MASTER'S THESIS OF NATURAL SCIENCE

Application of Artificial Neural Network to Prediction of Groundwater Level and to Evaluation of Influence Factors in Yangpyeong Riverside Area

인공 신경망 기법을 이용한 강변 지역 지하수위 예측 및 영향요인 평가

February 2018

Graduate School of Seoul National University

College of Natural Sciences

School of Earth and Environmental Sciences

LEE, SANG HOON

Application of Artificial Neural Network to Prediction of Groundwater Level and to Evaluation of Influence Factors in Yangpyeong Riverside Area

지도교수 이 강 근

이 논문을 이학석사 학위논문으로 제출함

2018년 2월

서울대학교 대학원

지구환경과학부

이 상 훈

이상훈의 석사 학위논문을 인준함

2017년 12월

위 원 장 이 준기 (인)

부 위 원 장 이 강근 (인)

위 원 김 덕진 (인)

ABSTRACT

Sustainable use of groundwater resource is crucial issues in recent years due to increasing use of groundwater. In the Yangpyeong riverside area, where anthropogenic activities such as operation of the groundwater heat pump system (GWHPs) and the water curtain cultivation (WCC) are concentrated, prediction of changes in the groundwater levels is necessary to suggest management plans for protecting the groundwater resource in quantitative aspects. As a means of the groundwater level forecasting, the ANN model was applied in this study. It was revealed that the surface water level, the WCC and the GWHPs operation showed higher correlation with the groundwater level fluctuation in the study area rather than the precipitation. Based on the abovementioned influence factors, network architecture was constructed in train and test period, and prediction of the groundwater level at 8 wells was performed. The forecasted groundwater levels showed good matches with the observed data with low RMSE values in range of 0.03~0.06 m. Additionally, the contribution and the relative importance of each influence factor were computed to decompose the combined effects. Weights method and PaD method which are pre-existing method to quantify the input variables in the ANN model were applied at first to calculate the contribution or the relative importance. An extraction method which can compare spatial and temporal contributions of each influence factor was developed in this study, and verified the suitability on the spatial and temporal comparison. As a results, the surface water level accounted for dominant effect on the groundwater level fluctuation (64.09~83.30 %), the WCC showed low effect (16.04~26.76 %), and the GWHPs had a little or no effect (0.29~6.26 %). Especially, effect of the WCC was changed drastically depending on times, showing greater contribution than that of the surface

water level in a period of the WCC operation in the winter season. Evaluation of influence factors on the groundwater level by the ANN will help understanding driving forces to change the groundwater level especially for an aquifer system which have complex and nonlinear features.

Key words: the groundwater level forecasting, artificial neural network model, influence factors, contribution, relative importance

LIST OF CONTENTS

ABSTRACT ..... **i**

LIST OF CONTENTS ..... **iii**

LIST OF FIGURES ..... **v**

LIST OF TABLES ..... **viii**

1 INTRODUCTION **1**

1.1 Research background **1**

1.2 Objectives and scope **4**

2 STUDY AREA **6**

2.1 Site description **6**

2.2 Hydro-geological characteristics **12**

3 METHODOLOGY **22**

3.1 Time series analysis **22**

3.2 Artificial Neural Network (ANN) **24**

3.3 Methods to study the contribution of variables in ANN models **30**

4 RESULTS AND DISCUSSION 36

4.1 Influence factors for the groundwater level fluctuation .. 36

4.2 Groundwater level forecasting using ANN models 56

4.3 Computing contribution of influence factors 65

5 CONCLUSION 85

6 REFERENCES 87

LIST OF FIGURES

Fig. 2-1. Location of the study area. 8

Fig. 2-2. Climatic data in Yangpyeong for 10 years from 2007 to 2016. 9

Fig. 2-3. Land use map for agriculture. 10

Fig. 2-4. (a) Location and (b) picture of the greenhouse performing the WCC 11

Fig. 2-5. Map of the study area with four lines (A-A', B-B', C-C', D-D'). 14

Fig. 2-6. Geological cross section for line A-A'. 15

Fig. 2-7. Geological cross section for line B-B'. 16

Fig. 2-8. Geological cross section for line C-C'. 17

Fig. 2-9. Geological cross section for line D-D'. 18

Fig. 2-10. Monthly maximum, mean, and minimum of the surface water level and the groundwater level from March 2014 to December 2015. 21

Fig. 3-1. Schematic diagram of a node. 26

Fig. 3-2. Schematic diagram of the ANN model. 27

Fig. 4-1. The groundwater level (YSN01) and the surface water level (Paldang dam) in 2015. 38

Fig. 4-2. Cross-correlation between the surface water level (Paldang dam) and the groundwater level (YSN01) for whole period and no WCC, GWHPs period. 40

Fig. 4-3. The groundwater level time series data at YSN01 showing the effect of the water curtain cultivation: (a) from November 2014 to March 2015, and (b) in December 2014. 43

Fig. 4-4. Result of spectral analysis and peaks at cycle of 24 hours for the groundwater level at YSN01. 45

Fig. 4-5. The groundwater level in response to the operation of the groundwater heat pump system from April 5 to April 15 in 2015. 48

Fig. 4-6. Change in the groundwater level by the groundwater heat pump system according to the distance from pumping/injection well. 49

Fig. 4-7. The groundwater level (YSN01) and the precipitation from March 8 to December 31 in 2014. 52

Fig. 4-8. Cross-correlation between the precipitation and the groundwater level at YSN01 for whole season and wet season. 54

Fig. 4-9. The precipitation and the discharge rate at the Paldang dam. 55

Fig. 4-10. Input data set and the train, test, and predict period for ANN model. 58

Fig. 4-11. Distribution of RMSE values for 125 combination of three model parameters. 60

Fig. 4-12. Results of the prediction of the groundwater level at (a) YSN01, (b) YSN03, (c) YSO01, (d) YSO03, (e) YSO07, (f) YSO08, (g) YSO11, (h) YSO12..... 63

Fig. 4-13. Relative importance obtained by weights method. 67

Fig. 4-14. Sum of square partial derivatives obtained by PaD method. 69

Fig. 4-15. Input data set for predict period. 70

Fig. 4-16. Comparison of sum of square partial derivatives for WCC on February and April in 2017. 71

Fig. 4-17. Controlled input data sets which have only effects of (a) surface water level, (b) WCC, and (c) GWHPs. 75

Fig. 4-18. Contribution obtained by extraction method. 76

Fig. 4-19. Relative importance obtained by extraction method. 77

Fig. 4-20. (a) 3-D bars of the contribution, and (b) pie charts of the relative importance plotted on the map. 78

Fig. 4-21. Temporal comparison of the contribution: (a) February, (b) March, and (c) April. 80

Fig. 4-22. Temporal comparison of the relative importance: (a) February, (b) March, and (c) April.....81

Fig. 4-23. Contribution or relative importance of the GWHPs obtained by three methods (a) for wells around the injection well; (b) for wells around the pumping well. 83

Fig. 4-24. Contribution or relative importance of the WCC on February and April in 2017 obtained by three methods. 84

LIST OF TABLES

Table 2-1 Statistical information of the surface water level. 19

Table 2-2 Statistical information of the groundwater level. 20

Table 4-1 Time lags and maximum cross-correlation coefficients between the surface water level and the groundwater level for whole period and no WCC, GWHPs period.....39

Table 4-2 Power spectrum of cycle of 24 hours for May 2014 and December 2014. 44

Table 4-3 Time lags and maximum cross-correlation coefficients between the precipitation and the groundwater level for whole season and wet season.....53

Table 4-4 Data sets and network sets of the ANN model. 57

Table 4-5 Results of exploration of model parameters for each well to perform the optimal prediction of the groundwater level.....61

Table 4-6 Prediction errors for each well. 64

Table 4-7 Classification of wells according to features of the contribution and the relative importance. 79

1 INTRODUCTION

1.1 Research background

Groundwater, which is stored in soil pore or fractures of rock beneath the land surface is one of the important natural resources. The quantity of groundwater could be increased by areal recharge from precipitation or losing stream and decreased by evapotranspiration or discharge to surface water naturally. However, since groundwater use has been increased for various purposes like agriculture, industry or drinking water in recent years, the issue related to sustainability on the groundwater use also has been raised (Alley et al., 1997; Gleeson et al., 2012). Excessive groundwater pumping can cause depletion of groundwater resources, and it leads to several environmental problems such as seawater intrusion, ecosystem change, and land subsidence (Foster and Chilton, 2003; Konikow and Kendy, 2005). Therefore, it is important to manage the groundwater sustainably and to plan thoroughly before the groundwater development. For effective management, prediction of the groundwater level which has a role of fundamental indicator to estimate the quantity of groundwater has been practiced (Daliakopoulos et al., 2005; Nayak et al., 2006; Wong et al., 2007; Yang et al., 2009).

For these reasons, many studies have tried to predict the groundwater level. Based on factors influence the groundwater, time series models such as autoregressive (AR) model, moving average (MA) model and autoregressive moving average (ARMA) model have been applied to groundwater level forecasting (Eriksson, 1970; Yi et al., 2004). These models are kinds of linear model, which considers linear regression or linear relationship between influence factors and the groundwater level. However, groundwater systems have complex features and nonlinearity so the groundwater

forecasting using time series models is questionable especially in complex environments (Yang et al., 2007). Numerical model such as MODFLOW, FEFLOW, and HydroGeoSphere is another tool to estimate the groundwater level (Brunner and Simmons, 2011; Trefry and Muffels, 2007; Wang et al., 2008). Because it is based on mathematical equation representing mechanism of groundwater flow, fluctuation of the groundwater level by some exogenous factors like precipitation or pumping can be calculated if all necessary hydraulic variables for governing equations are given. At last, an Artificial Neural Network (ANN) model, which was developed by a desire to imitate the functioning of human brain have been applied mostly to forecast the groundwater level in recent years (ASCE, 2000a; ASCE, 2000b; Daliakopoulos et al., 2005; Yoon et al., 2011). The ANN is one of empirical models, and it can solve the complex problems and consider nonlinearity. Empirical model has advantage that it can be used without detail hydraulic properties which are essential for the numerical model.

As a study for the ANN model, Daliakopoulos et al. (2005) compared seven combinations of ANN architectures and training algorithms to derive the optimum network for best prediction efficiency and accuracy to forecast the groundwater level. Yoon et al. (2016) applied a weight error function approach to improve stability and accuracy of ANN models for a long term groundwater level prediction. They predicted groundwater level fluctuating in response to the natural factors such as rainfall, surface water level or tidal effect in coastal area. But only a few studies (Mohanty et al., 2010) have performed groundwater level forecasting fluctuating by human activities like pumping and irrigation.

The study site is located at riverside area in Yangpyeong, and it was revealed that several factors such as the water curtain cultivation (WCC) and operation of the

groundwater heat pump system (GWHPs) affect the groundwater level in previous studies (Moon et al., 2012; Nam and Ooka, 2010; Park et al., 2015). Because the groundwater level fluctuates complexly by these factors, the groundwater level forecasting is essential for efficient management of the groundwater level in Yangpyeong area. Considering features of site in this study such as interaction with surface water, WCC and operation of the GWHPs, prediction of the groundwater level would be challenging.

Because the ANN model estimates the output from input variables, studies for understanding how much each variable affect the output have been conducted. There are several methods like weights method, PaD method, profile method, and perturbation method which obtain the contribution of input variables using functions of ANN (Dimopoulos et al., 1999; Gevrey et al., 2003; Olden and Jackson, 2002). Although evaluation of input variables using these methods is not performed actively in a field of hydrogeology, it would be significant work to evaluate diversifying influence factors of the groundwater level in the future.

1.2 Objectives and scope

In this research, time series of the groundwater level data in the Yangpyeong study area were applied as a main subject for three primary purposes. They are 1) understanding influence factors for the groundwater level change, 2) groundwater level forecasting using ANN models, and 3) evaluating the relative importance of influence factors.

First step, search for influence factors which affect the groundwater level performed as foundation for constructing the ANN model to predict the groundwater level. This is fundamental and significant work because determined factors would be applied to the ANN model as input variables. Time series analyses like cross correlation and spectral analysis were conducted to investigate the relationship between possible candidates for influence factor and the groundwater level.

Second step, constructing the ANN models has an objective to find an optimal network for accurate groundwater level forecasting. The ANN model was constructed once, and then parameters configuring the ANN model and input structures were determined for the optimal ANN model.

Last step, the contribution and the relative importance for each influence factors that were used as input data in ANN model will be calculated. The purpose of this work is to understand which factor contributes to the groundwater level change more and to manage groundwater efficiently. Also, if the contribution is calculated exactly, it is possible to understand how effects of influence factors distribute spatially and how they change over time.

This study suggests an effective ANN model for prediction of groundwater level at the study site in Yangpyeong riverside area. It will be helpful to manage and develop groundwater sustainably.

2 STUDY AREA

2.1 Site description

The study site is located at the Han River Environmental Research Center (HRERC) in a riverside area of the alluvial river island, Yangsu-ri, Yangpyeong-gun, South Korea. The river island is situated at a junction of two rivers of the North Han river and the South Han river (Fig. 2-1). The Chungpyeong dam and the Paldang dam is located from the study site upstream and downstream, for purposes of flood prevention, water supply and power generation. Therefore, a water level of the Han river is controlled by opening/closing event of a gate of the Paldang dam.

The weather observation data (air temperature and precipitation) of Yangpyeong-gun for 10 years (2007~2016) are represented in (Fig. 2-2) (Korea Meteorological Administration, 2017). The mean air temperature and the mean precipitation of each month are plotted by dashed line and bar, respectively. The highest temperature was recorded on August with 24.02℃ and the lowest temperature was observed on January with -3.37℃. The precipitation in a wet season (June-August) was about 880.19 mm showing very high percentage (61.78%) of the total annual precipitation in comparison to the precipitation in a dry season (December-February) of 66.15 mm (4.64%).

GWHPs is installed at the HRERC to cool in summer and to heat in winter. The GWHPs refers to the open loop ground source heat pump system, using groundwater as a source of heat. Therefore, it was installed at a riverside area to supply the groundwater stably. To pump and inject the water for the GWHPs operation, three geothermal wells, one test well and 18 monitoring wells were installed for

monitoring the system. When the GWHPs is operated, groundwater is pumped from an abstraction well (YSG03) and injected into injection wells (YSG01, and YSP) with rate of 300 m^3/day (Ministry of Land, Infrastructure and Transport, 2017). Because 9 wells were abandoned in June 2016, only 13 wells remain today and record the groundwater level continuously.

Figure 2-3 shows the land uses around the study area. At the eastern area of the study site, most surface lands are used as agricultural lands which is composed of farmland dominantly and paddy occasionally. Among them, some of parts indicated as yellow box in Fig. 2-4a is the location of greenhouses performing the WCC, and Fig. 2-4b is the picture of one of them. In winter season, agricultural activities are held in the greenhouse and the WCC is practiced to protect crops from being frozen to death. WCC is a way to keep warm in the greenhouse with pumping lots of groundwater which has higher temperature than air and forming water curtain around the greenhouse. Although the WCC sites are not investigated in detail such as distances from the study area and pumped amounts of groundwater, impact of the WCC would be shown in the groundwater level in the study area.

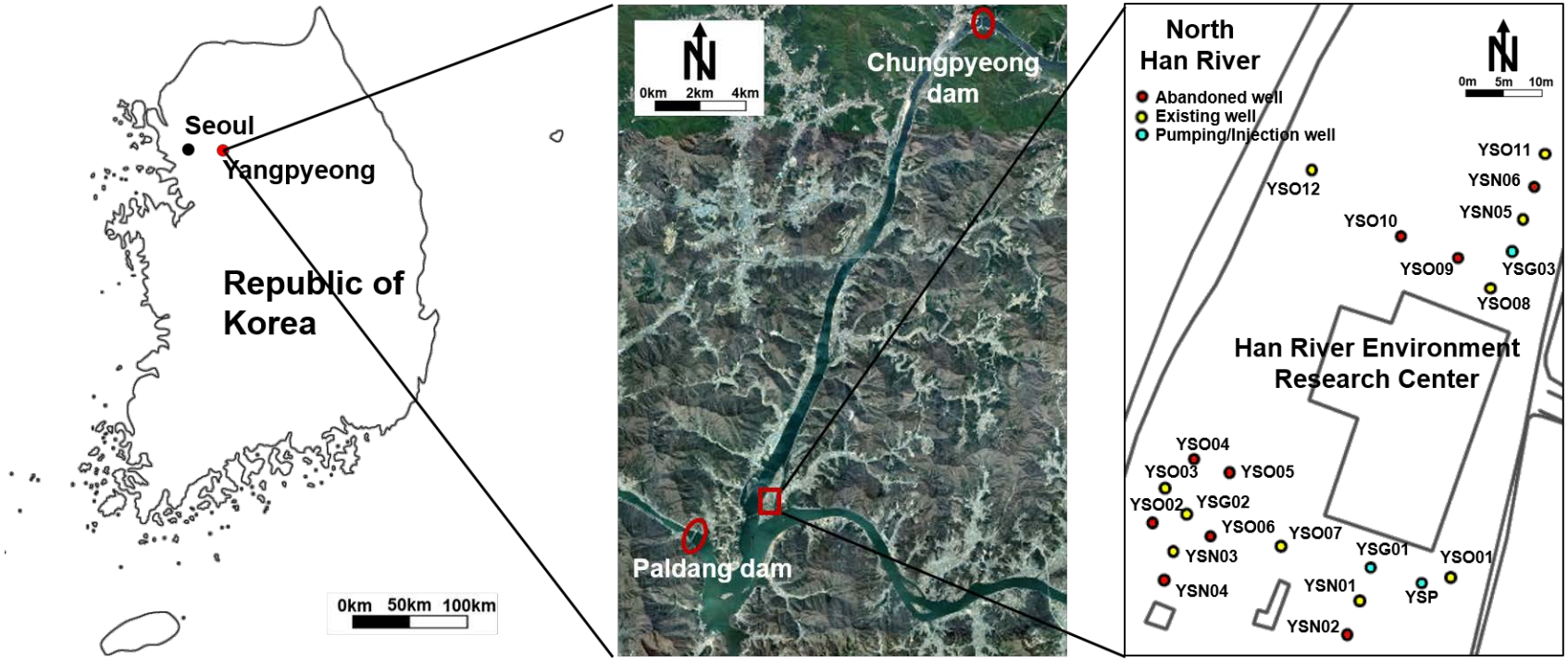

Fig. 2-1. Location of the study area.

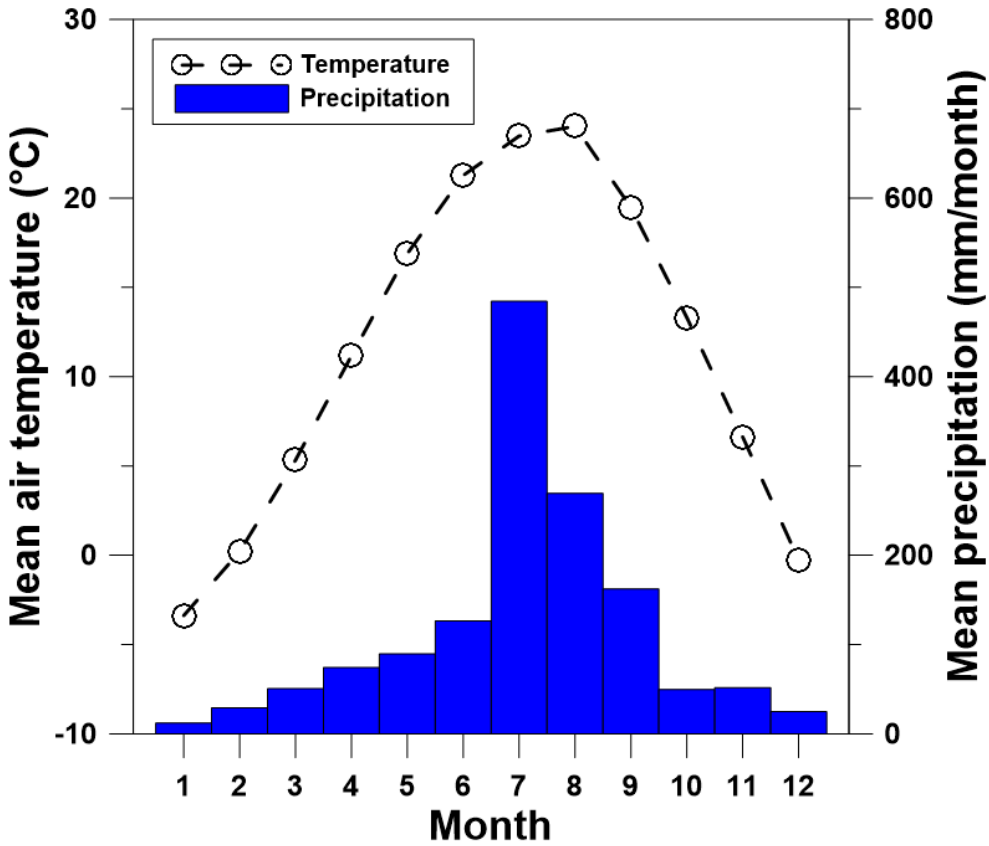

Fig. 2-2. Climatic data in Yangpyeong for 10 years from 2007 to 2016.

