PREDICTION OF GROUNDWATER LEVEL USING ARTIFICIAL NEURAL NETWORK AND MULTIVARIATE TIME SERIES MODELS

Md. Abrarul Hoque*[1] and Sajal Kumar Adhikary[2]

[1]Undergraduate Student, Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, e-mail: milon16khan@gmail.com
[2]Professor, Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, e-mail: sajal@ce.kuet.ac.bd

***Corresponding Author**

ABSTRACT

Groundwater is the major source of potable water supply in Bangladesh. The overextraction of groundwater and as a consequence, the continued depletion of groundwater level are causes numerous problems such as reducation in freshwater supply, increase in water scarcity, reduction in crop yields, degradation of water quality and impact on human health. Therefore, accurate prediction of grounwater level is of great importance for the efficient management of groundwater resources in Bangladesh. In this study, a framework of predicting the groundwater level fluctuations in the shallow aquifer of Bangladesh using is presented and demonstrated through a case study (Kushtia district of Bangladesh). For this purpose, a groundwater level observation station in each upazilla (sub-district) is selected under the study area. The time series groundwater level data collected on a weekly basis during the period from 1999 to 2006 from the observation station is used for the analysis, model development and prediction. Since most shallow aquifers in Bangladesh is unconfined in nature, the fluctuation of groundwater level is highly influenced by rainfall. With this consideration, both groundwater level and rainfall information are required to be taken into account for accurate prediction of groundwater level fluctuations.

In the current study, artificial neural network (ANN) and autoregressive integrated moving average with exogenous variable (ARIMAX) time series models are adopted in MATLAB platform for modelling and prediction of groundwater level fluctuations. In order to develop ANN and autoregressive integrated moving average (ARIMA) based univariate time series models, only groundwater level data is used. However, the rainfall data is used as an exogenous input to both ANN and ARIMAX based multivariate time series models. Finally, one-week-ahead groundwater level prediction is carried out using the adopted models and the performance of each model is checked through a number of performance evaluation criteria. The results indicate that ANN and ARIMAX based multivariate models give better prediction compared to the ANN and ARIMA based univariate time series models. It is also found that ANN based models generate the best prediction over the ARIMA and ARIMAX time series models and proves its superiority over the time series models for groundwater fluctuation modelling and prediciton. Overall, this study proves the fact that the inclusion of exogenous input is highly effective to achieve the enhanced prediction of groundwater level fluctuations in the field of groundwater hydrology.

Keywords: *Artificial Neural Network, ARIMAX, Time Series, Exogenous Input, Groundwater Level.*

1. INTRODUCTION

Groundwater is one of the major sources of potable water supply all over the world and also used in many purposes such as domestic, agricultural, industrial purposes but this resource is not unlimited. The continuous depletion of groundwater due to its excessive utilization and the fluctuations of ground water are major issues, which causes numerous problems such as water scarcity, impact on crop yields, degradation of water quality etc. In some areas, groundwater can be the only usable sources of water. Prediction of groundwater level could be supportive for the proper management and utilization of this resource in an efficient way and for the design of suitable groundwater improvement projects (Adhikary et al, 2012).

Predicting the groundwater level is vital for the effective planning of the conjunctive use of any aquifer with groundwater and surface water (Nayak et al., 2006). By ensuring the conjunctive use of groundwater and surface water one can get many benefits such as storage of the water will be economical, famine or drought can be easily overcome by the storage of groundwater. Furthermore, it ensures the sustainable utilization of this valuable resource for future generation (Taweesin et al, 2018). At present, Bangladesh is facing extreme pressures on its groundwater resource due to the rapid growth of population, and fast increase of the industrialization, where 80% of population is dependent on the groundwater source for their freshwater supplies. The presence of arsenic in groundwater makes it more challenging. Therefore, detailed analysis and prediction of this resource is indispensable for Bangladesh.

