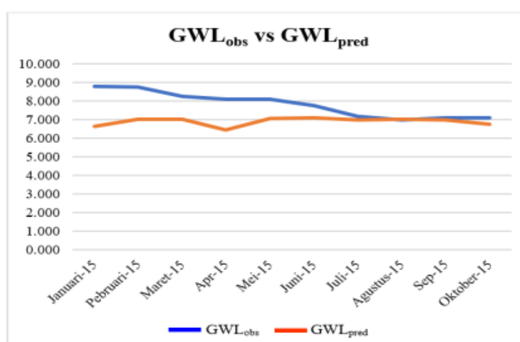

Figure 5. GWL_{obs} and GWL_{pred} in verification stage (MLR; Ngurah Rai)

3.3. Artificial Neural Network

The groundwater level fluctuations in the aquifer area in the Denpasar region were modeled using four architecture variations. The first variation was a 4-4-1 architecture with 4 input variables, 4 neurons hidden layer, and 1 output variable. The second variation was a 4-8-1 architecture with 4 input variables, 8 hidden layer neurons, and 1 output variable. The third variation was a 7-7-1 architecture with 7 input variables, 7 hidden layer neurons, and 1 output variable. The fourth variation was a 7-14-1 architecture with 7 input variables, 14 hidden layer neurons, and 1 output variable. Variations of network architecture are shown in Figures 6 and 7.

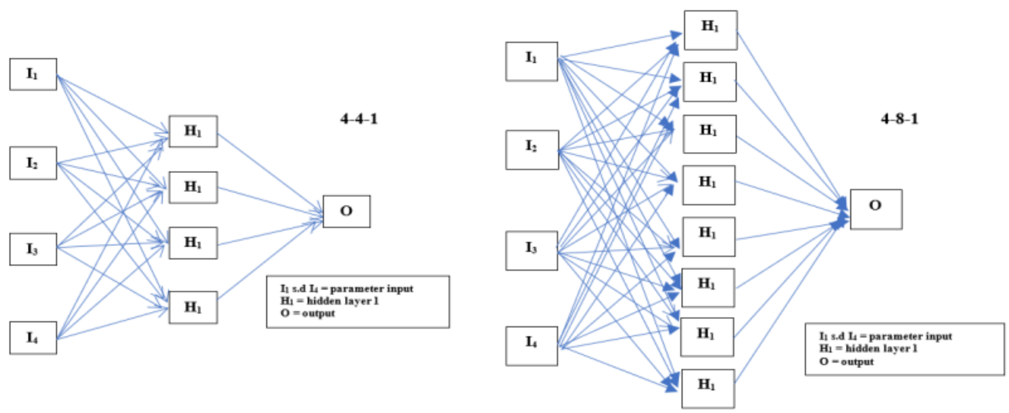

Figure 6. Variations of network architecture (4-4-1 and 4-8-1)

The input parameters for ANN model were selected based on previous studies using barometric pressure, evaporation, temperature, humidity, wind speed, bright sunshine, and groundwater level fluctuation data (GWL) [41]. The parameters of hydrometeorological variables, including barometric pressure, evaporation, temperature, humidity, wind speed, and bright sunshine data, were used simultaneously as variations of network architecture. Another variation used hydrometeorological input parameters based on data quality test results barometric pressure, evaporation, temperature, and precipitation parameters. The data used are hydrometeorology and GWL data for 36 months from 2017 to 2019, calibration and hydrometeorology, and GWL data for ten months between January and October 2015 as verification data.

Groundwater level fluctuations were modeled with the ANN approach using Matlab R2015a software. The aim was to facilitate and accelerate analysis to obtain a prediction model on Ubung and Ngurah Rai monitoring wells for the groundwater addition area in the Denpasar City aquifer. The analysis with the ANN approach involved the normalization process.