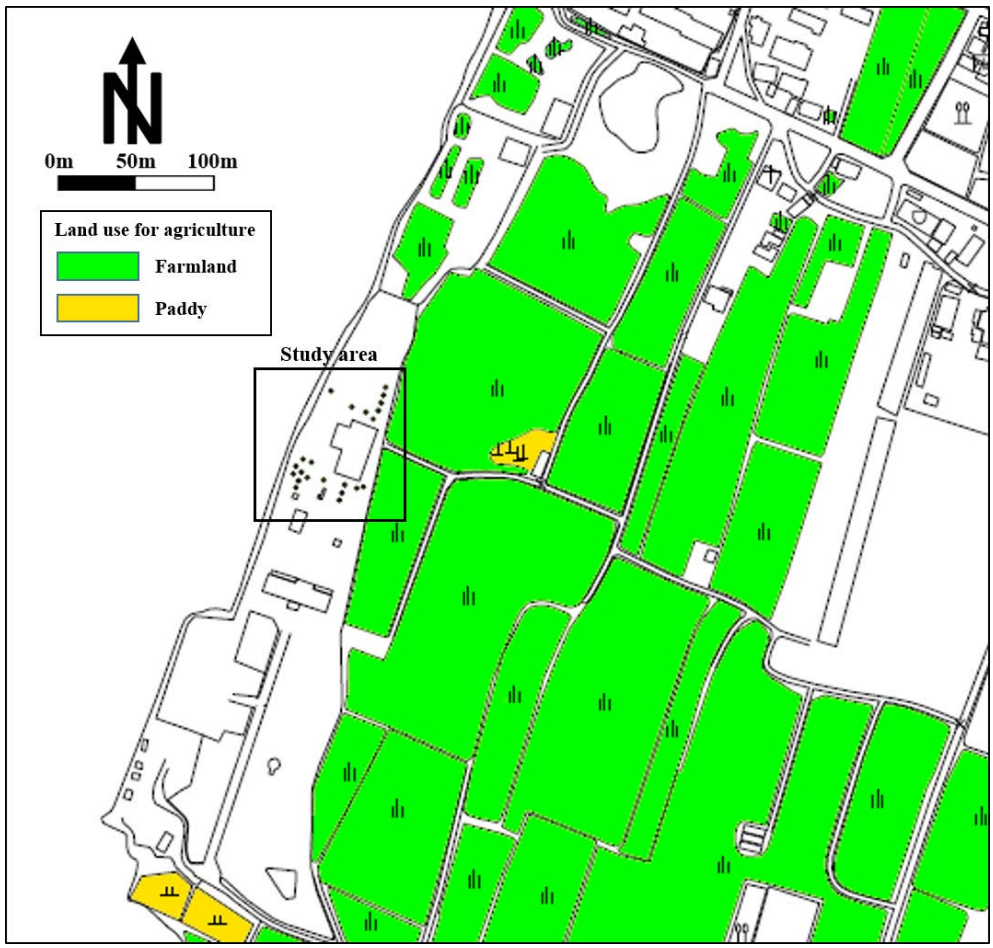

Fig. 2-3. Land use map for agriculture.

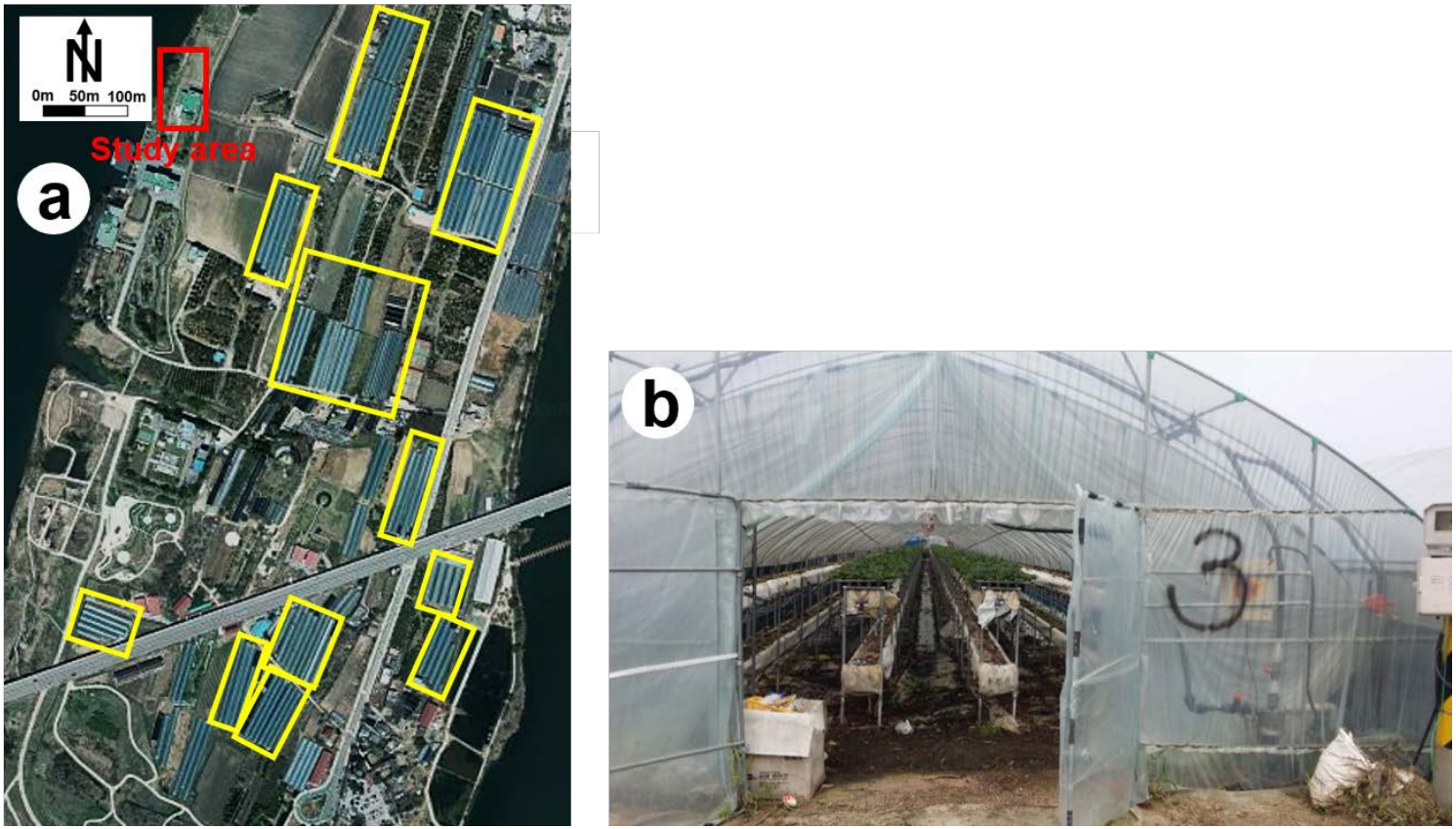

Fig. 2-4. (a) Location and (b) picture of the greenhouse performing the WCC.

2.2 Hydro-geological characteristics

In Yangsu-ri, which has an area of 8.27 km^2, total amount of water resource is $10{,}687\times10^3$ m^3/year. Groundwater recharge rate and development available groundwater are $1{,}926\times10^3$ m^3/year and $1{,}467\times10^3$ m^3/year. There are 137 facilities for use of groundwater in Yangsu-ri, 127 facilities are for agriculture and 10 facilities are for living respectively (Ministry of Land, Infrastructure and Maritime Affairs, 2008).

According to geological logs obtained from installed geothermal and monitoring wells for GWHPs, subsurface of the study area is majorly composed of silty sand, gravel and gneiss from top to bottom. Four lines are shown in Fig. 2-5, and cross sections along each line are represented in Fig. 2-6~Fig. 2-9, respectively. Top silty sand layer exists at about 0 to 10 m distance from the surface. The gravel layer is under the silty sand layer with 5 to 6 m thickness. Bedrock which mainly consists of banded gneiss belongs to Precambrian gyeonggi gneiss complex exists below the silty sand/gravel layer with a depth of 16~17 m from the surface (Ministry of Land, Infrastructure and Maritime Affairs, 2008).

According to cross sectional geologic maps across the study area, the alluvial layer with depth of 2 m to 16~18 m depth is divided into the silty sand layer and the gravel layer. Park et al. (2015) evaluated the hydraulic conductivity of the silty sand layer from laboratory and that of the gravel layer using a pumping test. The evaluated mean hydraulic conductivity of the silty sand is 3.962×10^{-5} m/s and the hydraulic conductivity of the gravel layer is 4.432×10^{-3} m/s which acts as a main aquifer in the study site (Park et al., 2015). An aquifer type of the study area corresponds to an unconfined aquifer, which distributed at the alluvial layer with depth to water of

about 2 m to 16~17 m. Because groundwater in this area is hydraulically connected to the Han river directly, water levels of the groundwater and the surface water have similar aspects to each other (Ministry of Land, Infrastructure and Transport, 2017). Table 2-1 and Table 2-2 show the basic statistical information of the surface water level and the groundwater level (measured at the lake Paldang and YSO12, respectively) for month from March 2014 to December 2015. In Fig. 2-10, maximum, mean, minimum values of the surface water level and the groundwater level are plotted on the same graph. The monthly mean groundwater level always was located at higher elevation than that of the surface water with difference in range of 0.26~0.61 m, which means that the groundwater flows towards the surface water body, mostly. Difference between the surface water level and the groundwater level is larger in summer (about 0.5~0.6 m) than that in winter season (about 0.2~0.3 m) due to the WCC operation.

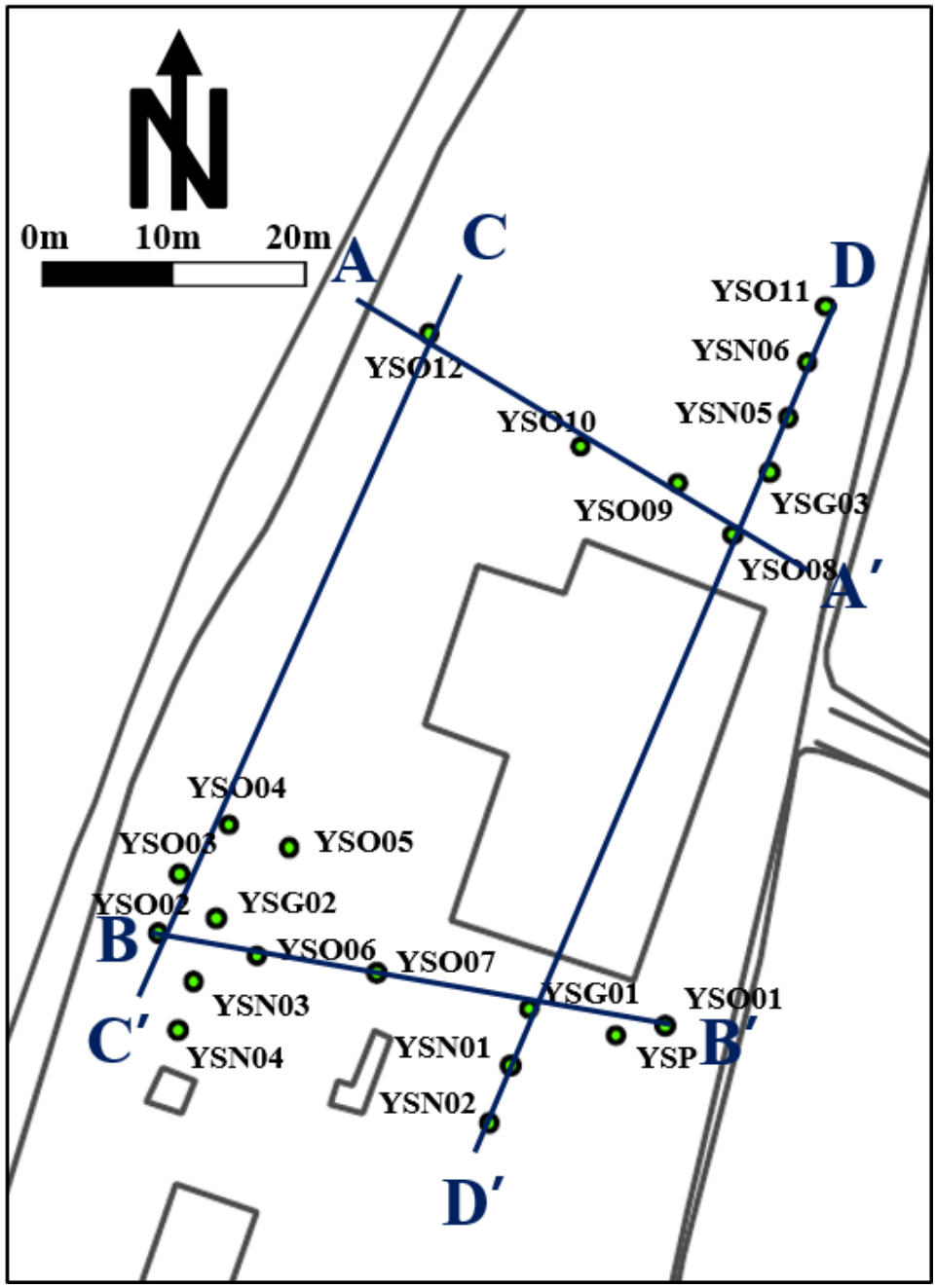

Fig. 2-5. Map of the study area with four lines (A-A', B-B', C-C', D-D').

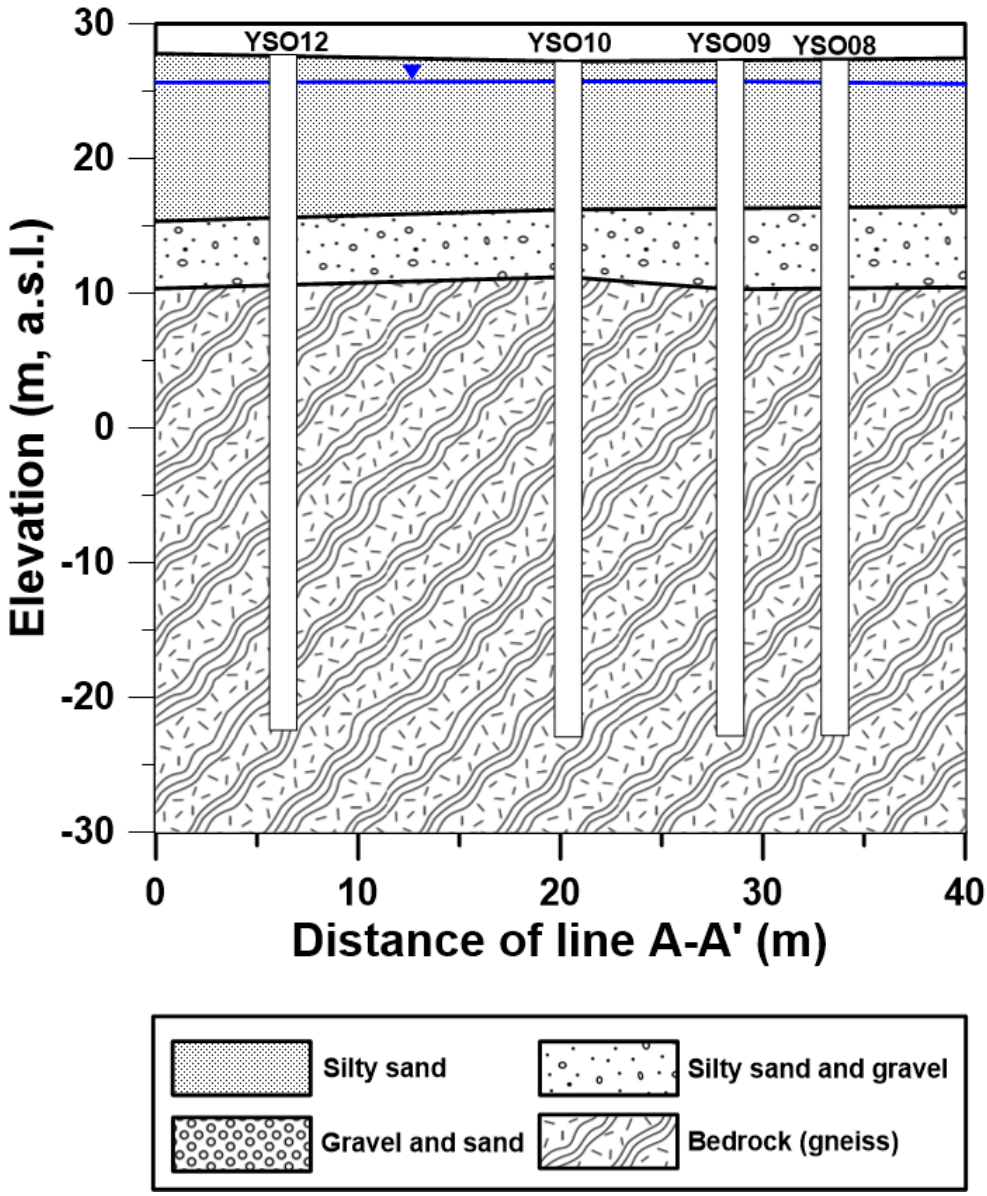

Fig. 2-6. Geological cross section for line A-A'.

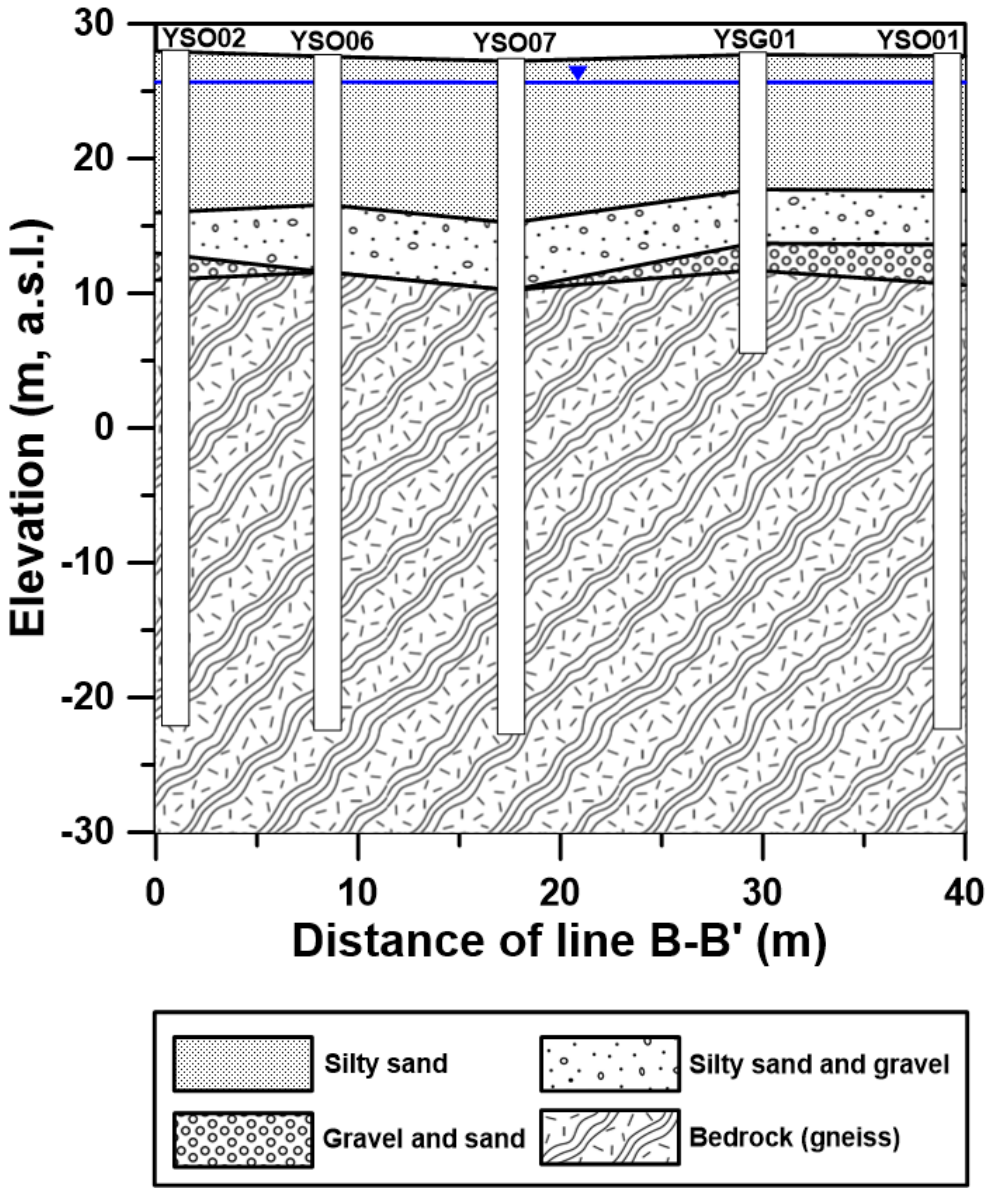

Fig. 2-7. Geological cross section for line B-B'.