Artificial neural network (ANN), an evolutionary data-driven technique, is frequently used to forecast the fluctuations of groundwater level. An important advantage of ANN models is their capability to adjust recurring alterations and also identify patterns of a complex natural system. ANN models are fast and dependable, also produce results analogous to conceptual models, and these models can mine the complex nonlinear relationships between the inputs and outputs, in a process deprived of the physics being explicitly provided (Adhikary et al., 2018). It has been proven to be effective in modelling any nonlinear function with an arbitrary degree of accuracy and the main advantage of this method over traditional methods is that it does not need the complex nature of the process under consideration to be explicitly defined in mathematical form, this makes ANN model is a good tool for modelling water level fluctuations (Nayak et al., 2006).

Forecasting of groundwater level by ANN technique is carried out by several researchers (e.g., Daliakopoulos et al, 2005; Sreekanth et al, 2009; Taormina et al, 2012). Due to its several advantages, ANN models have been applied in many fields including water table depth fluctuation forecasting (Patle et al, 2015; Coulibaly et al., 2001), rainfall forecasting (Luk et al., 2001), salinity estimation and forecasting in rivers (Maier & Dandy, 1996), streamflow forecasting (Adhikary et al., 2018), and rainfall-runoff modelling (Dawson and Wilby, 1998; Hsu et al., 1995). It is widely recognized that like many other factors, the groundwater level in shallow unconfined aquifer is highly influenced by rainfall and thus, rainfall should be considered as an exogenous input for grounwater level modelling and prediction. Therefore, the aim of the current study is to use ANN and multivariate time series models using groundwater level and rainfall variables for groundwater level fluctuation modelling and prediciton in Kushtia district of Bangladesh. In this study, a feed-forward multilayer perceptron (MLP) ANN model and and autoregressive integrated moving average with exogenous variable (ARIMAX) multivariate time series models are adopted for modelling and prediction of groundwater level fluctuations.

2. METHODOLOGY

2.1 Study Area and Datasets

Kushtia district of Bangladesh is selected as a study area in this study, which is showin in Figure 1. A larger part of the Ganges-Kobadak irrigation project (also known as the G-K project) is located in this area. Five groundwater (GWL) monitoring wells with a rainfall station in the study area are selected from each upazila (sub-district) of Kushtia to carry out the study. The frequency of GWL data collection is a week. So, weekly time series GWL data from 1999 to 2006 with 414 data points are collected from the Bangladesh Water Development Board (BWDB). As mentioned earlier, the rainfall will be used as an exogenous input for model development, rainfall data for the same period is collected from the BWDB. Details of the collected are presented in Table 1 and the locations of the GWL monitoring stations are shown in Figure 1.