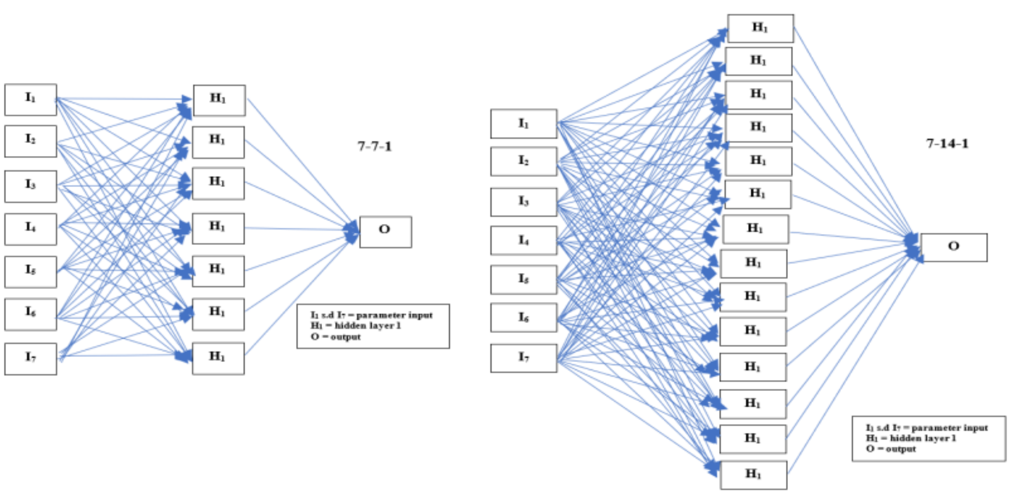

Figure 7. Variations of network architecture (7-7-1 and 7-14-1)

This involved data pre-processing or transformation following the range of activation functions applied, input data training and testing into the network architecture (4-4-1, 4-8-1, 7-7-1, and 7-14-1), and analysis to obtain the most optimal results based on the mean value of square error. Other analysis stages involved obtaining MSE network and correlation coefficient value (R) stage training, testing, and validating the overall variation of ANN model. A different ANN model for each monitoring well was obtained as follows:

Ubung monitoring well ANN model equation:

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

Ngurah Rai monitoring well ANN model equation:

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

Where GWL = groundwater level, X_{mn} = input variable value of 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.

ANN model for Ubung monitoring well used a network architecture of 4-4-1 with 4 input variables, 4 hidden layer neurons, and 1 output variable. The MSE model obtained a value of 0.0018388 in the 87th epoch with an overall R_{model} value of 0.95493. The calibration or training tests obtained an R of 0.955, R^2 of 0.912, MSE of 0.143, RMSE of 0.379, and MAE of 0.311. Furthermore, the verification test obtained an R of 0.891, R^2 of 0.794, MSE of 0.129, RMSE of 0.359, and MAE of 0.319. The 7-14-1 network architecture provided good values in modeling groundwater level fluctuations in Ubung monitoring well. This is seen from a $R_{training}$ value of 0.9674 and the $R_{testing}$ value of 0.7635, implying good reliability in modeling fluctuations in groundwater levels. The values show the reliability of ANN model with an 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 Ubung well.

ANN model for Ngurah Rai monitoring wells used a network architecture of 7-14-1 with 7 input variables, 14 hidden layer neurons, and 1 output variable. The MSE model value was 0.0010372 in the zero epoch with an overall model R-value of 0.95568. The calibration tests obtained an R of 0.9557, R^2 of 0.9133, MSE of 0.0483, RMSE of 0.2198, and MAE of 0.1266. In contrast, the verification test obtained an R of 0.2227, R^2 of 0.0496, MSE of 0.6621, RMSE of 0.8137, and MAE of 0.5985. These values show that ANN model with a 7-14-1 architecture could predict fluctuations in groundwater levels at Ngurah Rai monitoring well. Figure 10 compares GWL observation and prediction at the calibration stage and Figure 11 at the verification stage in the Ngurah Rai well.

The relationship between the $GWL_{observation}$ and the $GWL_{prediction}$ of ANN modeling with the network architecture 4-4-1 at Ubung monitoring well shows a correlation coefficient (R) of 0.955 and a coefficient of determination (R^2) of 0.912 at the calibration stage. The verification stage showed a correlation coefficient (R) of 0.891 and a coefficient of determination (R^2) of 0.794, implying a robust correlation. The relationship between GWL observation and prediction of ANN modeling with the network architecture 7-14-1 at Ngurah Rai monitoring well produced a correlation coefficient (R) of 0.9557 and coefficient of determination (R^2) of 0.9113 at the calibration stage (training). The results are different at the verification or testing stage, with a low correlation coefficient R of 0.2227 and a coefficient of determination (R^2) of 0.0496. However, the value is still better than other ANN models, meaning the 7-14-1 architecture predicts MAT fluctuations in the Ngurah Rai monitoring well.