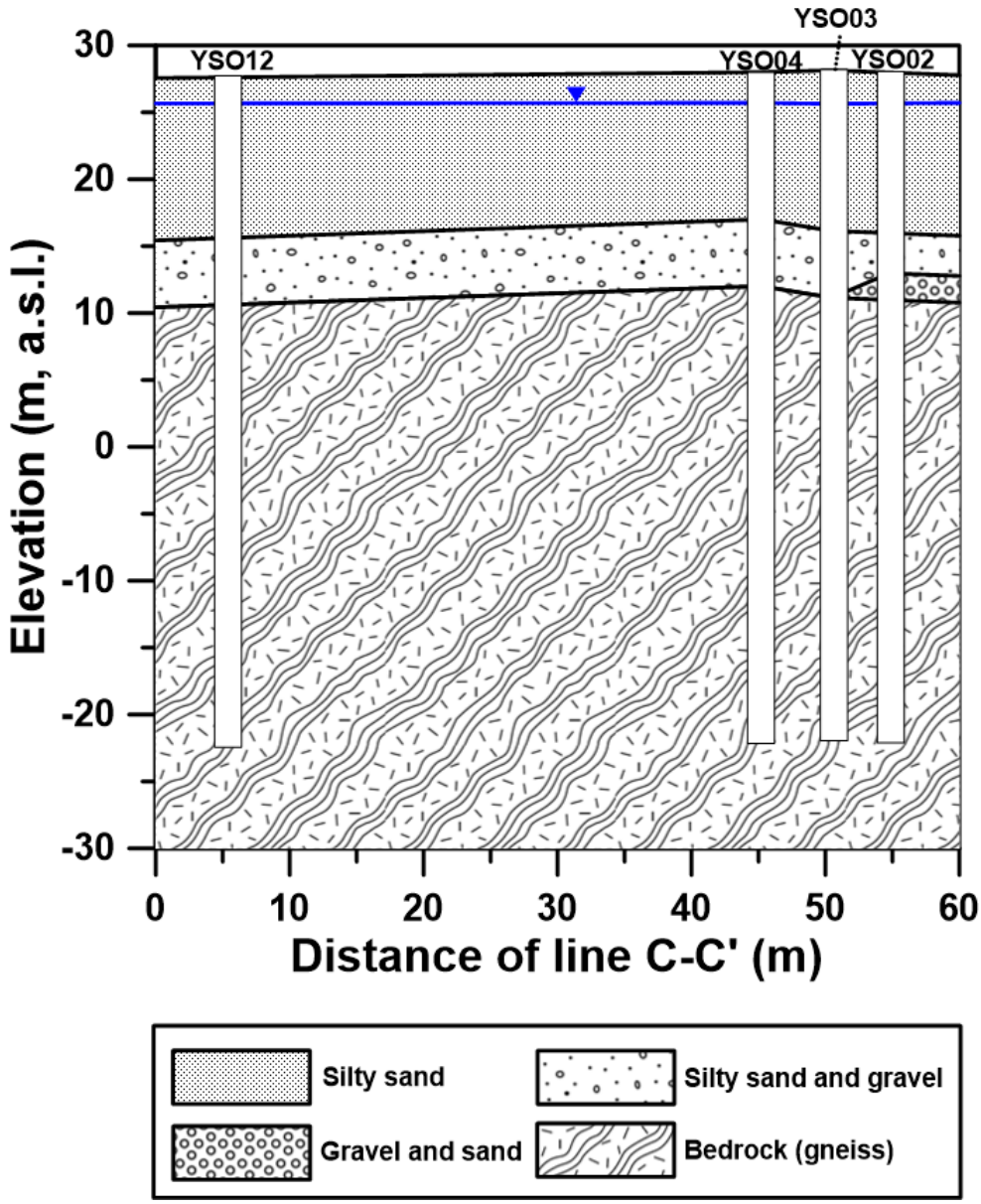

Fig. 2-8. Geological cross section for line C-C'.

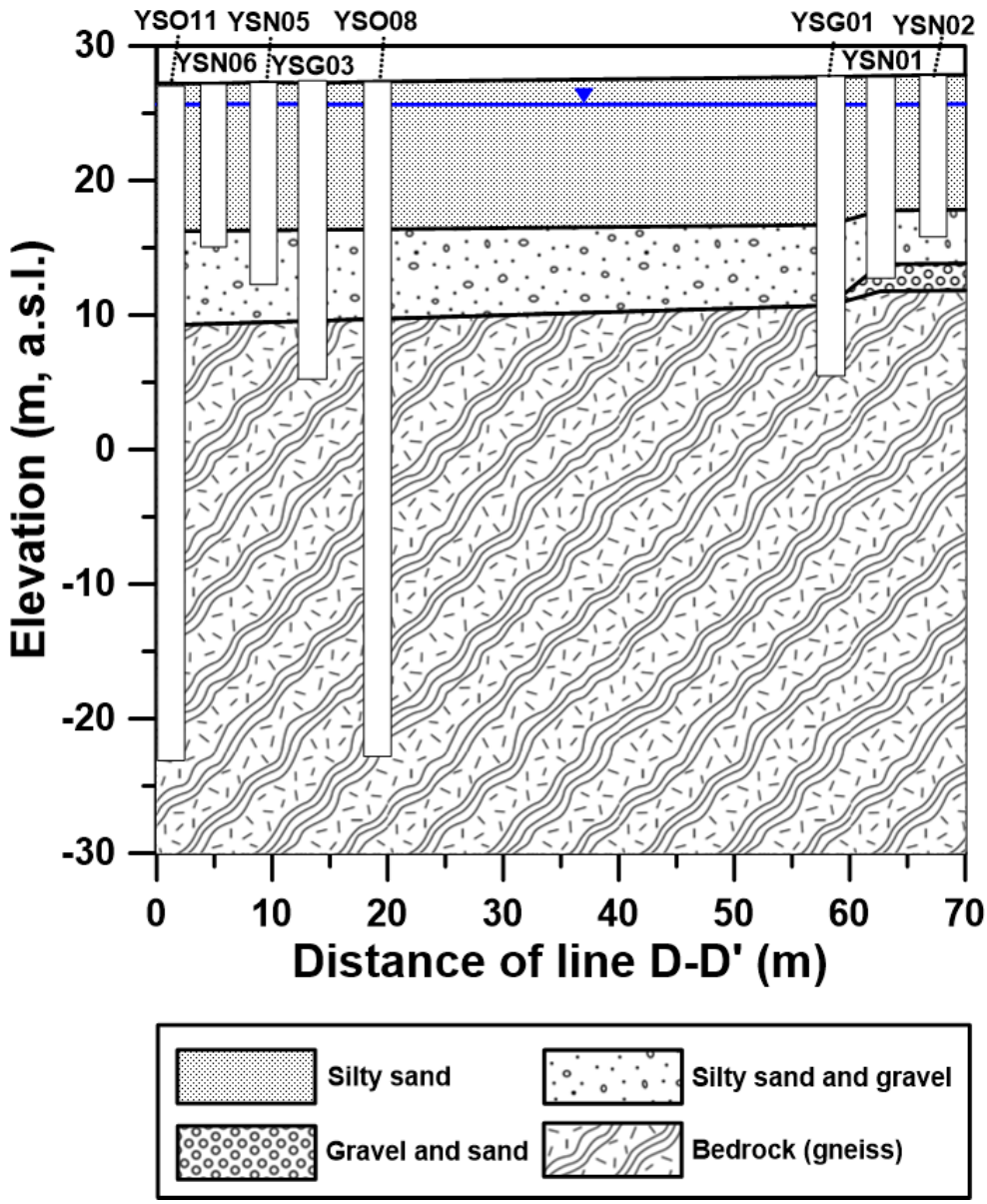

Fig. 2-9. Geological cross section for line D-D'.

Table 2-1 Statistical information of the surface water level.

Time	Minimum (m)	Mean (m)	Maximum (m)	Std. (m)*
14/03	25.00	25.13	25.21	0.04
14/04	24.95	25.13	25.32	0.07
14/05	25.03	25.19	25.35	0.06
14/06	24.95	25.14	25.34	0.08
14/07	24.78	25.09	25.31	0.12
14/08	24.84	25.13	25.35	0.10
14/09	24.74	25.14	25.35	0.14
14/10	25.01	25.22	25.37	0.06
14/11	24.98	25.19	25.36	0.09
14/12	24.85	25.12	25.35	0.12
15/01	24.88	25.05	25.19	0.06
15/02	24.76	25.03	25.38	0.12
15/03	24.93	25.24	25.37	0.10
15/04	24.18	25.22	25.37	0.11
15/05	24.94	25.13	25.37	0.08
15/06	24.95	25.18	25.36	0.09
15/07	24.85	25.15	25.39	0.10
15/08	25.01	25.19	25.37	0.08
15/09	25.09	25.24	25.37	0.06
15/10	25.13	25.30	25.43	0.06
15/11	25.13	25.31	25.44	0.06
15/12	25.12	25.24	25.39	0.06

Std.*: Standard deviation

Table 2-2 Statistical information of the groundwater level.

Time	Minimum (m)	Mean (m)	Maximum (m)	Std. (m)*
14/03	25.28	25.51	25.65	0.09
14/04	25.47	25.62	25.79	0.06
14/05	25.54	25.71	25.84	0.05
14/06	25.49	25.72	25.92	0.09
14/07	25.36	25.61	25.82	0.11
14/08	25.42	25.68	25.87	0.08
14/09	25.31	25.64	25.85	0.11
14/10	25.45	25.69	25.83	0.06
14/11	25.44	25.61	25.83	0.07
14/12	25.11	25.40	25.74	0.15
15/01	25.11	25.32	25.57	0.10
15/02	25.04	25.34	25.76	0.14
15/03	25.43	25.66	25.89	0.12
15/04	25.32	25.72	25.87	0.11
15/05	25.48	25.68	25.91	0.09
15/06	25.46	25.76	25.94	0.11
15/07	25.46	25.76	25.97	0.09
15/08	25.51	25.74	25.92	0.08
15/09	25.51	25.69	25.83	0.06
15/10	25.64	25.78	25.89	0.05
15/11	25.56	25.75	25.90	0.09
15/12	25.32	25.51	25.78	0.09

Std.* : Standard deviation

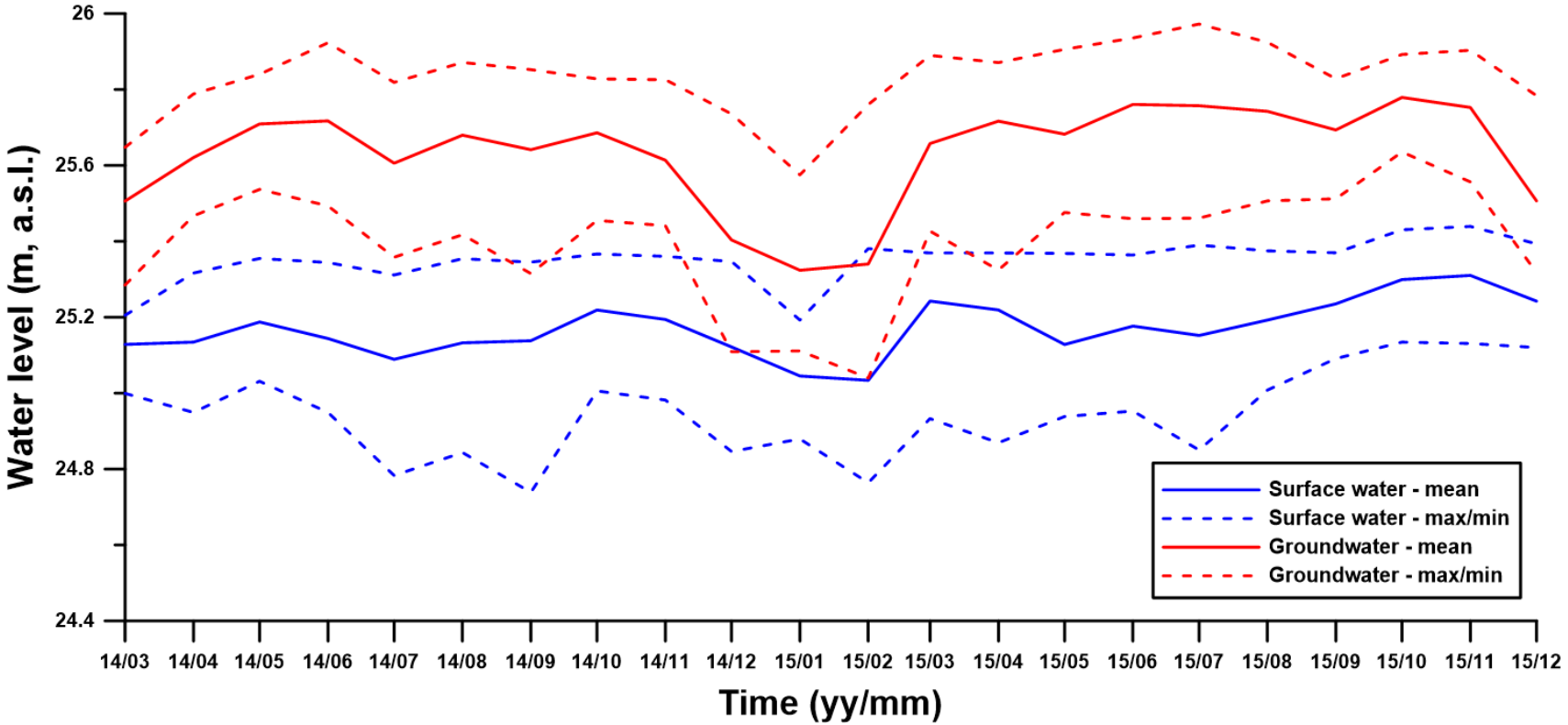

Fig. 2-10. Monthly maximum, mean, and minimum of the surface water level and the groundwater level from March 2014 to December 2015.

3 METHODOLOGY

3.1 Time series analysis

3.1.1 Cross correlation analysis

The cross-correlation is performed to analyze the relation of the groundwater level and various influence factors such as precipitation, evapotranspiration, surface water level and so on (Jeong et al., 2010; Padilla and Pulido-Bosch, 1995). Where time series of the influence factor is x_n, and time series of the groundwater level is y_n, cross-covariance can be expressed as follows.

$$\mathrm{c_{xy}}(t) = \frac{1}{N}\sum_{n-1}^{N-t}(x_n - \bar{x})(y_{n+t} - \bar{y}) \qquad \text{(Eq. 3.1)}$$

where t is the lag time, N is the number of samples in time series data, $\bar{x}$ and $\bar{y}$ are mean value of the x_n and y_n respectively. The cross-correlation function of x_n and y_n with time lag t is

$$\gamma_{\mathrm{xy}}(t) = \frac{c_{xy}(t)}{\sigma_x \sigma_y} \qquad \text{(Eq. 3.2)}$$

The cross-correlation represents the similarity of two time series, although it doesn't mean that they should have relationship actually. When the relationship of two is universal, lag time t refers to the time at maximum cross-correlation coefficient and it means that the influence factor highly affects the other at the time t.

3.1.2 Spectral analysis

Spectral analysis is a method for finding periodic characteristic from the time series data. By conducting Fourier transform, function in the time domain is converted to function in the frequency domain (Davis, 1973). In this study, Discrete Fourier Transform (DFT) is performed because the groundwater level time series data were obtained by sampling frequency of 0.5 minute. The function of DFT is as follows.

$$\mathrm{X_k} = \sum_{n=0}^{N-1} x_n \cdot e^{-2\pi i k n/N} \qquad (k = 0, 1, \cdots, N-1) \qquad \text{(Eq. 3.3)}$$

where x_n is the value of the signal at time n and X_k is the function in frequency domain. Because the value of X_k is composed of the real number and the image number, the absolute values of X_k which are called power spectrum is used for spectral analysis. Peaks in power spectrum represent the periodic characteristic of the time series data. Period of the time series data can be calculated as following equation.

$$\mathrm{T_n} = 1/f_n \qquad \text{(Eq. 3.4)}$$

where T_n is the period corresponding to the nth peak, and f_n is the frequency of the nth peak.

3.2 Artificial Neural Network (ANN)

In this study, the groundwater levels were predicted using an artificial neural network (ANN) model. The ANN is a mathematical structure patterned after the parallel processing of the human brain. In the present study, a feed forward network trained by a back propagation algorithm (Rumelhart et al., 1986) is employed to construct an ANN model. Descriptions for the mathematical aspects and detailed concepts of ANN are as follows (ASCE, 2000a; Daliakopoulos et al., 2005; Yoon et al., 2016; Yoon et al., 2015).

3.2.1 Feed Forward Network

In general, the feed forward network is composed of three layers: input layer, hidden layer, output layer. Each layer has a certain number of nodes and each node in a layer is connected to other nodes in the next layer with a specific value of weight and bias (Fig. 3-1). Fig. 3-2 represents a typical feed forward network with one hidden. Mathematical expressions of signal transmission in feed forward network are as follows.

$$\mathrm{y_h} = f_y\left(\sum_{i=1}^{ni} w_{ih}x_i + b_h\right) \qquad \text{(Eq. 3.5)}$$

$$\mathrm{z_o} = f_z\left(\sum_{h=1}^{nh} w_{ho}y_h + b_o\right) \qquad \text{(Eq. 3.6)}$$

where x_i, y_h, and z_o represent the *i*th, *h*th and *o*th nodal value in input layer, hidden layer and output layer respectively, *ni* and *nh* denote the number of nodes in input layer and hidden layer respectively, w_{ih} *and* w_{ho} represent the weight connecting *i*th

input node and *h*th hidden node, respectively. b_h and b_o are bias of *h*th hidden node and *o*th output node, respectively. f_y is activation function in hidden layer, and f_z is the activation function in output layer. In this study, log-sigmoid function was used as an activation function of hidden layer to reflect nonlinearity, and linear function was applied as an activation function of output layer. Equation of log-sigmoid function is expressed as Eq. 3.7.

$$f(x) = \frac{1}{1 + e^{-x}} \qquad \text{(Eq. 3.7)}$$

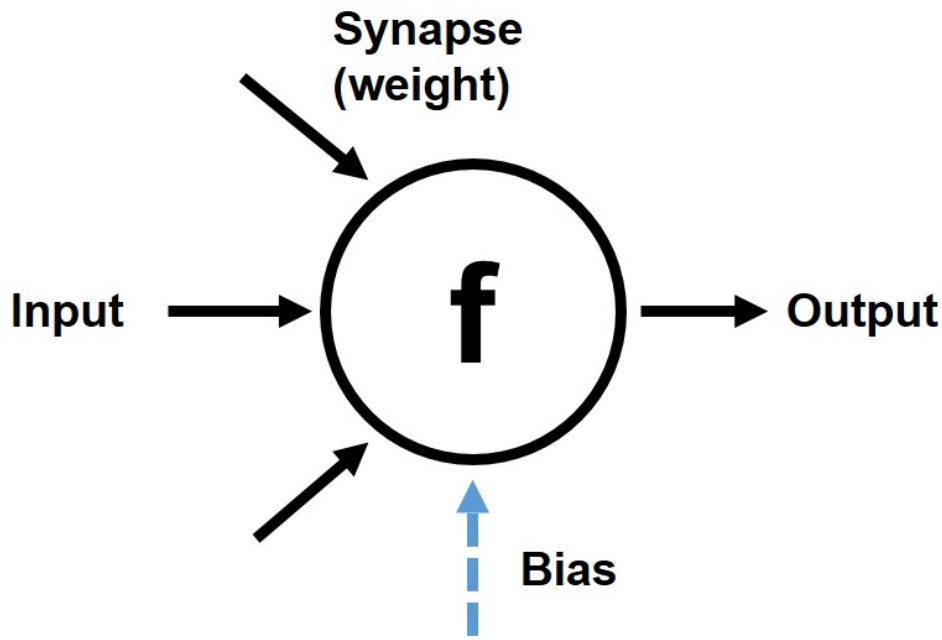

Fig. 3-1. Schematic diagram of a node.