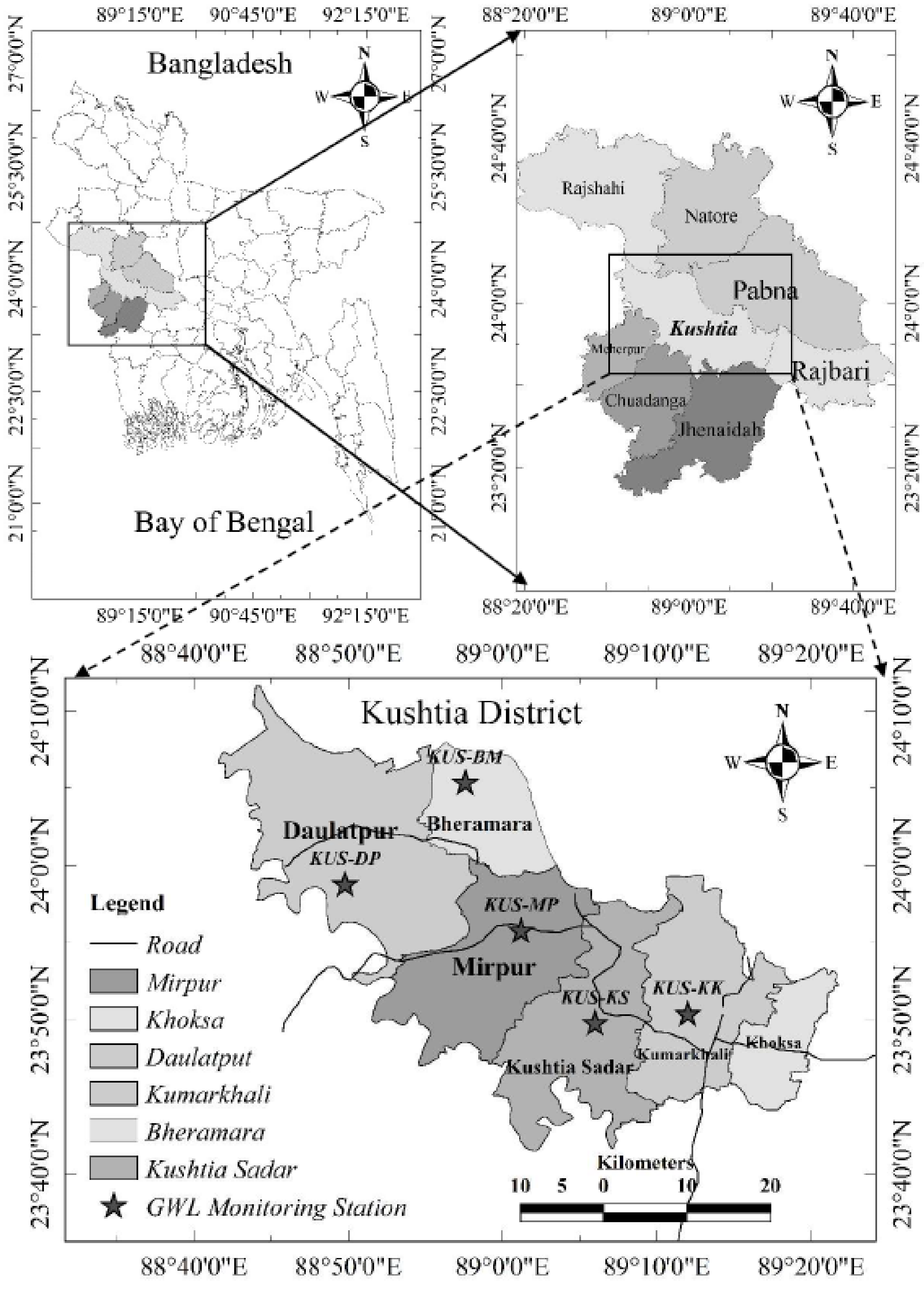

Figure 1: Study area with the location of GWL monitoring stations

Table 1: Details of GWL and Rainfall monitoring stations

ID	Latitude	Longitude	Location
Well ID			
KUS-BM	24.09	88.96	Bheramara, Kushtia
KUS-DP	23.98	88.83	Daulatpur, Kushtia
KUS-KS	23.83	89.10	KushtiaSadar Kushtia,
KUS-KK	23.84	89.20	Kumarkhali, Kushtia
KUS-MP	23.93	89.02	Mirpur, Kushtia
Rainfall Station ID			
RF	24.05	88.99	Mirpur, Kushtia

2.2 ANN Model

In this study, one week ahead GWL modelling and prediction is carried out using the ANN model. The steps involved are preparation of datasets, division of data for model training, validation and testing purposes, selection of the best ANN architectures and finally validation and testing of the developed ANN model. Different ANN architectures have been analyzed to identify the best ANN architecture for modeling testing and evaluation of the model performance. The GWL data is used for developing univariate ANN model whereas rainfall data is used as an exogenous variable along with the GWL data to develop the multivariate ANN model in the current study. Finally, the performance of each model is evaluated through some model performance criteria. The ANN model development is performed in the neural network toolbox of MATLAB platform.