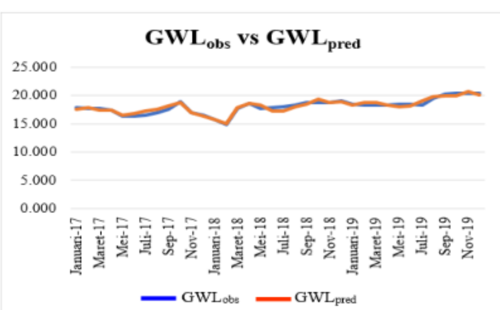

Figure 8. GWL_{obs} and GWL_{pred} in calibration stage (ANN 4-4-1; Ubung)

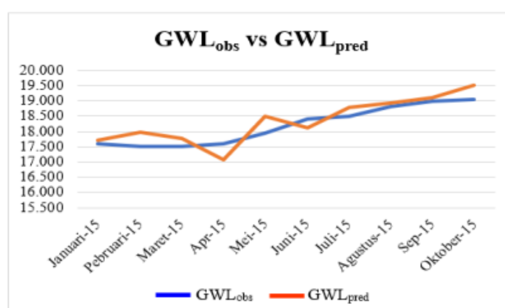

Figure 9. GWL_{obs} and GWL_{pred} in verification stage (ANN 4-4-1; Ubung)

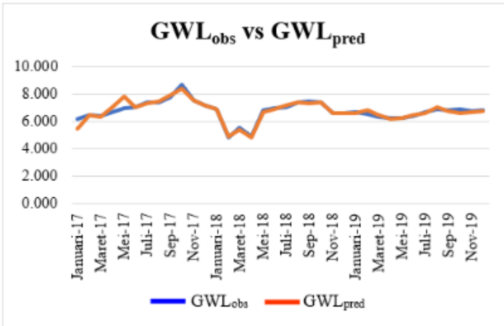

Figure 10. GWL_{obs} and GWL_{pred} in calibration stage (ANN 7-14-1; Ngurah Rai)

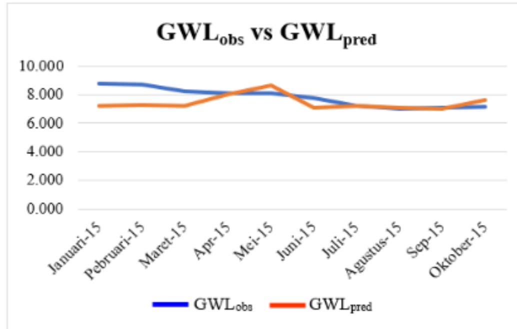

Figure 11. GWL_{obs} and GWL_{pred} in verification stage (ANN 7-14-1; Ngurah Rai)

3. Conclusion

This study aimed to predict groundwater level fluctuation at Ubung and Ngurah Rai monitoring wells in Denpasar. It used hydrometeorology data on barometric pressure, evaporation, temperature, humidity, wind speed, bright sunshine, and precipitation with MLR and ANN models. The developed ANN model has three-layer structures with one input layer having 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 ANN method. The 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 a quality data test comprising the consistency, trend absence, outlier, stationary, and persistence tests. MLR model obtained R^2 value of 0.606 for Ubung and 0.257 for Ngurah Rai. In contrast, ANN model obtained R^2 value of 0.912 for Ubung and 0.9133 for Ngurah Rai. The results showed that ANN model has a high R^2 between the predicted and the observed groundwater level. Based on the model performance, MSE, RMSE, and MAE values of ANN model were lower than MLR. Furthermore, MLR model for Ubung monitoring well obtained an MSE value of 0.6325, RMSE of 0.7953, MAE of 0.6122. ANN models obtained an MSE value of 0.143, RMSE of 0.379, and MAE of 0.311. The MLR model for Ngurah Rai monitoring well obtained an MSE value of 1.3406, RMSE of 1.1579, and MAE value of 0.9152. ANN model obtained an MSE value of 0.0483, RMSE of 0.2198, and MAE of 0.1266. This shows that ANN provides a more efficient prediction model than the MLR model.