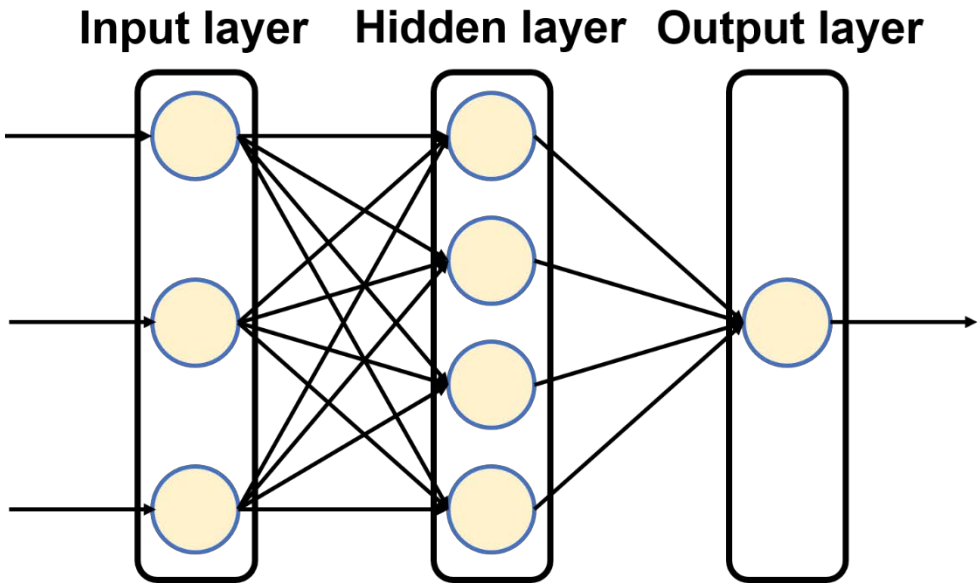

Fig. 3-2. Schematic diagram of the ANN model.

3.2.2 Back Propagation Algorithm

When the weights of nodes are given initially, predicted value can be calculated in order from the input layer to the output layer. Training, also referred to as learning, is a process to find the optimal weights and bias that minimize error between observed value and predicted value that has form of following equation.

$$E^P = \sum_{n=1}^{N} (t_n - z_n)^2 \qquad \text{(Eq. 3.8)}$$

where E^P is the error function of the Pth feed forward process and t_n is the desired target value. The back propagation algorithm updates the weights and nodal bias by gradient descent method using learning rate and momentum. In the opposite direction to feed forward process, training process is practiced from output layer to input layer. At first, the incoming connection weights and the bias value of node in output layer are updated as follows.

$$w_{ho}{}^2 = w_{ho}{}^1 + \alpha \cdot \left(-\frac{\partial E^1}{\partial w_{ho}{}^1} \right) \qquad \text{(Eq. 3.9)}$$

$$w_{ho}{}^{P+1} - w_{ho}{}^P = \beta \cdot (w_{ho}{}^P - w_{ho}{}^{P-1}) + (1-\beta) \cdot \alpha \cdot \left(-\frac{\partial E^P}{\partial w_{ho}{}^P} \right) \qquad \text{(Eq. 3.10)}$$

$$(P \geq 2)$$

$$b_o{}^2 = b_o{}^1 + \alpha \cdot \left(-\frac{\partial E^1}{\partial b_o{}^1} \right) \qquad \text{(Eq. 3.11)}$$

$$b_o{}^{P+1} - b_o{}^{P} = \beta \cdot \left(b_o{}^{P} - b_o{}^{P-1}\right) + (1-\beta) \cdot \alpha \cdot \left(-\frac{\partial E^P}{\partial b_o{}^{P}}\right) \quad (P \geq 2) \qquad \text{(Eq. 3.12)}$$

where, w^P is weight that is updated at *P*th iteration, α is the learning rate and β is the momentum. And then, the incoming connection weights and the bias value of nodes in hidden layer are updated as follows.

$$w_{ih}{}^{2} = w_{ih}{}^{1} + \alpha \cdot \left(-\frac{\partial E^1}{\partial w_{ih}{}^{1}}\right) \qquad \text{(Eq. 3.13)}$$

$$w_{ih}{}^{P+1} - w_{ih}{}^{P} = \beta \cdot (w_{ih}{}^{P} - w_{ih}{}^{P-1}) + (1-\beta) \cdot \alpha \cdot \left(-\frac{\partial E^P}{\partial w_{ih}{}^{P}}\right) \quad (P \geq 2) \qquad \text{(Eq. 3.14)}$$

$$b_h{}^{2} = b_h{}^{1} + \alpha \cdot \left(-\frac{\partial E^1}{\partial b_h{}^{1}}\right) \qquad \text{(Eq. 3.15)}$$

$$b_h{}^{P+1} - b_h{}^{P} = \beta \cdot \left(b_h{}^{P} - b_h{}^{P-1}\right) + (1-\beta) \cdot \alpha \cdot \left(-\frac{\partial E^P}{\partial b_h{}^{P}}\right) \quad (P \geq 2) \qquad \text{(Eq. 3.16)}$$

3.3 Methods to study the contribution of variables in ANN models

Most processes occurring in hydrogeology are intensively complicated and very nonlinear. This is also one of reason why ANN models were appropriate to be used to predict the fluctuation of the groundwater level in this study. As much of studies applying the ANN models have been conducted in various fields as well as hydrogeology, several methods were also developed to quantify variables and to understand the relationship between inputs and the output (Dimopoulos et al., 1999; Gevrey et al., 2003; Goh, 1995; Olden and Jackson, 2002; Sung, 1998; Yao et al., 1998). Two methods (weights method and PaD method) were applied in this study, but they were not proper to spatial comparison and temporal comparison. To overcome these limitation, extraction method to compute the contribution and the relative importance was developed.

3.3.1 'Weights' method

Weights connect each and every node in architecture of the ANN. Through learning process, weights that would work best for the prediction are determined. Using these determined connection weights, a weights method calculates the relative importance for each of input factor, suggesting the importance of the input factor on the output behavior as a percentage (Gevrey et al., 2003; Goh, 1995; Olden and Jackson, 2002). The weights method computes relative importance through following procedure.

(1) Calculate contribution of each input node to the output (B_{ih}) via each hidden node by multiplying the input-hidden connection weight (w_{ih}) by the hidden-

output connection weight (w_{ho}):

$$\mathrm{B_{ih}} = w_{ih} \times w_{ho} \quad \text{(Eq. 3.17)}$$

(2) At a hidden node, relative contribution (R_{ih}) is absolute contribution of each input node divided by sum of absolute contribution of every input node.

$$\mathrm{R_{ih}} = \frac{|B_{ih}|}{\sum_i |B_{ih}|} \quad \text{(Eq. 3.18)}$$

(3) Add up the total relative contributions (R_{ih}) of each input node (U_i).

$$\mathrm{U_i} = \sum_h R_{ih} \quad \text{(Eq. 3.19)}$$

(4) Calculate relative importance of each input factor (RI_i).

$$\mathrm{RI_i} = \frac{U_i}{\sum_i V_i} \times 100 \ \ (\%) \quad \text{(Eq. 3.20)}$$

3.3.2 'PaD' method

Although the ANN architecture is complicated, calculation process from input variables to output can be expressed by a formulation. A PaD method computes the contribution and relative importance by calculating the partial derivatives of ANN output with respect to the input (Dimopoulos et al., 1999; Gevrey et al., 2003). The partial derivatives of the output z_{on} with respect to input x_{in} (n = 1,…,N and N the total number of observations) are:

$$\mathrm{d_{in}} = D_n \sum_{h=1}^{nh} w_{ho} I_{hn} (1 - I_{hn}) w_{ih} \quad \text{(Eq. 3.21)}$$

where, D_n is the derivative of the output node with respect to its input, I_{hn} is the value of the *h*th hidden node for *n*th observation.

For each input variable, SSD (Sum of the Square partial Derivatives) denote the contribution of it.

$$\mathrm{SSD_i} = \sum_{n=1}^{N} (d_{in})^2 \qquad \text{(Eq. 3.22)}$$

3.3.3 ‘Extraction’ method

Extraction method is devised in this study to decompose each influence of input variables in the ANN model. Principle of the extraction method is to compare output derived from controlled data set which contains only one effect of influence factor with corresponding input data.

As a first step, three controlled input data sets were prepared to compute the contribution of each of three influence factors. Each data set is composed of data for three influence factors, but only one data remains intact and the others are fixed to a specific value according to data type. This step is to produce input data set which represents only one influence factor by eliminating effects of the other factors. Every influence data can be classified into two types, type A and B. Data type A has no base value, varying continuously like water level, temperature, and atmospheric pressure. Type B has specific value whenever an event occurs, and has base value if not. Factors such as precipitation rate and pumping rate would belong to type B. Data for type A should be fixed to average value of training data, and data for type B is fixed to its base value in process of eliminating the effect. Among three influence factors in this study, surface water level is classed as type A, while WCC and GWHPs are type B. As a result of this process, Fig.4-17a to c show input data sets which have only effect of surface water level, WCC, and GWHPs, respectively.

Next step is to obtain the output by applying these input data sets to ANN model constructed in section 4.2 for prediction. Three outputs obtained from prediction model refer to pattern of the groundwater level fluctuation formed by each influence factor.

Before calculating the contribution, each input data should be normalized using the minimum and maximum values of each data, so that the variables range between 0 and 1 in order to make it possible to compare different unit or property of input factors. And then, terms for 'strength' and 'frequency' are calculated separately. Mathematical equations of them differ slightly according to data type. Strength term can be expressed as given below.

$$\mathrm{S_i} = \frac{\sum_{n=1}^{N-1}\left|z_{o(n+1)} - z_{on}\right|}{\sum_{n=1}^{N-1}\left|x_{i(n+1)} - x_{in}\right|} \qquad (Type\ \ A) \qquad \text{(Eq. 3.23)}$$

$$\mathrm{S_i} = \frac{\sum_{n=1}^{N}\left|z_{on} - z_{o(base)}\right|}{\sum_{n=1}^{N}\left|x_{in} - x_{i(base)}\right|} \qquad (Type\ \ B) \qquad \text{(Eq. 3.24)}$$

where, z and x are output value and normalized input value, respectively and $z_{(base)}$ or $x_{(base)}$ is the base value when there is no event of corresponding factor. And mathematical expressions of frequency term are given

$$\mathrm{Q_i} = 1 - \frac{n_{0i}}{N-1} \qquad (Type\ \ A) \qquad \text{(Eq. 3.25)}$$

$$\mathrm{Q_i} = 1 - \frac{n_{0i}}{N} \qquad (Type\ \ B) \qquad \text{(Eq. 3.26)}$$

where, n_{0i} is the number that the value of $|x_{i(n+1)}\text{-}x_{in}|$ in type A or $(x_{in}\text{-}x_{i(base)})$ in type B is zero.

The contribution of an influence factor can be computed by the product of strength term and frequency term.

$$\mathrm{C_i} = S_i \times Q_i \qquad \text{(Eq. 3.27)}$$

Relative importance is obtained from proportion of the contribution among total contribution of every influence factor.

$$\mathrm{RI_i} = \frac{|C_i|}{\sum_i |C_i|} \times 100 \ \ (\%) \qquad \text{(Eq. 3.28)}$$

4 RESULTS AND DISCUSSION

4.1 Influence factors for the groundwater level fluctuation

4.1.1 Surface water level

Figure 4-1 shows the surface water level measured at Paldang dam and the groundwater level of YSN01 in 2015. Cross-correlation analyses between the surface water level and the groundwater level measured at every wells except YSG01, YSG02, and YSG03 were performed (Table 4-1). Fig. 4-2 shows the result of cross-correlation of YSN01, a blue line is the cross-correlation between the water level of lake Paldang and the groundwater level measured in 2015. A red dashed line shows the cross-correlation between the water level of lake Paldang and the groundwater level measured from May 2015 to July 2015 when WCC and the GWHPs were not operated. The reason why two periods are distinguished respectively is to verify impacts of the WCC and the GWHPs on the groundwater level. Maximum cross-correlation coefficients between the surface water level and the groundwater level are in range of 0.6053~0.8746 at time lag of 0 hour for whole period, and 0.6834~0.9693 at time lag of 0 day when the WCC and the GWHPs were not operated. Though the surface water level and the groundwater level shows very high correlation for whole period, coefficient of period when the WCC and the GWHPs are not operated is much higher than coefficient of whole period. It means that the groundwater level in the study area is dominantly controlled by the surface water level especially at periods when the WCC and the GWHPs are not operated. However, such basic pattern that the groundwater level follows the surface water level is disturbed when the WCC and the GWHPs are operated. Therefore, the surface water level has to be selected as input variable for prediction of the

groundwater level preferentially, and the WCC and the GWHPs should be considered additionally.

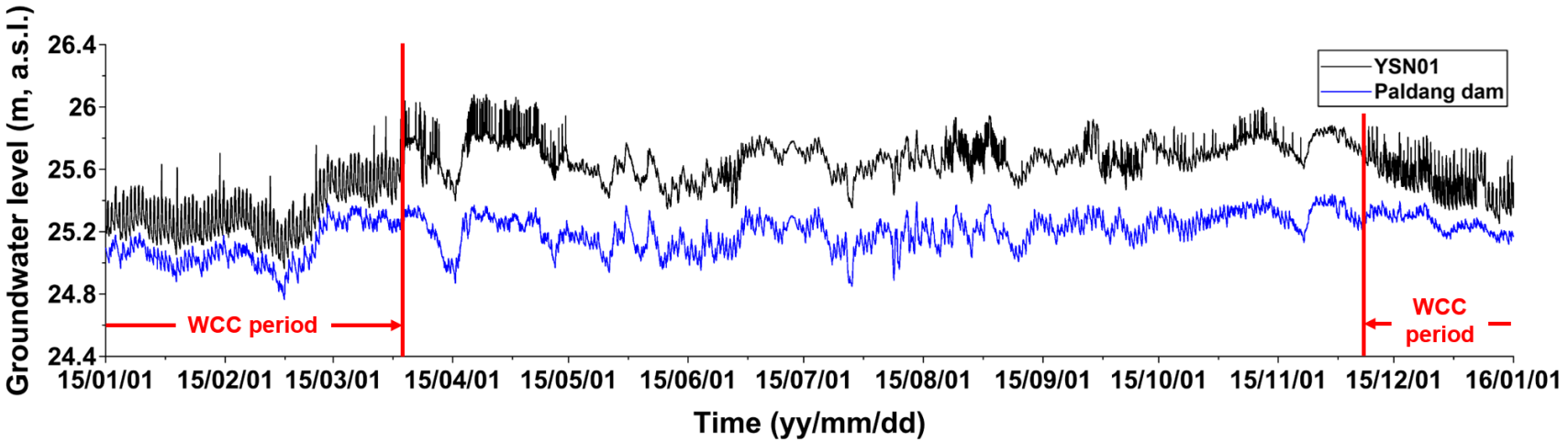

Fig. 4-1. The groundwater level (YSN01) and the surface water level (Paldang dam) in 2015.

Table 4-1 Time lags and maximum cross-correlation coefficients between the surface water level and the groundwater level for whole period and no WCC, GWHPs period.

	Whole period		No WCC, GWHPs period	
	Time lag (hours)	**Coefficient**	**Time lag (hours)**	**Coefficient**
YSO01		0.6684		0.8616
YSO02		0.8490		0.9643
YSO03		0.8746		0.9574
YSO04		0.8400		0.9460
YSO05		0.8270		0.9551
YSO06		0.8077		0.9648
YSO07		0.7888		0.9438
YSO08		0.6735		0.8047
YSO09		0.6248		0.8058
YSO10	0	0.7001	0	0.8963
YSO11		0.6774		0.8247
YSO12		0.7090		0.9175
YSN01		0.7586		0.9206
YSN02		0.7720		0.9287
YSN03		0.8283		0.9491
YSN04		0.8183		0.9693
YSN05		0.6348		0.7625
YSN06		0.6620		0.8111
YSP		0.6053		0.6834

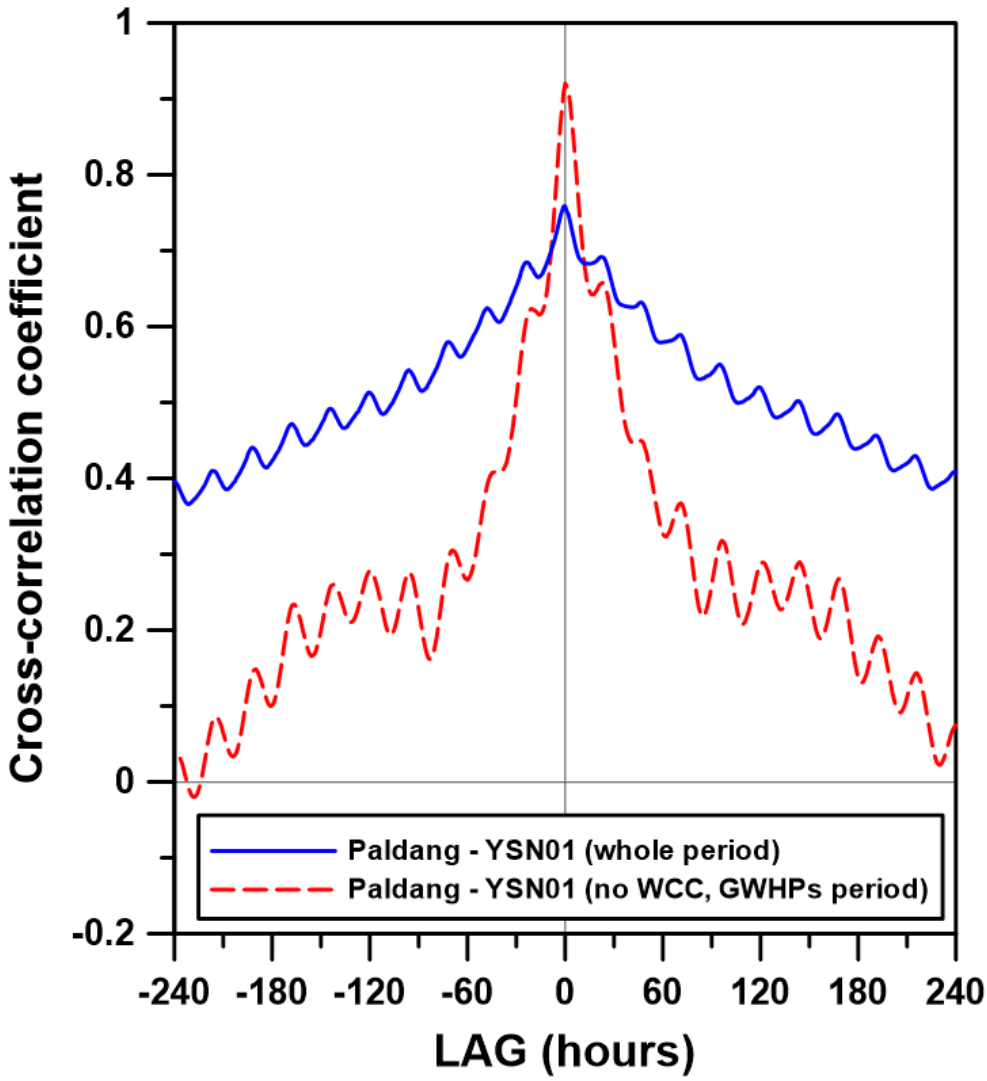

Fig. 4-2. Cross-correlation between the surface water level (Paldang dam) and the groundwater level (YSN01) for whole period and no WCC, GWHPs period.

4.1.2 Water Curtain Cultivation

In the eastern side of the study area, the WCC is practiced from sunset to sunrise in winter season (from middle of November to middle of March) to keep warm in the greenhouse through pumping of large amounts of groundwater. Therefore, effect of the WCC causes the diurnal fluctuation of the groundwater level as shown in Fig. 4-3. Diurnal fluctuation of the groundwater level occurred showing decrease of the groundwater level after pumping at sundown. After sunrise, pumping of the groundwater is stopped and the groundwater level is recovered.

To verify the variability of the groundwater level changes due to the WCC, spectral analysis was performed with data of May 2014 when the WCC was not practiced and December 2014 when the WCC was practiced. For every wells except geothermal wells and test well, power of spectrums corresponding to 24 hours which refer to the cycle of WCC are shown in Table 4-2. Fig. 4-4 represents the result of spectral analysis for YSN01 representatively, red line is power spectrum of data measured in May 2014 and blue line is for December 2014. Power spectrum of December 2014 recorded much higher peak at 24 hours than data of May 2014 for every wells. These results imply that the WCC has a significant effect on the groundwater level with the diurnal cycle.

Because the WCC causes a decline of the groundwater level, the daily mean groundwater level in winter is lower than the other season as mentioned in 2.2. As shown in Fig. 2-10, the daily mean groundwater level in winter season (WCC period) decrease drastically and the gap between the groundwater level and the surface water level is narrowing because the surface water level is not affected by the WCC.