In this study, a multilayer perceptron (MLP) feed forward ANN model is adopted. The architecture of a typical ANN model consists of an input layer, a hidden layer and an output layer, which is shown in Figure 2. In this study, sigmoid activation function is used in the hidden layer of ANN model with the linear activation function in the output layer. The input and output variables are standardized between 0 and 1, to make them fall within a specified range following the ANN modelling framework.

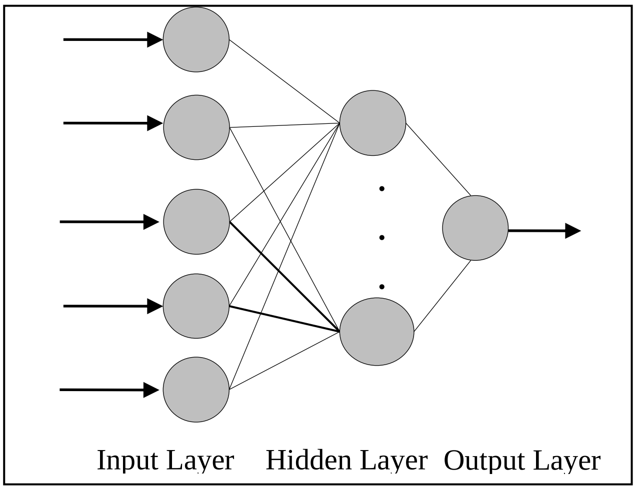

Figure 2: A typical architecture of an ANN model

In ANN modelling, data are generally divided into three categories including training, validation and testing. In this study, the data are divided in the proportion of 70% for training the ANN model, 15% for the ANN model validation and the remaining 15% for testing of the ANN model. The training dataset is used to train the ANN model and the validation dataset is used during the training process to avoid overtraining problem in the model, while validation error rises for a specified number of iterations then the training is stopped (Adhikary et al., 2018). In general, large numbers of data points are used in training because it can capture the patterns exist in the dataset. ANN models are trained by adjusting the values of the connection weights between the network components (Sreekanth et al., 2009). Back-propagation algorithm is used to train the ANN model, it adjusts the weights and biases in the backward direction, which are fixed random value to the same results every time. There are many training algorithms have been developed for back propagation. The widely used Levenberg-Marquardt (LM) back propagation algorithm is used in this study. To stop the back-propagation 1000 epochs are used.

2.3 ARIMAX Model

An ARIMAX multivariate time series model is simply a combinaton of a univariate autoregressive integrated moving average (ARIMA) time series model and exogenous variables, X. Since ARIMA model is a univarate model, it does not contain any exogenous variable. On the other hand, the ARIMAX models contain one or more exogenous variables. The ARIMAX model has some certain advantages over ARIMA model and thus widely used for model development and prediction purposes. It is suitable for forecasting when data is multivariate, stationary or non-stationary, contains any type of data form such as seasonality, periodicity, trend etc. A few studies have been conducted for grounwater level modelling and prediction in Bangladesh (e.g., Adhikary et al., 2012) but no studies have been found using ARIMAX model for prediciton purpose. Therefore, an attemp has been made in the current study to adopt ARIMAX model for modelling and prediciton of groundwater level fluctuations. The development of ARIMA and ARIMAX models are performed in the econometric toolbox of MATLAB platform.

In this study, one week ahead GWL modelling and prediction is also carried out using the ARIMA and ARIMAX models. The GWL data is used for developing univariate time series model using ARIMA modelling technique whereas rainfall data is used as an exogenous variable along with the GWL data to develop the multivariate time series model using the ARIMAX modelling technique in this study. Finally, the performance of each model is evaluated through some model performance criteria.