The modeling results showed that ANN is superior to MLR models. The groundwater level prediction model with the ANN approaches provides an excellent correlation and determination coefficient value in Ubung and Ngurah Rai monitoring wells. In predicting groundwater level fluctuation, ANN model is quick, more accurate, and reliable than MLR due to the account of non-linearities. Also, ANN is simple to use due to its power to deal with multivariate and complicated problems.

Acknowledgments

The authors thank the supervisory review teams for the constructive comments and suggestions for manuscript improvement.

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, 2007. 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, 2007. 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, 2007. 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, 2007. 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, 2007. 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, 2007. 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

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

[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, 2007

[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, 2007. 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, 2007. 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, 2007. DOI: 10.1007/s12665-014-3608-8

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

[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, 2007. 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, 2007. 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, 2007. 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, 2007

[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, 2007. 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, 2007. 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, 2007. 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, 2007. 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, 2007

[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. 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, 2007. DOI: 10.1007/s10040-013-1029-5

[27] BPS Kota Denpasar, *Kota Denpasar Dalam Angka 2021*, Badan Pusat Statistik Kota Denpasar, Denpasar

[28] Sudadi, P., Setiadi, H., Denny, B.R., Arief, S., Ruchijat, S., Hadi, S., *Peta Hidrogeologi Lembar Pulau Bali*Bandung

[29] Anonim, *Laporan Akhir Pekerjaan Pembuatan Peta Zonasi Pemanfaatan Air Tanah Provinsi Bali Tahun Anggaran 2014*Kota Denpasar

[30] Purbo-Hadiwidjojo, M.M., Samodra, H., Amin, T.C., *Peta Geologi Lembar Bali, Nusa Tenggara*Bandung

[31] Soewarno, *Analisis Data Hidrologi Menggunakan Metode Statistik Dan Stokastik*, Graha Ilmu, Yogyakarta

[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, 2007. 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, 2007. 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. DOI: 10.7717/peerj.7065

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

[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, 2007. DOI: 10.1016/j.aca.2010.12.028

[37] Ghozali, I., *Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23* (Edisi VIII.), Badan Penerbit Universitas Diponegoro, Semarang

[38] Haykin, S., *Neural Networks - A Comprehensive Foundation* (Second.), Prentice Hall, Ontario

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

[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, 2007. DOI: 10.1016/j.eswa.2011.11.036

[41] 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, 2007

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

ORIGINALITY REPORT

11%	6%	12%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	P D H Ardana, I W Redana, M I Yekti, I N Simpen. "Prediction of Groundwater Level in The Shallow Aquifer Using Artificial Neural Network Approach", IOP Conference Series: Earth and Environmental Science, 2021 Publication	3%
2	Sasmita Sahoo, Madan K. Jha. "Predição dos níveis de água subterrânea usando técnicas de regressão linear múltipla e redes neuronais artificiais: uma avaliação comparativa", Hydrogeology Journal, 2013 Publication	2%
3	cyberleninka.org Internet Source	2%
4	Submitted to University of Melbourne Student Paper	1%
5	Submitted to Universitas Pendidikan Indonesia Student Paper	1%

6	Sasmita Sahoo, Madan K. Jha. "On the statistical forecasting of groundwater levels in unconfined aquifer systems", Environmental Earth Sciences, 2014 Publication	1 %
7	Shen Yu Wang, Dan Jiang Chen, Yin Zhong Ye. "Research on Fault Prediction Method of Power Electronic Circuits Based on Particle Swarm Optimization RBF Neural Network", Applied Mechanics and Materials, 2014 Publication	1 %
8	"The Influence of Institutional Ownership, Dividend Policy, Profitability, and Company Sizes on Debt Policy", International Journal of Innovative Technology and Exploring Engineering, 2019 Publication	1 %
9	www.coursehero.com Internet Source	1 %

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	On		