As a result, the WCC makes the groundwater level to fluctuate considerably with diurnal cycle, and causes different pattern of the groundwater level from the surface water level. Therefore, it is sufficiently entitled to be one of input factor in ANN model in addition to the surface water level.

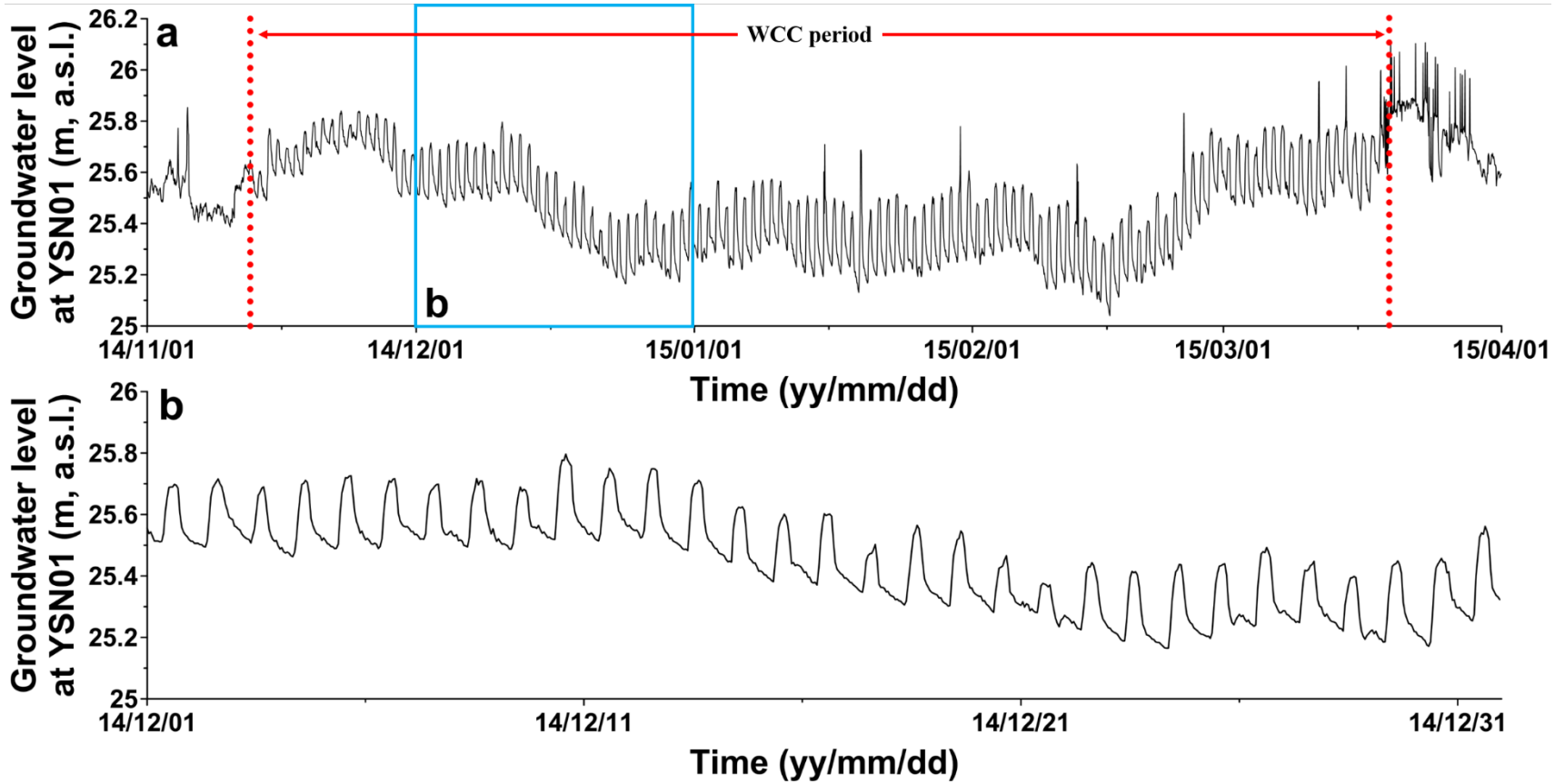

Fig. 4-3. The groundwater level time series data at YSN01 showing the effect of the water curtain cultivation: (a) from November 2014 to March 2015, and (b) in December 2014.

Table 4-2 Power spectrum of cycle of 24 hours for May 2014 and December 2014.

	May 2014	December 2014
YSO01	0.0056	0.0584
YSO02	0.0059	0.0284
YSO03	0.0056	0.0296
YSO04	0.0057	0.0321
YSO05	0.0055	0.0346
YSO06	0.0053	0.0335
YSO07	0.0052	0.0409
YSO08	0.0056	0.0580
YSO09	0.0057	0.0557
YSO10	0.0057	0.0510
YSO11	0.0053	0.0579
YSO12	0.0053	0.0436
YSN01	0.0055	0.0477
YSN02	0.0055	0.0467
YSN03	0.0056	0.0307
YSN04	0.0055	0.0308
YSN05	0.0050	0.0580
YSN06	0.0056	0.0579

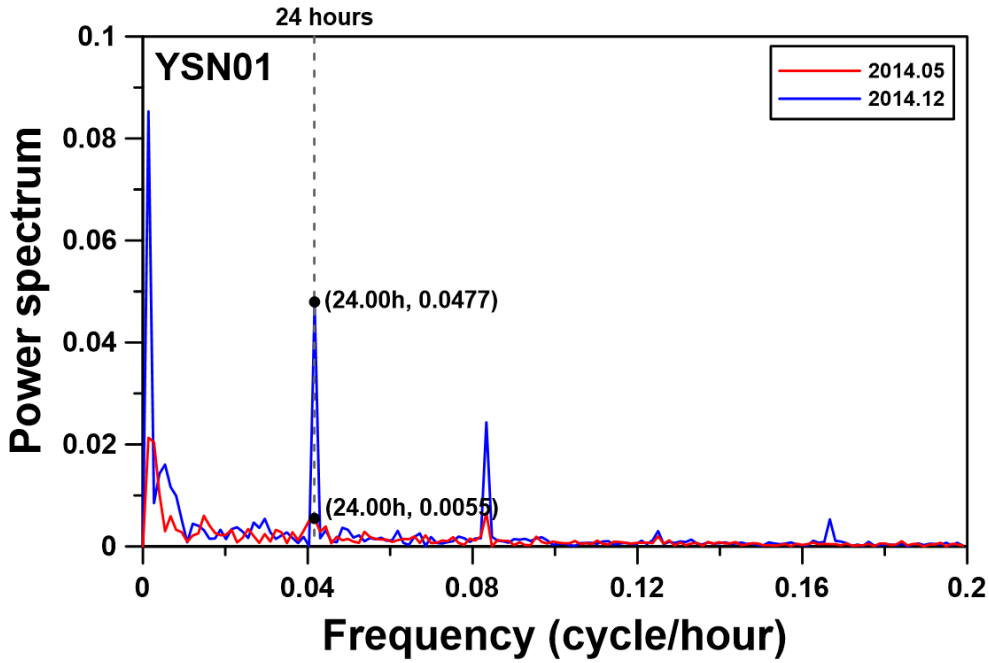

Fig. 4-4. Result of spectral analysis and peaks at cycle of 24 hours for the groundwater level at YSN01.

4.1.3 Groundwater Heat Pump system (GWHPs)

The operation of the GWHPs makes drastic change of the groundwater level in a few minutes and it leads to the increase of the average daily groundwater levels at the wells (YSN01, YSN02, YSN03, YSN04, YSO01, YSO02, YSO03, YSO04, YSO05, YSO06, YSO07, and YSG02 wells) around the pumping well (YSG03) and the decrease of it (YSN05, YSN06, YSO08, YSO09, YSO10, YSO11, YSO12) around the injection wells (YSG01, YSP). Fig 4-5 shows the groundwater level fluctuation response to the operation of GWHPs from April 5 to 15 in 2015 when the GWHPs was operated actively. A blue line is the groundwater level measured at the YSN01 beside the injection well and a red line represents the groundwater level measured at the YSO08 beside the pumping well. Unlike the WCC, whether the groundwater level is increased or decreased by the GWHPs is decided according to the distance from the pumping/injection wells, so it is necessary to identify the influence range for each rise and drop of the groundwater level.

Figure 4-6 illustrates changes of the groundwater level for each wells according to the distance from the pumping/injection well affected by the operation of GWHPs. Magnitudes of changes in the groundwater level diminish as the distance from the pumping/injection wells increases, following the logarithmic function. The groundwater level changes recorded the greatest increase (+0.1773 m) at YSO01 and the largest drop (-0.2587 m) at YSN05. Result shows that the magnitude of changes in the groundwater level around the abstraction well are bigger than the groundwater level around the injection wells in all distance from pumping/injection well. It is because that there are two injection wells, so pumped water is injected with splitting into YSG01 and YSP, while the water is pumped from only one pumping well, even though total amounts of pumped water and reinjected water are definitely same.

From the results, the absolute maximum change of the groundwater level due to the GWHPs is 0.2587 m, and the magnitude of fluctuation is dependent on the well location. Because the groundwater level fluctuation by the GWHPs is not negligible and varies spatially, GWHPs should be used as an input variable.

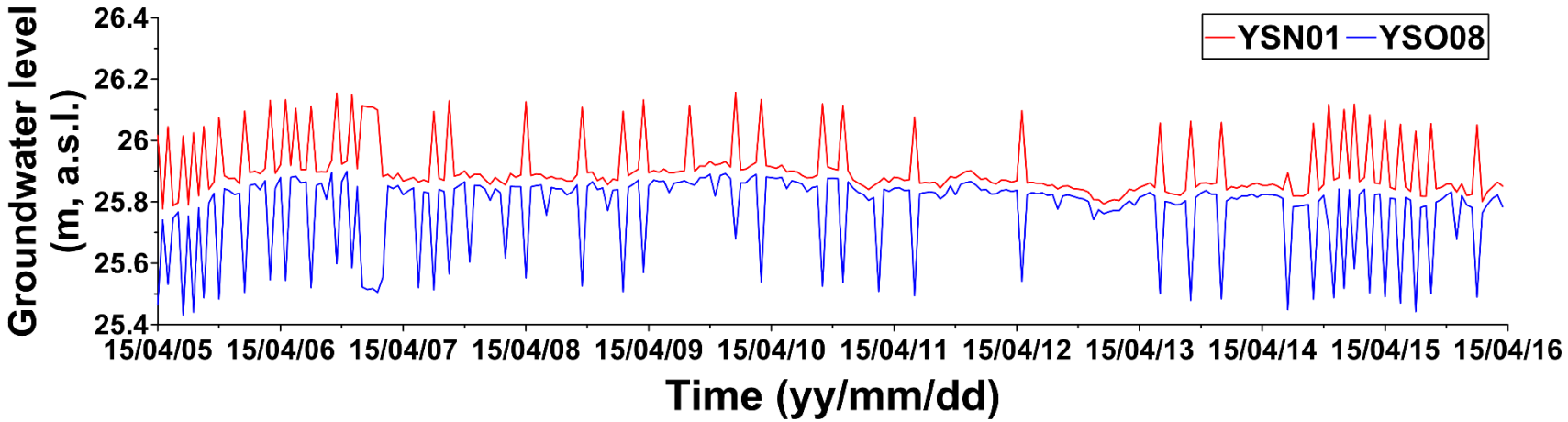

Fig. 4-5. The groundwater level in response to the operation of the groundwater heat pump system from April 5 to April 15 in 2015.

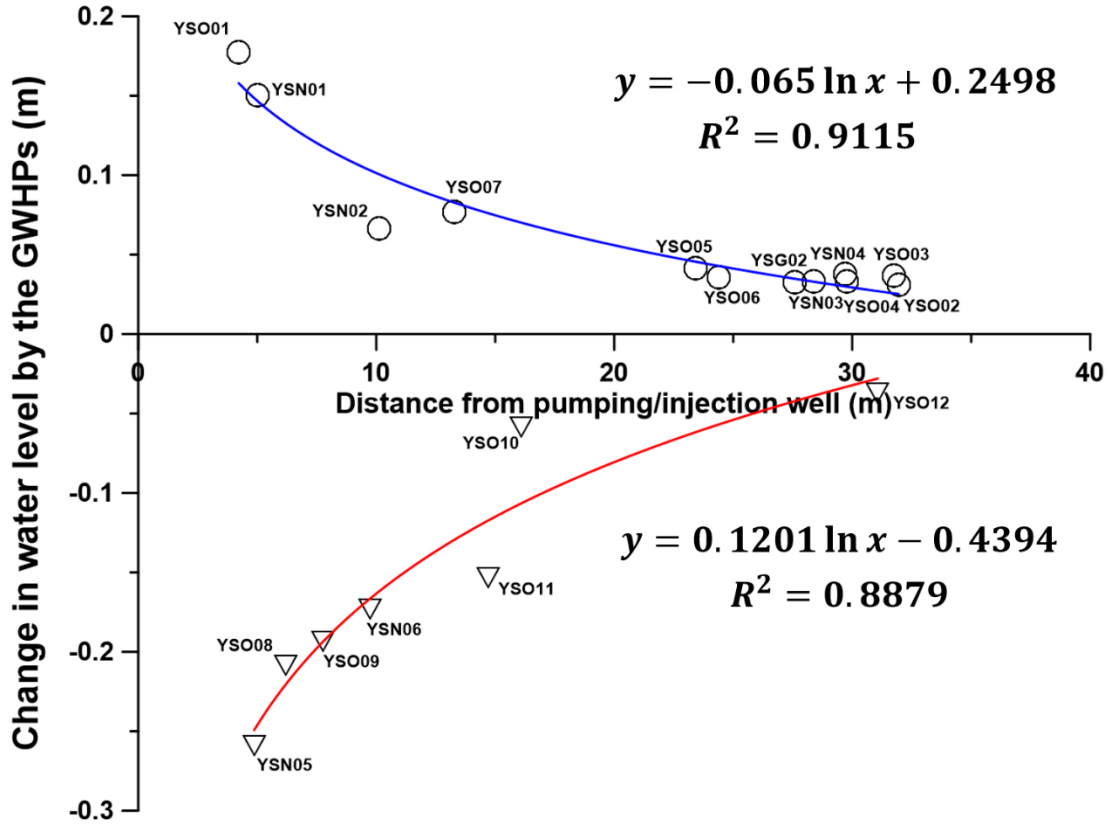

Fig. 4-6. Change in the groundwater level by the groundwater heat pump system according to the distance from pumping/injection well.

4.1.4 Precipitation

Rainfall is one of well-known influence factors that affect the groundwater level especially in an alluvial aquifer (Jan et al., 2007). Fig. 4-7 represents the groundwater level of YSN01 and the precipitation observed from March 8 in 2014 to December 31 in 2014.

To evaluate the relationship between the precipitation and the groundwater level, cross-correlation was performed for the time series data measured at every wells except YSG01, YSG02, and YSG03. Data measured in 2014 when the GWHPs was not operated were applied at first, and then cross-correlation with data of wet season (July, August) in 2014 was additionally performed to evaluate the change of correlation in wet season. The results of cross-correlation are shown at Table 4-3, and Fig. 4-8 represents the result for YSN01 representatively. For all wells, maximum cross-correlation coefficients are in range of 0.1188~0.1646 with a time lag of 5 days in the whole season, and in range of 0.1415~0.2381 with time lag of 4 days in only for the wet season. It could seem that the groundwater level would rise after 4~5 days of the rainfall event and response of the groundwater level to rainfall in the wet season is slightly larger and faster than that in whole season. Because an upper part of the aquifer is composed of alluvial silty sand layer and the groundwater exists about 2 m below the surface, a short lag time and a high coefficient are expected for the relation between the precipitation and the groundwater level of the study area. However, the results of the cross-correlation in this study presented the long time lag and the low coefficient in the wet season as well as in the whole season. Based on the weak relation between the precipitation and groundwater level above, the groundwater level in the study area is unlikely to be affected by precipitation event directly.

Despite a permeable geological condition for infiltrating of rainfall, the reason for the week correlations between the precipitation and the groundwater level can be found at the characteristic of lake Paldang. Water level of lake Paldang is controlled by opening and closing events of the Paldang dam, in which stored water in lake Paldang is discharged to prevent flood whenever it rains. For this reason, the groundwater level doesn't rise by the precipitation event and it rather decreases at heavy rain period as shown in Fig. 4-8. Fig. 4-8 represents negative cross-correlation (-0.13~-0.05) between the precipitation and the groundwater level for wet season at time lag of 0~2 days. Fig. 4-9 represents the precipitation and discharge rate of lake Paldang, demonstrating that the precipitation leads to the opening of the Paldang dam and hardly affects the groundwater level.

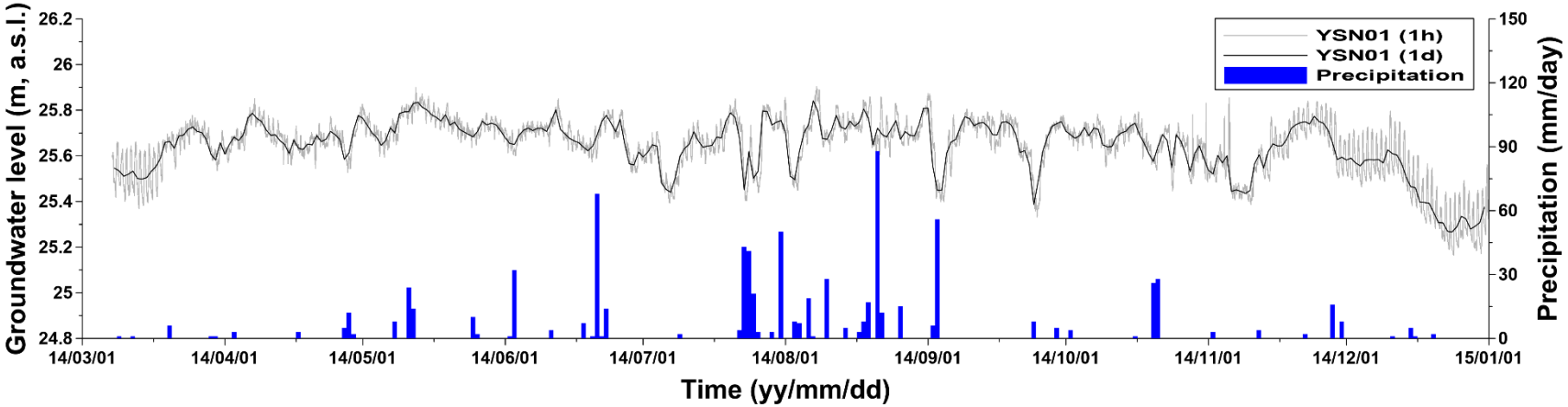

Fig. 4-7. The groundwater level (YSN01) and the precipitation from March 8 to December 31 in 2014.

Table 4-3 Time lags and maximum cross-correlation coefficients between the precipitation and the groundwater level for whole season and wet season.