2.4 Model Evaluation Criteria

Different performance evaluation criteria have been used to assess the efficacy of the adopted modelling techniques. The mean squared error (MSE) and Nash-Sutcliffe efficiency (NSE) expressed in Eqs. (1)-(2) are used in this study to evaluate the performance of the adopted ANN and ARIMAX models in groundwater level modelling and prediction. The model with the lowest MSE and with the highest NSE is selected as the best model in each case.

$$NSE = 1 - \frac{\sum_{q=1}^{n}[Y_{obs}(q) - Y_{est}(q)]^2}{\sum_{q=1}^{n}[Y_{obs}(q) - \overline{Y}_{est}(q)]^2} \qquad (1)$$

$$MSE = \frac{\sum_{q=1}^{n}[Y_{obs}(q) - Y_{est}(q)]^2}{n} \qquad (2)$$

Where Y_{obs} is the observed data, Y_{est} is the estimated data, $\overline{Y}_{est}$ is the mean value of the estimated data and n is the number of observations.

In order to evaluate the ARIMA and ARIMAX models, two commonly used criteria such the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are additionally used along with the MSE. The model with the lowest MSE, AIC and BIC gives the best model.

3. RESULTS AND DISCUSSION

In this study, the GWL and rainfall data are prepared with a five weeks lag, which generates a total of ten inputs for multivariate case and five inputs for univariate case and by using that data, one week ahead GWL modelling and prediction is carried out using the ANN, ARIMA and ARIMAX models. The model can be characterized by the structure (architecture), training methods or algorithms, neurons, number of layers, inputs, weights, outputs and activation functions. In this study, only one single layer is used because it is often adequate for fitting multi-dimensional mapping problems with satisfactory neurons (Wu et al., 2005). Large number of neurons in the hidden layer may cause over fitting (Adhikary et al, 2018). Hence, the optimum number of hidden neurons should be determined during the ANN model development. The optimum number of neurons in the hidden layer is determined by trial and error method by varying the number of hidden neurons from 2 to 10, which was found satisfactory in this study.

For all five stations, ANN models with exogenous input (rainfall in this study) are developed. The models are trained and validated and finally, tested. The performance of each model is evaluated based on the evaluation criteria indicated above. The best ANN models identified for all GWL monitoring stations for both multivariate and univariate cases with their performance measures is presented in Table 2. It can be seen from the table that the best ANN model for KUS-BM, KUS-DP, KUS-KK, KUS-KS and KUS-MP are 10-6-1, 10-3-1, 10-2-1, 10-4-1 and 10-6-1, respectively.

Table 2: Details of the ANN multivariate models for all five GWL stations in the study area

Well ID	**ANN model architecture**	**Model performance**	
		MSE	**NSE**
Multivariate ANN models			
KUS-BM	10-6-1	**0.0316**	**0.9848**
KUS-DP	10-3-1	0.0595	0.9567
KUS-KK	10-2-1	0.0849	0.9817
KUS-KS	10-4-1	0.0538	0.9653
KUS-MP	10-6-1	0.0326	0.9843
Univariate ANN models			
KUS-BM	5-9-1	**0.0291**	**0.9860**
KUS-DP	5-7-1	0.0392	0.9714
KUS-KK	5-9-1	0.0702	0.9849
KUS-KS	5-10-1	0.0754	0.9515

The models developed for all five selected GWL stations in the study area using the ARIMA or ARIMAX techniques are presented in Table 3. As can be seen from the table, the best time series model for multivariate case is ARIMAX (3,0,2) whereas the best univariate time series model is ARIMA (2,0,1). It is also identified that the performance of the ANN models are better than the ARIMA or ARIMAX models. This justifies the application of ANN modelling technique in the current study.

Table 3: Details of the ARIMAX and ARIMA time series models for all five GWL stations

Well ID	Model structure	Model performance		
		MSE	AIC	BIC
ARIMAX multivariate time series models				
KUS-BM	ARIMAX (3,0,3)	0.03738	-166.485	-130.318
KUS-DP	ARIMAX (3,0,2)	0.04555	-87.1871	-55.0384
KUS-KK	ARIMAX (2,0,1)	0.15456	411.9365	436.0627
KUS-KS	ARIMAX (1,0,3)	0.06162	35.10959	63.27372
KUS-MP	ARIMAX (3,0,2)	0.03685	-174.354	-142.206
ARIMA univariate time series models				
KUS-BM	ARIMA (2,0,1)	0.04159	-131.619	-111.491
KUS-DP	ARIMA (2,0,1)	0.06493	52.7996	72.929
KUS-KK	ARIMA (3,0,1)	0.15599	417.701	441.856
KUS-KS	ARIMA (2,0,3)	0.06813	76.755	104.936
KUS-MP	ARIMA (2,0,1)	0.04006	-147.101	-126.971