	Whole season (2014)		Wet season (2014)	
	Time lag (days)	**Coefficient**	**Time lag (days)**	**Coefficient**
YSO01		0.1525		0.2042
YSO02		0.1188		0.1415
YSO03		0.1487		0.1795
YSO04		0.1595		0.1849
YSO05		0.1573		0.2039
YSO06		0.1535		0.2037
YSO07		0.1608		0.2076
YSO08		0.1501		0.2229
YSO09		0.1646		0.2157
YSO10	5	0.1592	4	0.2089
YSO11		0.1533		0.2197
YSO12		0.1572		0.2070
YSN01		0.1530		0.2104
YSN02		0.1611		0.2073
YSN03		0.1468		0.1988
YSN04		0.1562		0.2082
YSN05		0.1608		0.2381
YSN06		0.1593		0.2231
YSP		0.1508		0.2123

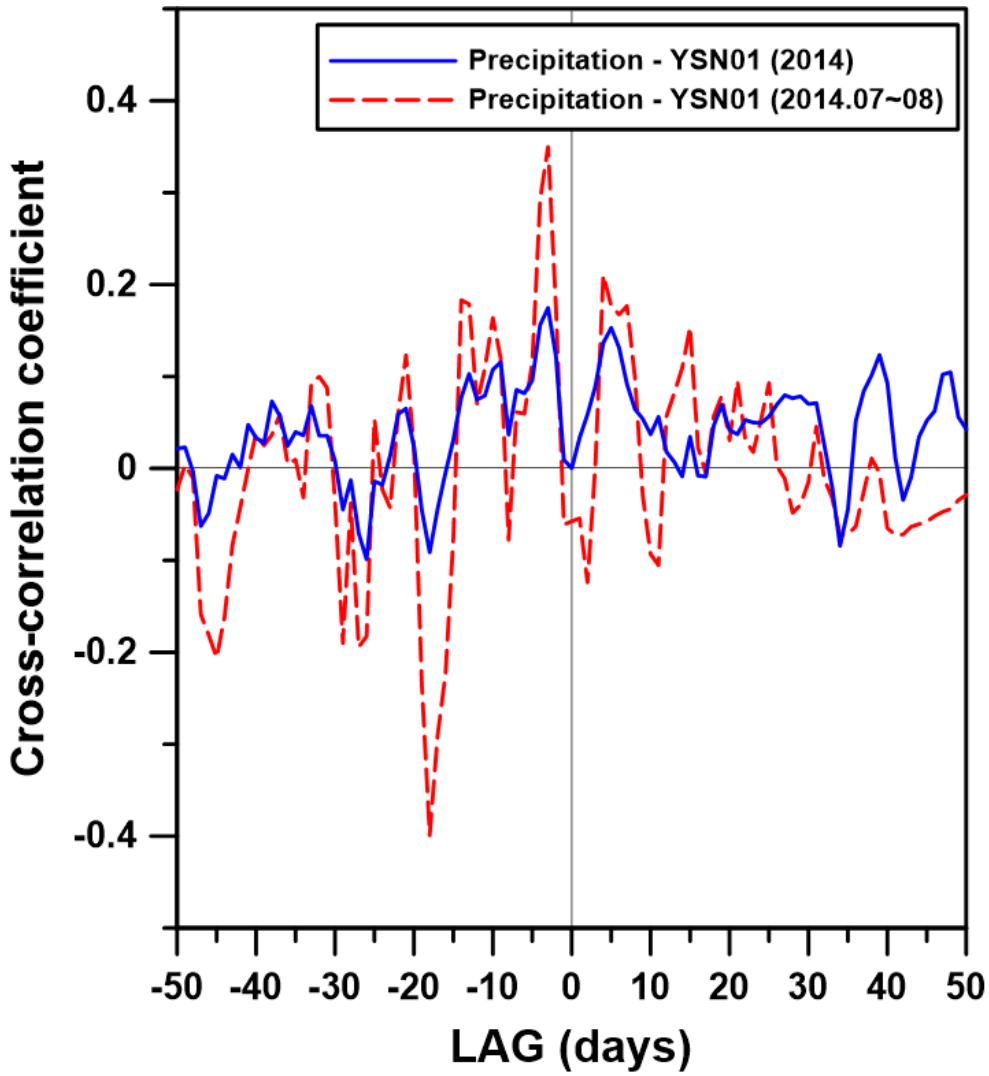

Fig. 4-8. Cross-correlation between the precipitation and the groundwater level at YSN01 for whole season and wet season.

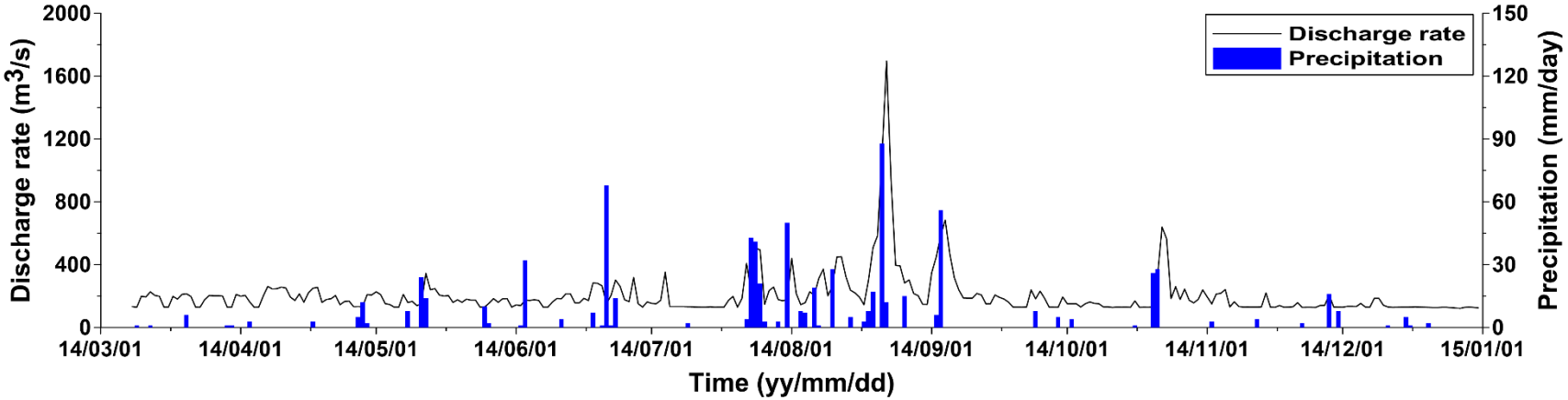

Fig. 4-9. The precipitation and the discharge rate at the Paldang dam.

4.2 Groundwater level forecasting using ANN models

4.2.1 Network and data set

Prior to performing the ANN model, information about network and data set are represented in Table 4-4. Time interval of every input and target data is 1 hour, and train period is from February 1 in 2016 to June 30 in 2016. From October 1 in 2016 to January 15 in 2017 is determined for a test period and from January 16 in 2017 to April 30 in 2017 is for a predict period. The groundwater levels of 8 wells (YSN01, YSN03, YSO01, YSO03, YSO07, YSO08, YSO11, and YSO12) which have sufficient groundwater level data among existing wells are used for the target data. To predict the groundwater level change in the study area, the ANN model with feed forward neural network and back propagation algorithm for training was applied. There are five values for each of three parameters (number of hidden nodes, learning rate, momentum), so total 125 combinations of parameters were applied for determining one combination of parameters leading optimal prediction. From results of 4.1, data related to the surface water level, WCC, and the GWHPs were chosen as the input data set for the ANN model. Figure 4-10 illustrates the input data of them, water level data measured at the Paldang dam was used as input data representing the surface water level. Data for the WCC was set as 1 during pumping groundwater and 0 at no pumping period, assuming pumping rate is constant for every events of the WCC. Data for the GWHPs was obtained from recorded history of the operating time by a heat pump.

Table 4-4 Data sets and network sets of the ANN model.

Data set	Training	Number	3624
		Period	2016. 2. 1 ~ 2016. 6. 30
	Testing	Number	2568
		Period	2016. 10. 1 ~ 2017. 1. 15
	Predicting	Number	2520
		Period	2017. 1. 16 ~ 2017. 4. 30
Network set	Network structure		Feed Forward Network
	Training algorithm		Back Propagation Algorithm
	Number of hidden nodes		4, 5, 6, 7, 8
	Learning rate		0.001, 0.003, 0.005, 0.007, 0.009
	Momentum		0.1, 0.3, 0.5, 0.7, 0.9

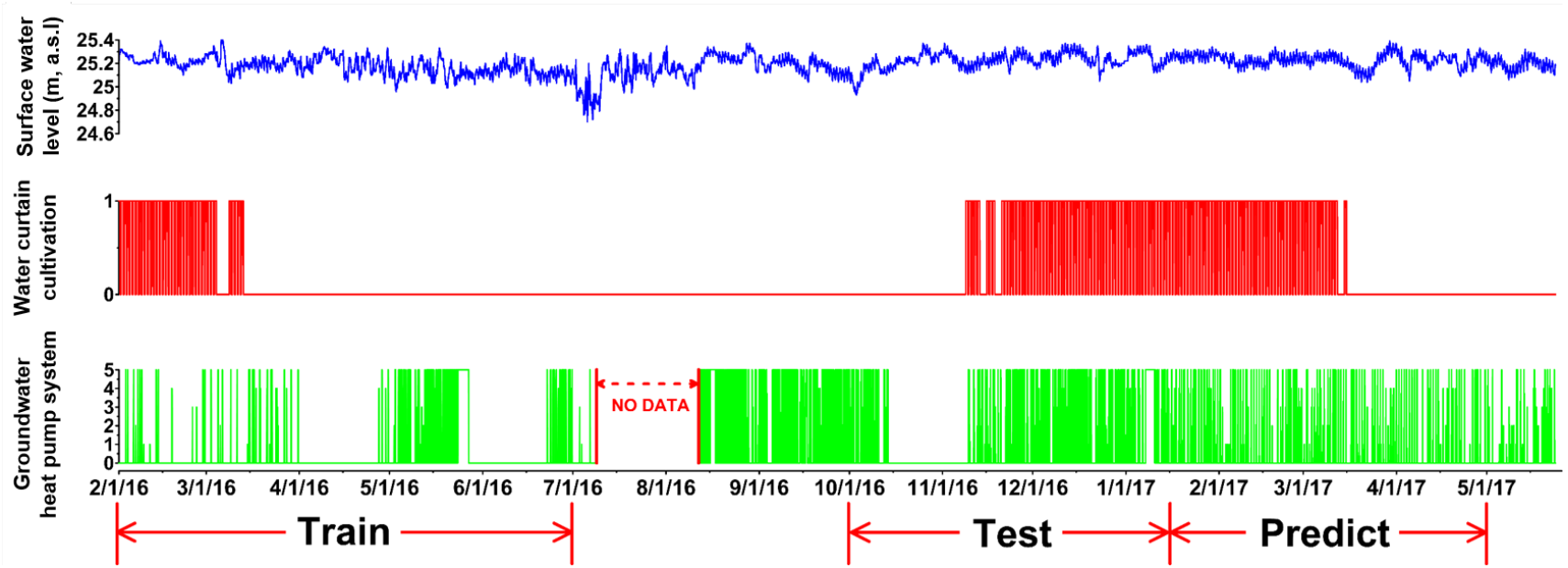

Fig. 4-10. Input data set and the train, test, and predict period for ANN model.

4.2.2 Exploration of model parameters

In the train period, weights and biases in network are all obtained by training algorism for 125 combinations of model parameter. Among these 125 combinations of the parameter, an optimal combination can be determined to make the prediction more accurate. Therefore, exploration of the model parameters for finding the best one was performed in a test period by trial and error method. By comparing RMSE values of estimation using data in the test period, one combination of model parameter which has the lowest RMSE is selected to be applied for a predict period. Exploration of model parameters was conducted for every wells separately.

Figure 4-11 illustrates distribution of the RMSE values according to model parameters, for the case of YSN01 representatively, and Table 4-5 represents the results of parameter exploration for all wells. In Fig. 4-11, RMSE tends to be lower as learning rate and momentum are smaller, while the number of hidden nodes does not regard RMSE value. However, Table 4-5 shows all different determined model parameters depending on the well, meaning RMSE can show various patterns according to the well or location. A combination of model parameter which is selected as a result of exploration for each well would be used for the prediction of the groundwater level in predict period.

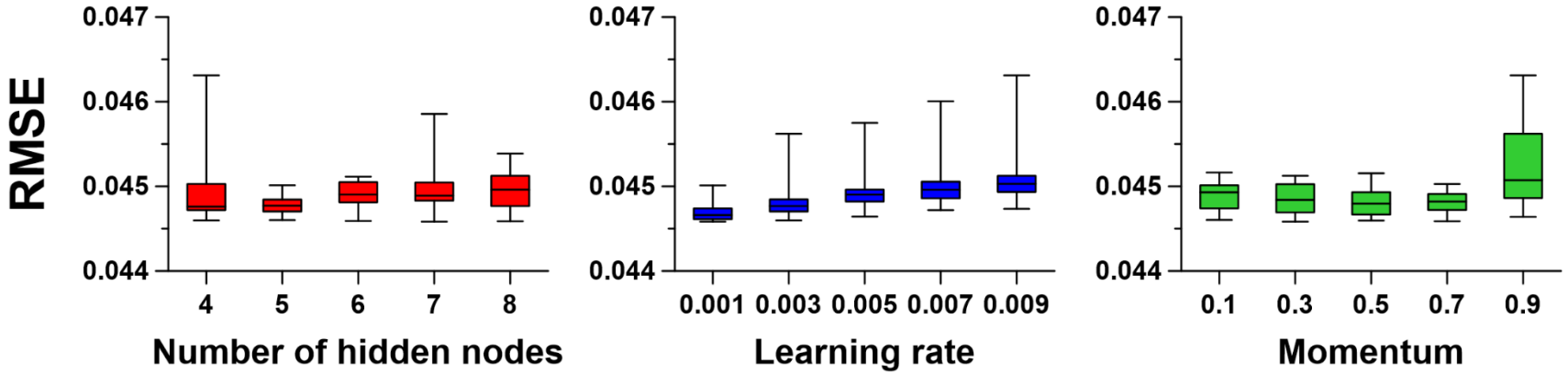

Fig. 4-11. Distribution of RMSE values for 125 combination of three model parameters.

Table 4-5 Results of exploration of model parameters for each well to perform the optimal prediction of the groundwater level.

Parameter	YSN01	YSN03	YSO01	YSO03	YSO07	YSO08	YSO11	YSO12
NH	7	7	6	4	6	7	7	4
LR	0.001	0.003	0.001	0.003	0.001	0.009	0.007	0.003
MM	0.3	0.1	0.5	0.1	0.5	0.5	0.5	0.9

NH: Number of Hidden nodes
LR: Learning Rate
MM: Momentum

4.2.3 Prediction of the groundwater level

After constructing the optimal ANN model through the train and test periods, prediction of the groundwater level at the selected 8 wells was performed using data in the predict period. Fig. 4-12 represents results of the groundwater level forecasting for the 8 wells. Black dot denotes the observed groundwater level, and a red line refers to the estimated value by the ANN model. Although all of 8 wells used same input data set as shown in Fig. 4-10, they led to all different prediction results. It is because that weights and biases from train period, and network architecture from test period were determined differently for each well.

Prediction errors of the 8 wells are represented in Table 4-6. The groundwater levels at 8 wells were predicted with ME (Mean Error) in range of -0.0250~0.0057, with RMSE (Root Mean Square Error) between 0.0331 and 0.0562, with MAPE (Mean Absolute Percentage Error) in range of 0.1070~0.1956, CORR (Correlation coefficient) of 0.9053~0.9221, and NSE (Nash-Sutcliffe Efficiency) of 0.7585~0.7138. Wells around the pumping well and the injection well (YSN01, YSO01, YSO08, YSO11) tend to show rather poor prediction comparing other wells. Moreover, these wells are closer to WCC site than the other wells. These imply that the stronger effects of the WCC and the GWHPs could cause the prediction error to be increased. It is because that the input data for WCC during operating period was set as constant value 1 with assuming consistency of pumping rate. Another reason is that the GHWPs is operated in minutes, and the groundwater level changes so quickly to estimate remarkably accurate level. Despite these bad conditions, results showed acceptable prediction of the groundwater level compared to other studies for groundwater level forecasting using artificial neural network (Mohanty et al., 2010; Taormina et al., 2012; Yoon et al., 2011)

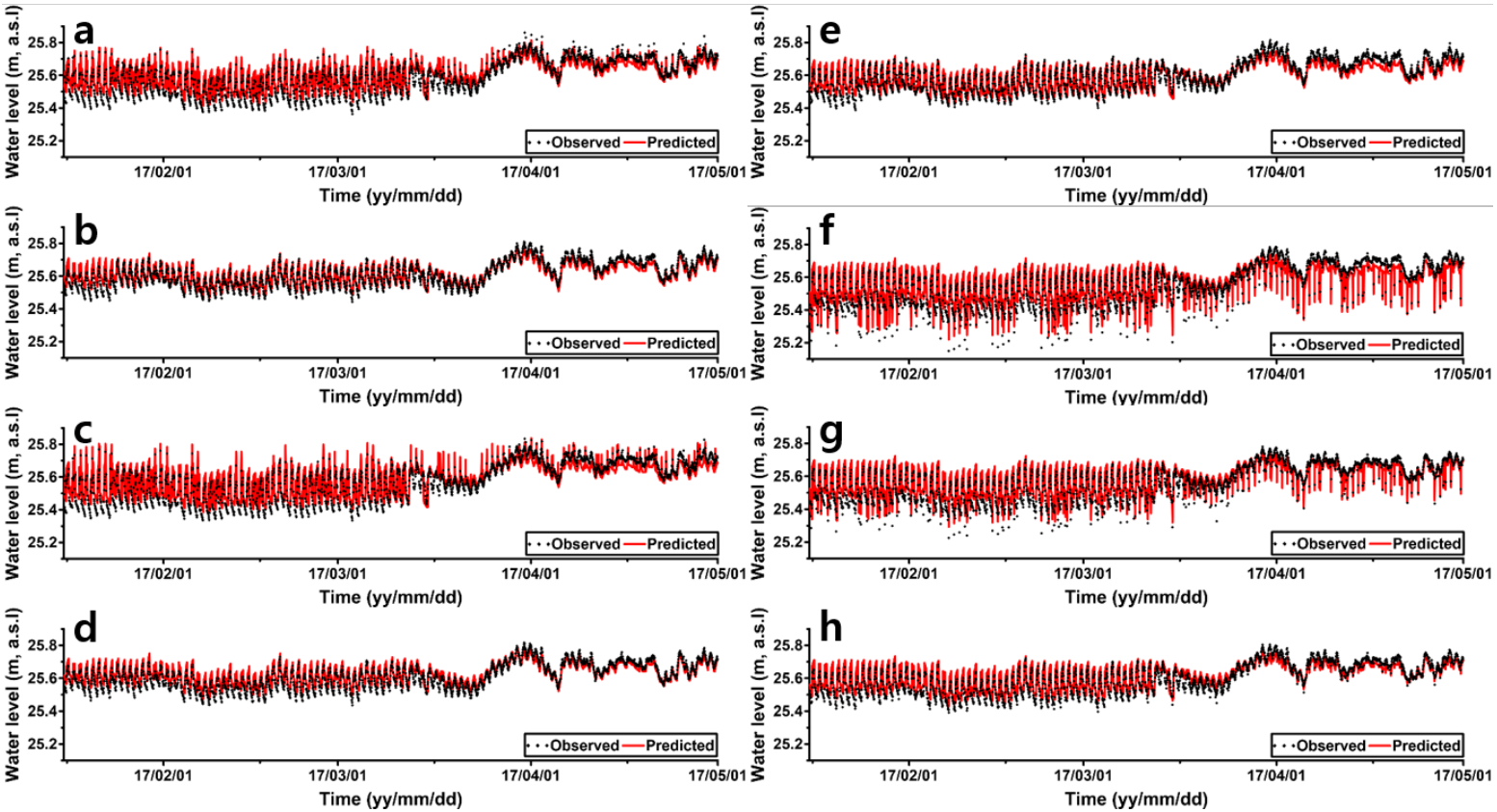

Fig. 4-12. Results of the prediction of the groundwater level at (a) YSN01, (b) YSN03, (c) YSO01, (d) YSO03, (e) YSO07, (f) YSO08, (g) YSO11, (h) YSO12.

Table 4-6 Prediction errors for each well.

Error Index*	YSN01	YSN03	YSO01	YSO03	YSO07	YSO08	YSO11	YSO12
ME (m)	-0.0168	-0.0048	-0.0121	-0.0189	0.0057	-0.0208	-0.0250	-0.0169
RMSE (m)	0.0475	0.0331	0.0517	0.0378	0.0391	0.0562	0.0504	0.0420
MAPE (%)	0.1629	0.1070	0.1739	0.1285	0.1237	0.1956	0.1632	0.1338
CORR	0.9105	0.9083	0.9053	0.9181	0.9086	0.9165	0.9221	0.9151
NSE	0.7872	0.8105	0.8039	0.7585	0.8138	0.7955	0.7872	0.7913

ME: Mean Error
RMSE: Root Mean Square Error
MAPE: Mean Absolute Percentage Error
CORR: Correlation coefficient
NSE: Nash-Sutcliffe Efficiency

4.3 Computing contribution of influence factors

Contribution can be interpreted in various ways according to the purpose of use or types of data. Therefore, it was necessary to establish a meaning of the contribution for application in this study. Contribution refers to how much a factor affects the target in general, but for time series data, it should also contain the meaning of frequency, i.e. how often a factor affects the target. If obtained contributions satisfy the concept mentioned above, they will be able to be compared spatially and temporally, and this will help understanding features of the groundwater level and influence factors in the study area. Decomposition of influence factors causing the groundwater level fluctuation complexly can make easy to know the effects of influence factors and to manage them. Previous studies have developed several methods to compute the contribution or relative importance of input data in ANN model already, among them, following two methods were applicable for calculation of contribution with this study data (Dimopoulos et al., 1999; Gevrey et al., 2003; Goh, 1995; Olden and Jackson, 2002).

4.3.1 Results of weights method

Figure 4-13 illustrates results of weights method for every wells. Results of eight wells shows different relative importance to each other, but insignificant differences between three influence factors in almost wells are generated. In other word, there are minor differences between relative importances of three influence factors, so features of them are not noticeable. Furthermore, they reveal significant defect in spatial comparison. In comparison between YSO01 and YSO07, the WCC's relative importance of YSO07 is much greater than that of YSO01 even though YSO07 is farther from the WCC site than YSO01. Also, the GWHPs' relative importance of

YSO08 is slightly lower than that of YSO12 though YSO08 is much closer to the pumping well. Results of weights method did not satisfy the suitability for spatial comparison which is one of crucial point in computing the contribution. This limitation of weights method may come from computational procedure. The relative importance is obtained using weights connecting nodes, however, they are just calculated by multiplication and addition. Because there are several components which make the process of signal in the ANN more complicated such as activation functions and biases, sophisticated computing method should be devised considering these points. For these reasons, weights method is not suitable for evaluation of influence factors.

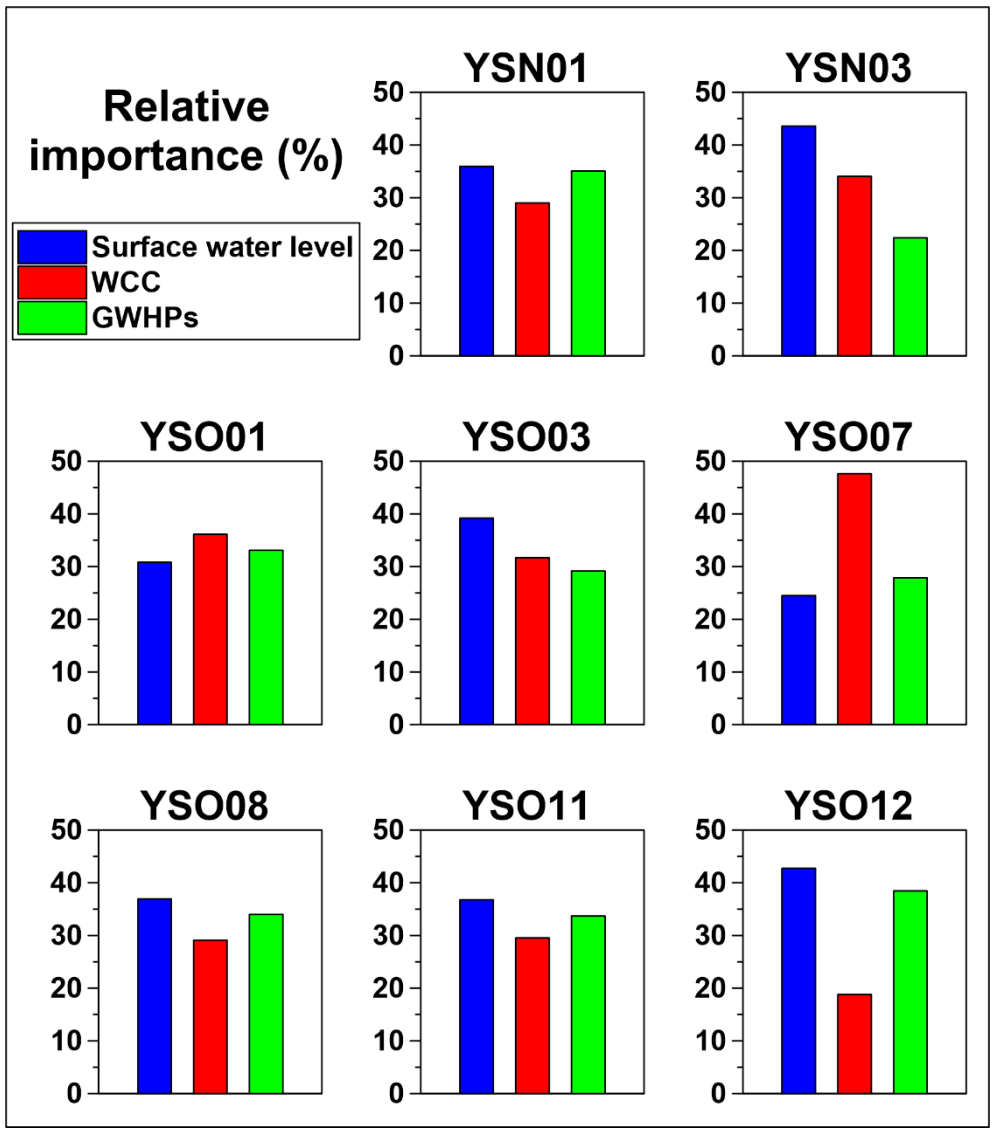

Fig. 4-13. Relative importance obtained by weights method.

4.3.2 Results of PaD method

The output of PaD method is Sum of Square partial Derivatives (SSD) which refers to the contribution of each input factor at every well, the results of PaD method are shown in Fig. 4-14. Comparing with weights method, there is standardized pattern which represent the characteristic of the study site. Effect of surface water level is the highest in every wells, and it seems that values of SSD vary according to well location. Especially, YSN03 and YSO03 which are located closely to the river have much higher contribution of the surface water level than that of the other wells. Also, wells around pumping well or injection well (YSN01, YSO01, YSO08, YSO11) have some effect of GWHPs while SSD values of the other wells are too low. In these respects, PaD method is more reliable to compare spatially than weights method.

However, PaD method also has a limitation in temporal comparison. Fig. 4-15 represents input data only for the prediction period, showing big difference in WCC data around middle of March. Because the WCC is not operated after the middle of March, there is only zero in input data for WCC in that period. To verify that this difference is reflected in results of PaD method, SSDs for February and April were obtained separately expecting that there is no value of SSD for the WCC in April. Fig. 4-16 illustrates SSD values for the WCC in February and April, but they are very similar to each other. Evaluation of influence factors based on time series data should be possible to compare temporally, because input variables can vary in time. Therefore, PaD method is not suitable for time-periodic evaluation of influence factors.

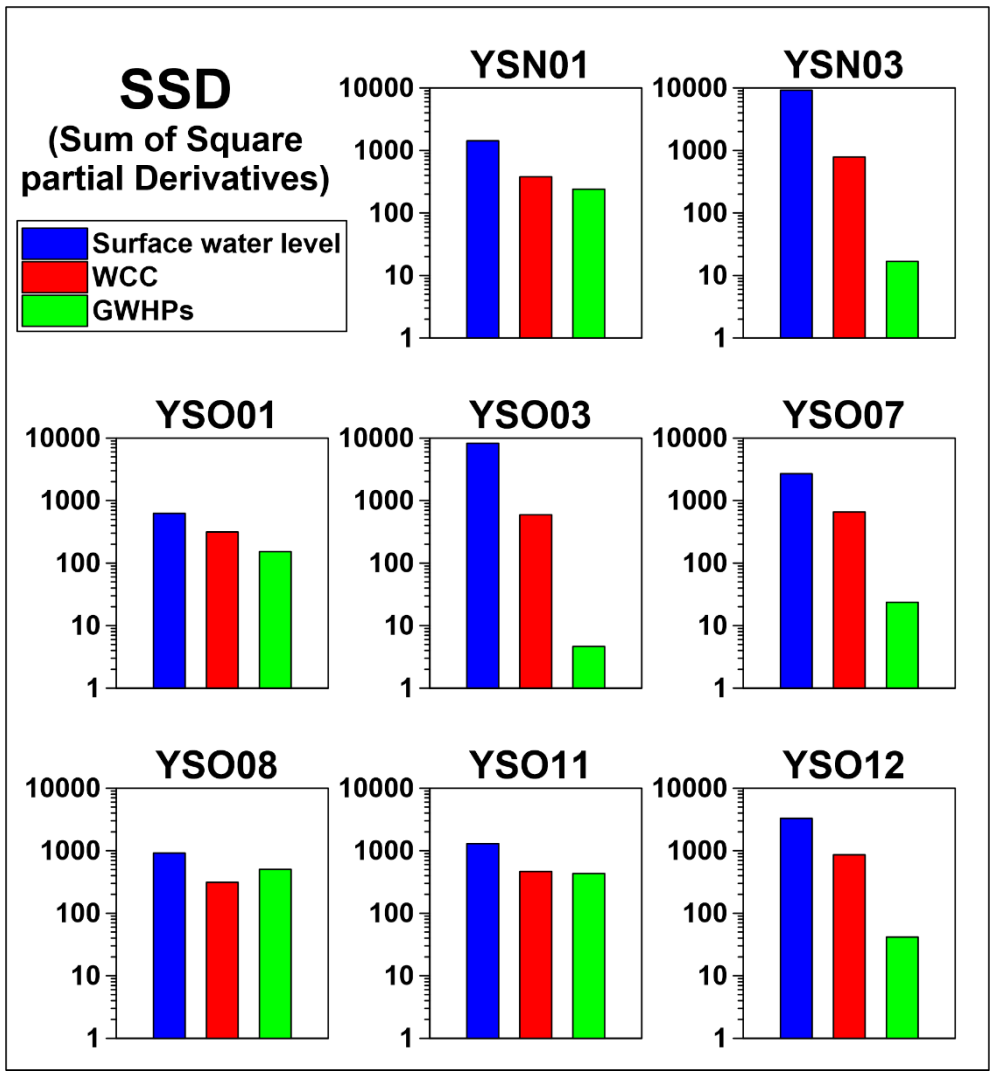

Fig. 4-14. Sum of square partial derivatives obtained by PaD method.

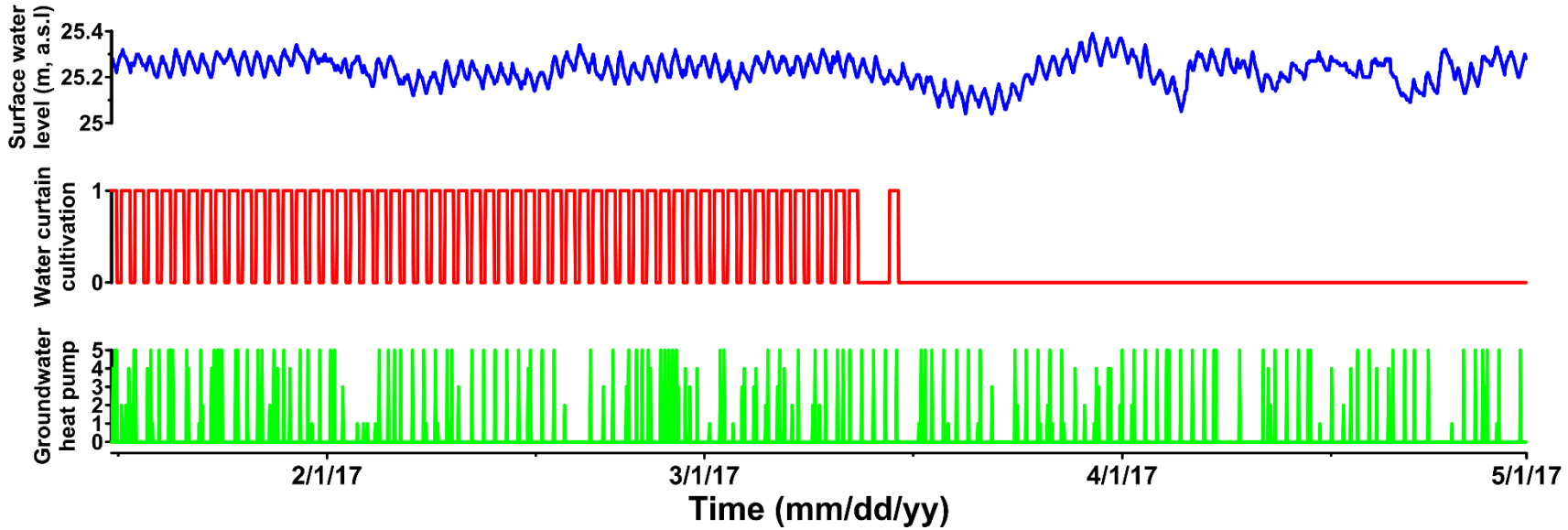

Fig. 4-15. Input data set for predict period.

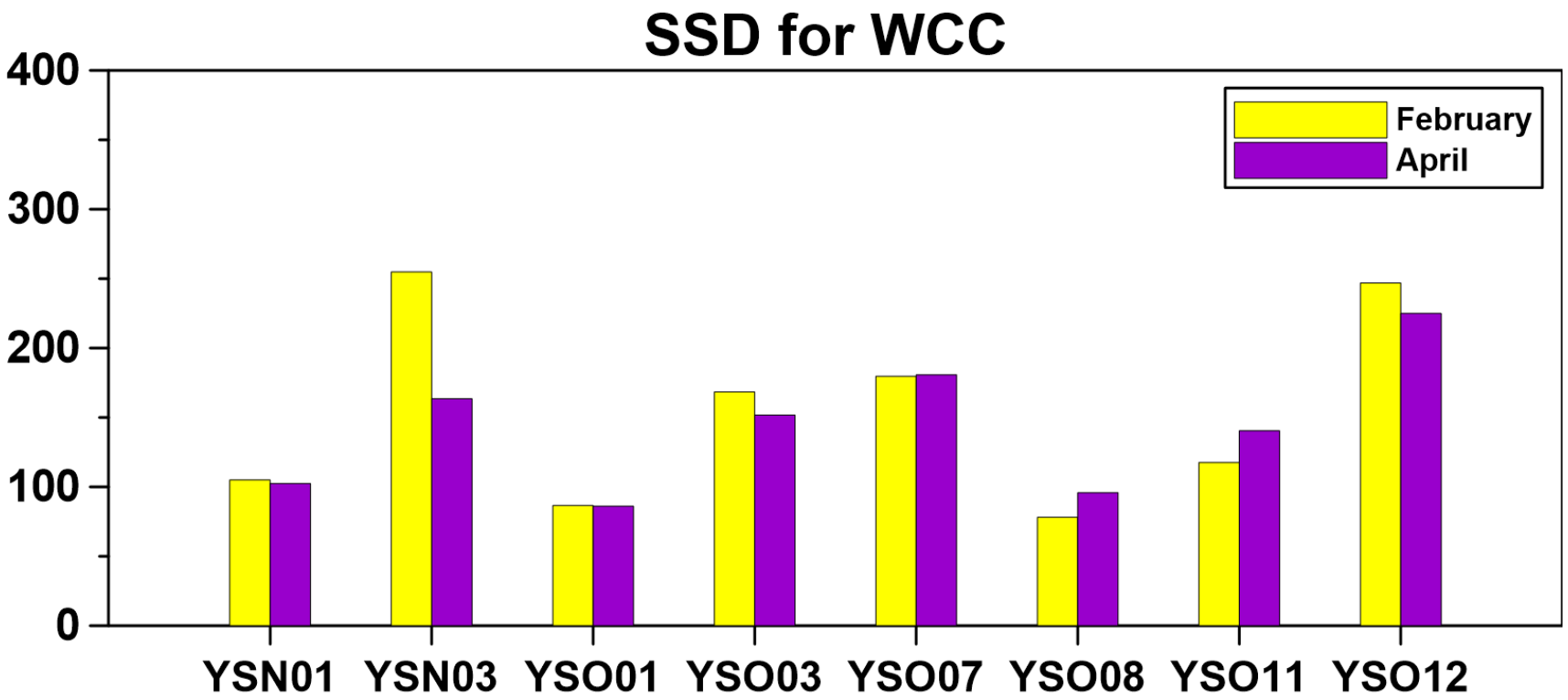

Fig. 4-16. Comparison of sum of square partial derivatives for WCC on February and April in 2017.

4.3.3 Results of extraction method

Evaluation of influence factor using weights method and PaD method was not suitable for spatial or temporal comparison. In order to make up for defects of these two methods, new approach calculating the contribution and the relative importance of input variable was developed and applied in this study.

Figure 4-18 and Fig. 4-19 represent the contribution and the relative importance computed by extraction method, respectively. What is unique about the contribution is that it shows negative value on some occasion. The negative value of contribution denotes that the influence factor has negative correlation with the output. For example, contribution of WCC in all wells should be negative because operation of WCC cause decrease of the groundwater level by pumping water. On the other hand, contribution of surface water level is always positive because correlation between the surface water level and the groundwater level is close to 1.

All the contribution and the relative importance describe the surface water level as the most dominant influential factor, meaning the groundwater level fluctuation has a higher correlation with surface water level than the other factors. The effect of WCC is rather low in every wells, the GWHPs have little or no contribution.

To compare them spatially, contribution and relative importance were plotted on the map with 3-D bar and pie chart (Figure 4-20a and b). They show reasonable evaluation of influence factors when considering the relative orientation and distance of human activities. Consulting the spatial comparison of the contribution and the relative importance, 8 wells were classified into three groups according to feature of them (Table 4-7). Wells which are on the west side, beside the Han river belong to group A (YSN03, YSO03, YSO07, YSO12). They have very high effect of surface

water level, low WCC, and little or no GWHPs influence. Well number YSO08 and YSO11 are group B, showing rather higher effect of WCC than group A due to closer distance from WCC site. And they have some negative contribution of GWHPs because there is pumping well (YSG03) around them. Group C (YSN01, YSO01) is somewhat similar to group B but has difference in that their contributions of GHWPs are positive. It is because wells in group C are around the injection well (YSG01, YSP).

After then, the contribution and the relative importance for three months (February, March, and April) were calculated to verify its ability of temporal comparison. Fig. 4-21 illustrates the contribution in order of month, and Fig. 4-22 shows the relative importance. As referred in section 4.3.2 and Fig. 4-15, WCC is practiced daily during winter season, and stopped after the middle of March. This feature is revealed well at both of Fig. 4-21 and Fig. 4-22, implying the suitability of extraction method to temporal comparison. Especially, contribution of WCC in February is higher than that of surface water level except for three wells (YSN03, YSO03, YSO07). In March, figures show decreased contribution and relative importance of WCC compared to those of February because the WCC is completely stopped on March 16. The crucial point is that there is no contribution of the WCC for every well in April when the WCC wasn't operated. Unlike PaD method, newly developed method can perform periodically adjusted evaluation successfully, making it possible to compare the contribution temporally.

The groundwater level fluctuation is due to several complex factors having each different feature. Therefore, it is difficult to catch the effect of them separately with only observed data. The process of prediction and evaluation of influence factors using the ANN model made it possible to decompose the combined effects of all

factors. Evaluation of influence factors will help understanding formation of the groundwater level and managing it efficiently. Also, it could be fundamental reference for other studies about interaction between the surface water and the groundwater, impact of the operation of GWHPs or WCC, and so on.

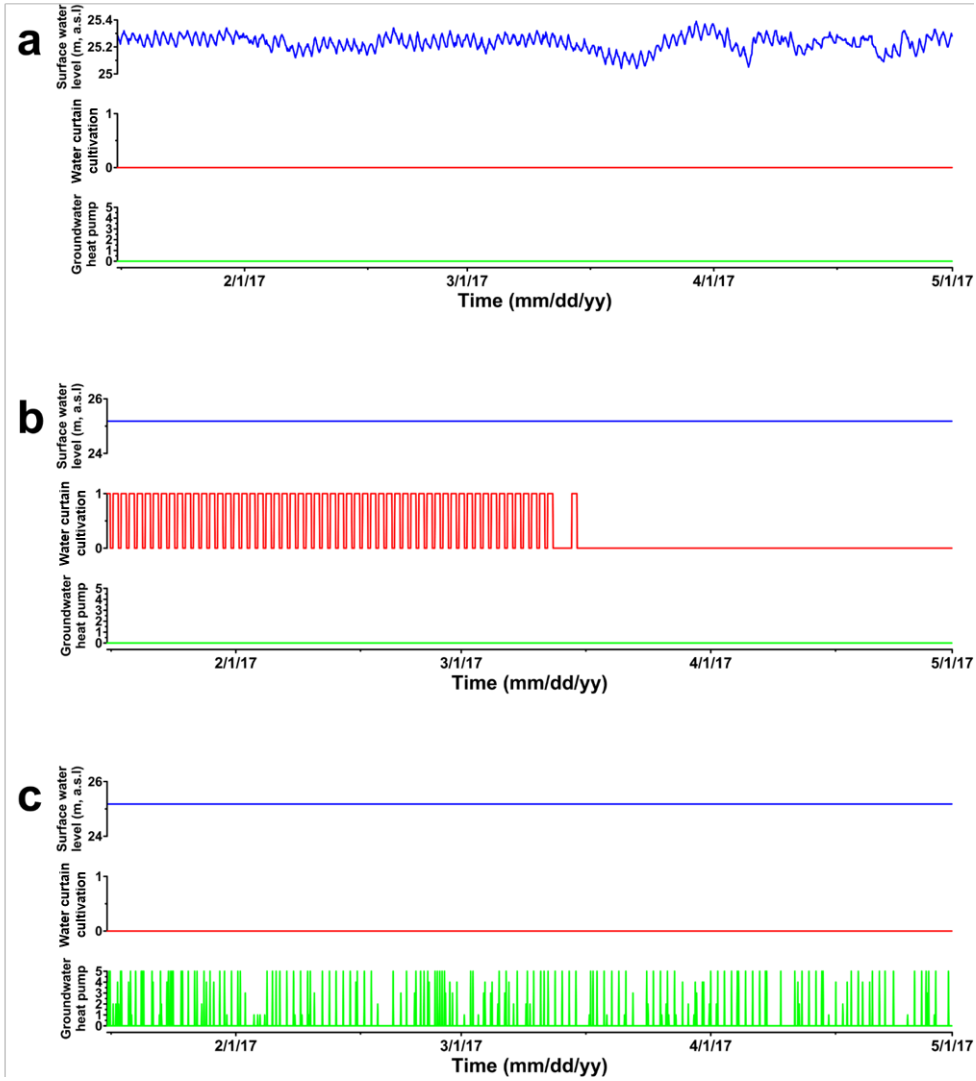

Fig. 4-17. Controlled input data sets which have only effects of (a) surface water level, (b) WCC, and (c) GWHPs.