4. CONCLUSIONS

In this current study, an attempt has been made for modelling and prediction of groundwater level fluctuation for five selected monitoring wells in Kushtia district of Bangladesh. For this purpose, artificial neural network (ANN) and autoregressive integrated moving average (ARIMA) with exogenous variable (ARIMAX) multivariate time series models are adopted. Model development and analysis are carried out in the neural network toolbox for ANN model and in the econometric toolbox for the ARIMA and ARIMAX models in MATLAB platform. In order to develop ANN and autoregressive integrated moving average (ARIMA) based univariate time series models, only groundwater level data is used. However, the rainfall data is used as an exogenous input to both ANN and ARIMAX based multivariate time series models. Finally, one-week-ahead groundwater level prediction is carried out using the adopted models and the performance of each model is checked through a number of performance evaluation criteria. The results indicate that ANN and ARIMAX based multivariate models generate better prediction compared to the ANN and ARIMA based univariate time series models. It is also found that ANN based models give the best prediction over the ARIMA and ARIMAX time series models and proves its superiority over the time series models for modelling and prediction of groundwater level fluctuations.

REFERENCES

Adhikary, S. K., Muttil, N., & Yilmaz, A. G. (2018). Improving streamflow forecast using optimal rain gauge network-based input to artificial neural network models, *Hydrology Research*, 49(5), 1559-1577.

Adhikary, S. K., Rahman, M., & Gupta, A. D. (2012). A stochastic modelling technique for predicting groundwater table fluctuations with time series analysis, *International Journal of Applied Science and Engineering Research*, 1(2), 238-249.

Coulibaly, P., Anctil, F., Aravena, R., & Bobée, B. (2001). Artificial neural network modeling of water table depth fluctuations, *Water resources research*, 37(4), 885-896.

Daliakopoulos, I. N., Coulibaly, P., &Tsanis, I. K. (2005). Groundwater level forecasting using artificial neural networks, *Journal of hydrology*, 309(1-4), 229-240.

Dawson, C. W., &Wilby, R. (1998). An artificial neural network approach to rainfall-runoff modelling, *Hydrological Sciences Journal*, 43(1), 47-66.

Hsu, K. L., Gupta, H. V., &Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process, *Water resources research*, 31(10), 2517-2530.

Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting, *Mathematical and Computer modelling*, 33(6-7), 683-693.

Maier, H. R., & Dandy, G. C. (1996). The use of artificial neural networks for the prediction of water quality parameters, *Water resources research*, 32(4), 1013-1022.

Nayak, P. C., Rao, Y. S., & Sudheer, K. P. (2006). Groundwater level forecasting in a shallow aquifer using artificial neural network approach, *Water resources management*, 20(1), 77-90.

Patle, G. T., Singh, D. K., Sarangi, A., Rai, A., Khanna, M., &Sahoo, R. N. (2015). Time series analysis of groundwater levels and projection of future trend, *Journal of the Geological Society of India*, 85(2), 232-242.

Sreekanth, P. D., Geethanjali, N., Sreedevi, P. D., Ahmed, S., Kumar, N. R., &Jayanthi, P. D. (2009). Forecasting groundwater level using artificial neural networks, *Current Science*, 96(7), Paper ID: 00113891.

Taormina, R., Chau, K. W., &Sethi, R. (2012). Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon, *Engineering Applications of Artificial Intelligence*, 25(8), 1670-1676.

Taweesin, K., Seeboonruang, U., &Saraphirom, P. (2018). The influence of climate variability effects on groundwater time series in the lower central plains of Thailand, *Water*, 10(3), 290.