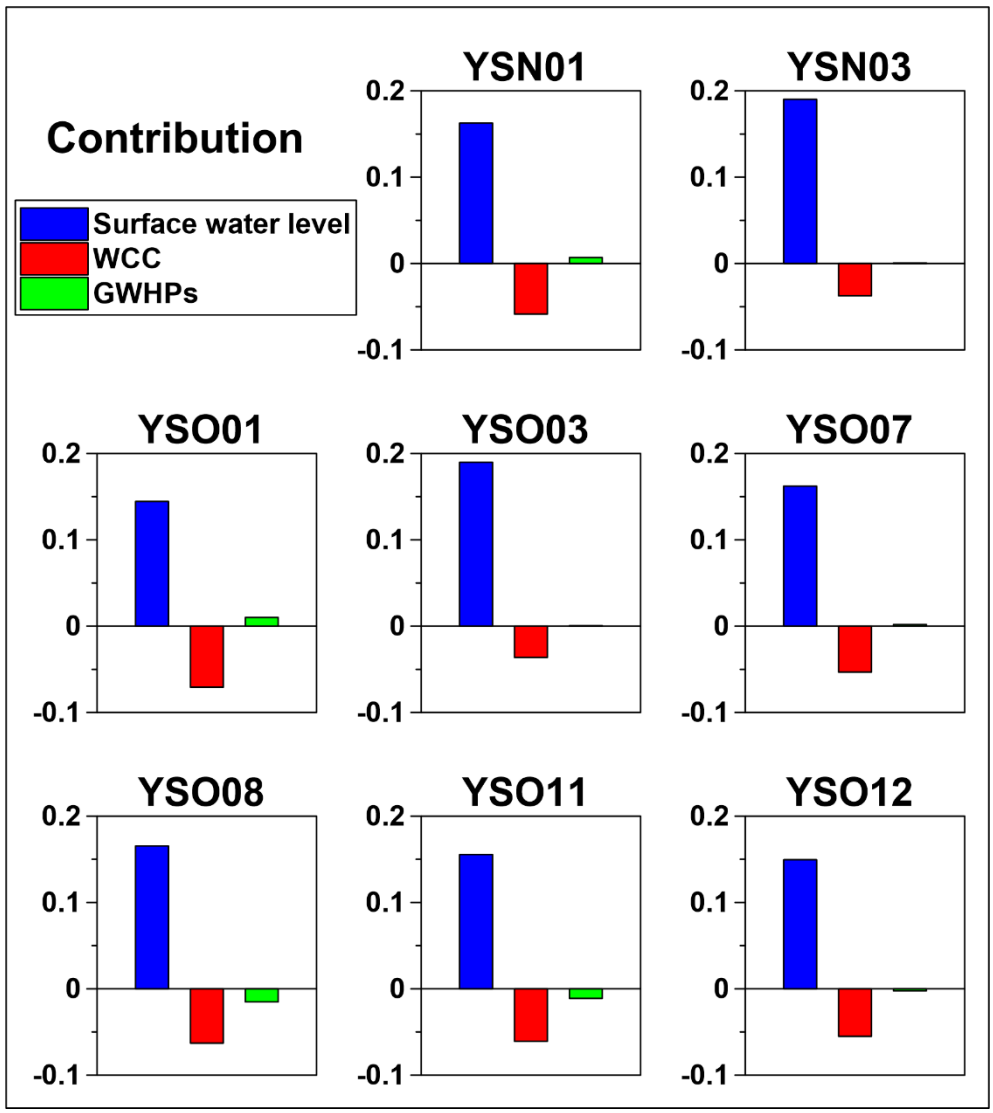

Fig. 4-18. Contribution obtained by extraction method.

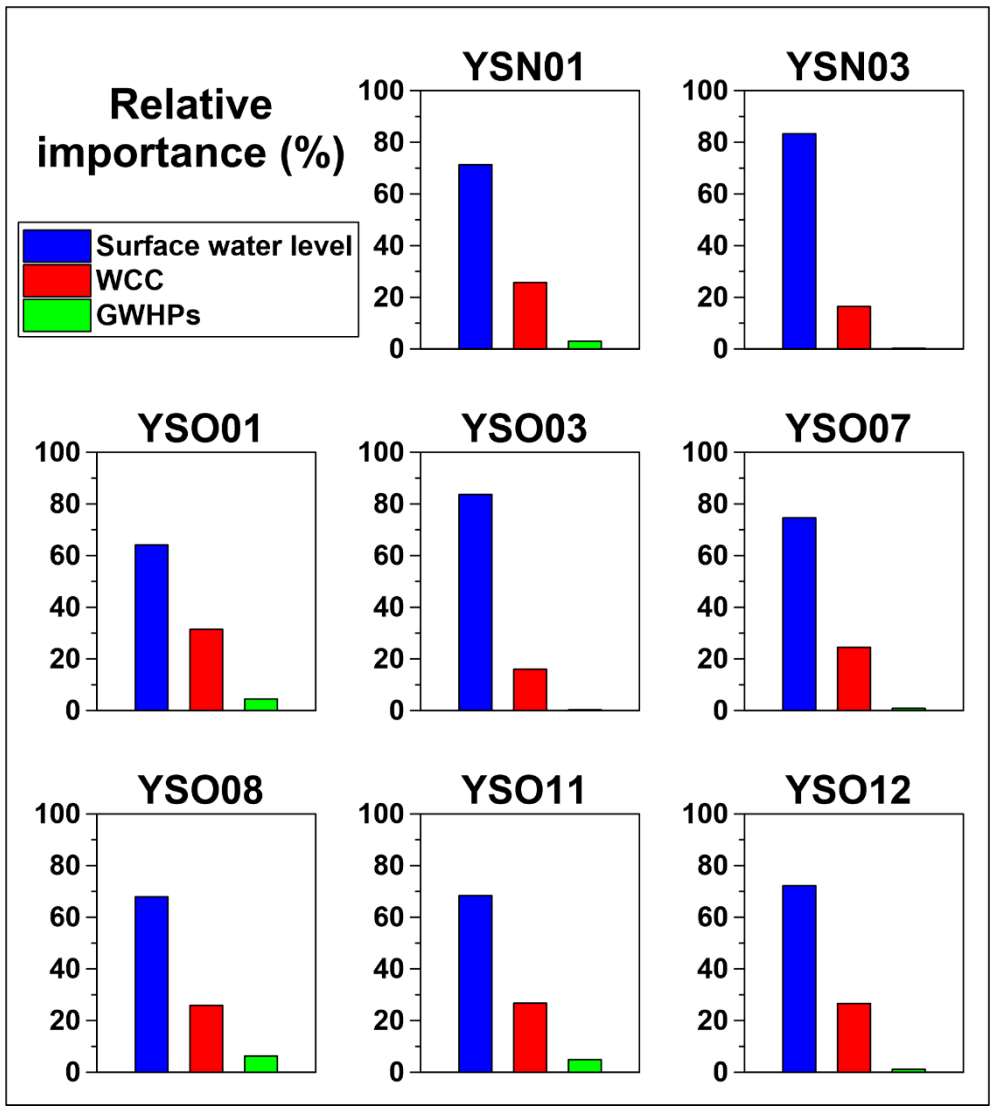

Fig. 4-19. Relative importance obtained by extraction method.

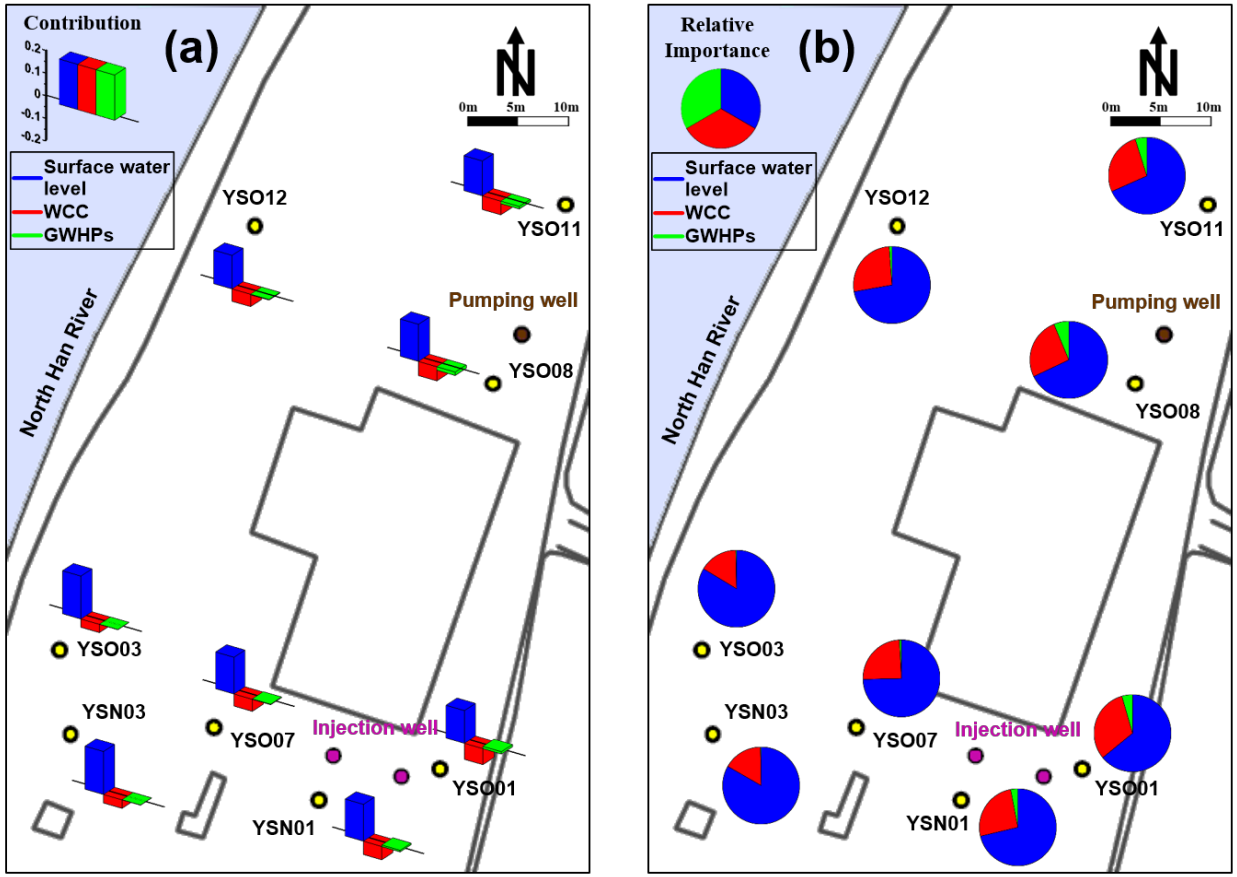

Fig. 4-20. (a) 3-D bars of the contribution, and (b) pie charts of the relative importance plotted on the map.

Table 4-7 Classification of wells according to features of the contribution and the relative importance.

Groups	Well number	Feature
Group 1	YSN03, YSO03, YSO07, YSO12	Very high effect of surface water level Rather low effect of WCC Little or no effect of GWHPs
Group 2	YSO08, YSO11	High effect of surface water level Slightly higher effect of WCC than group A Small effect of GWHPs (negative)
Group 3	YSN01, YSO01	High effect of surface water level Slightly higher effect of WCC than group A Small effect of GWHPs (positive)

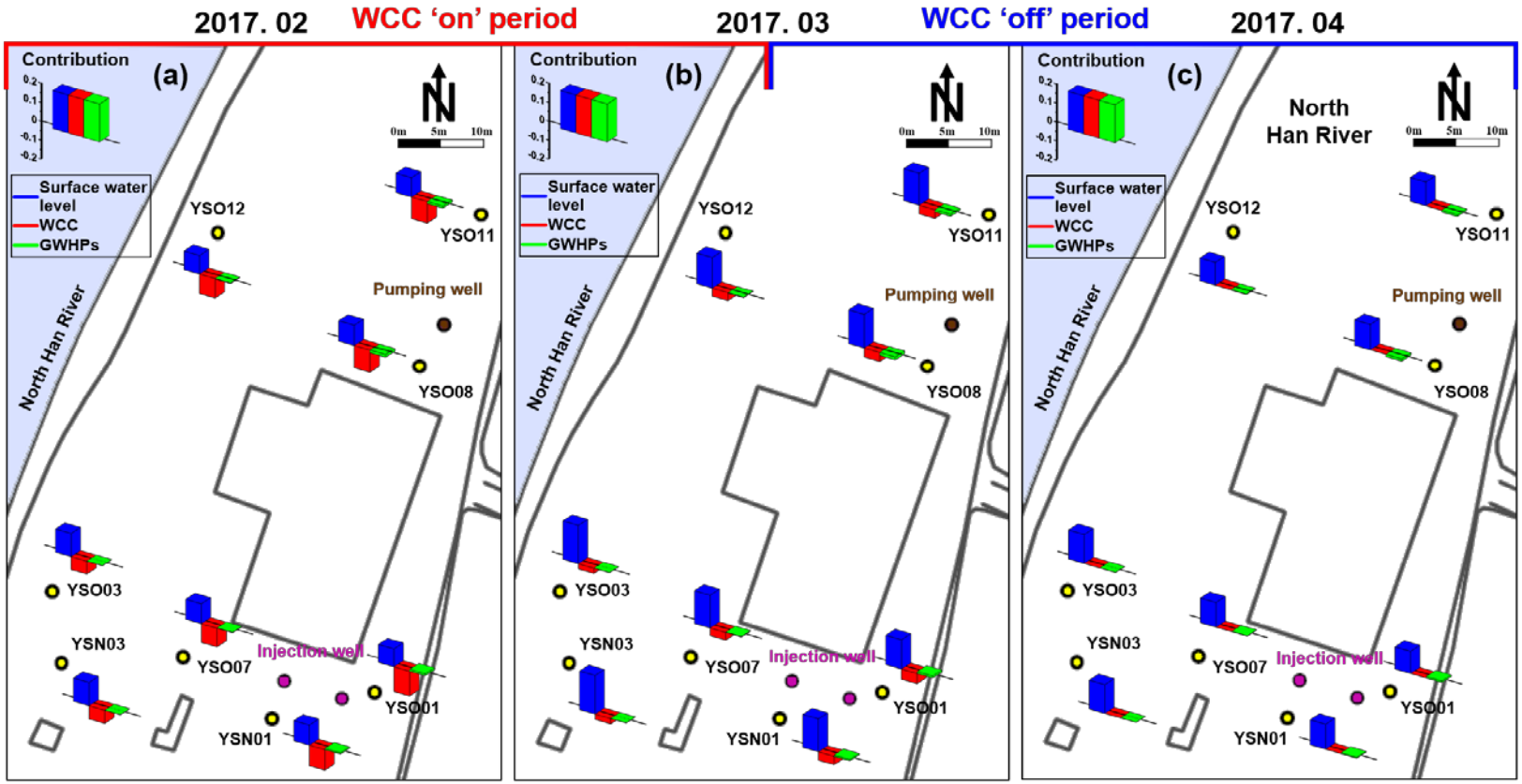

Fig. 4-21. Temporal comparison of the contribution: (a) February, (b) March, and (c) April.

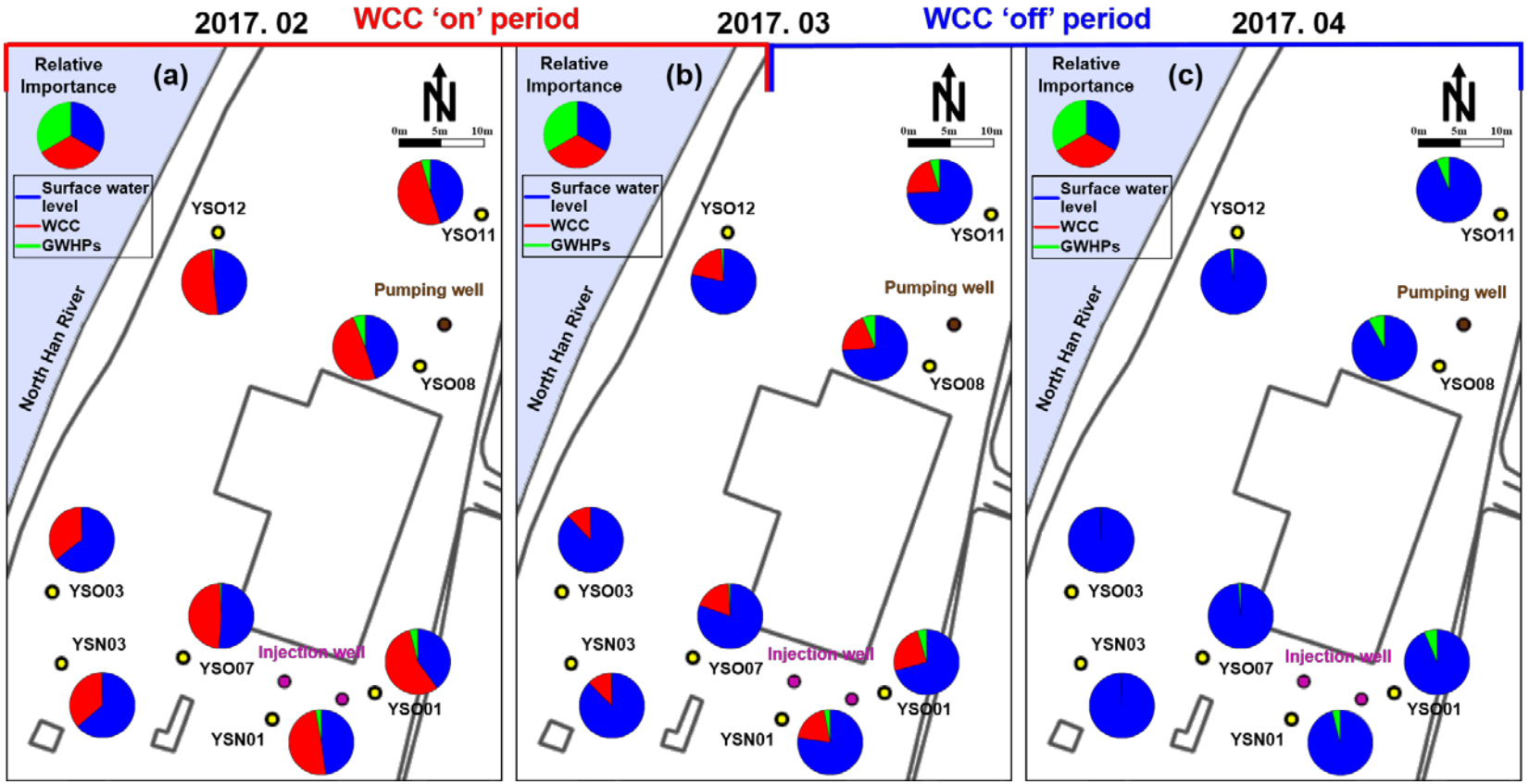

Fig. 4-22. Temporal comparison of the relative importance: (a) February, (b) March, and (c) April.

4.3.4 Comparison of three methods

Results of computing the contribution using three methods (weights method, PaD method, extraction method) were directly compared to know which is most pertinent method for evaluation of input variables. Fig. 4-23a and b show contributions of the GWHPs resulted from three methods in order of distance from the injection well and the pumping well, respectively. Relative importance calculated by weights method was independent of distance from the injection/pumping well. SSD computed by PaD method intends to decrease along with the distance from the injection/pumping well. However, there is exceptional well (YSN01) which showed higher contribution than that of YSO01 although YSN01 is farther from the injection well than YSO01. On the other hand, contribution from extraction method was always inversely proportional to the distance from the injection/pumping well with no exception. Fig. 4-24 represents contributions of the WCC on February and April, 2017. Weights method showed same result between two months because weights method calculates the relative importance using weights determined in train period. Contribution obtained by PaD method had similar value between February and April, showing higher contribution on April in some wells (YSO07, YSO08, YSO11) though there was no WCC on April, 2017. Therefore, weights method and PaD method are unsuitable for evaluating the influence factors reflecting the time-periodic characteristics. The extraction method, however, calculated contribution represents appropriate difference between two months as mentioned in section 4.3.3. Overall, the extraction method is superior to the other two methods for evaluating the influence factors based on time series data.

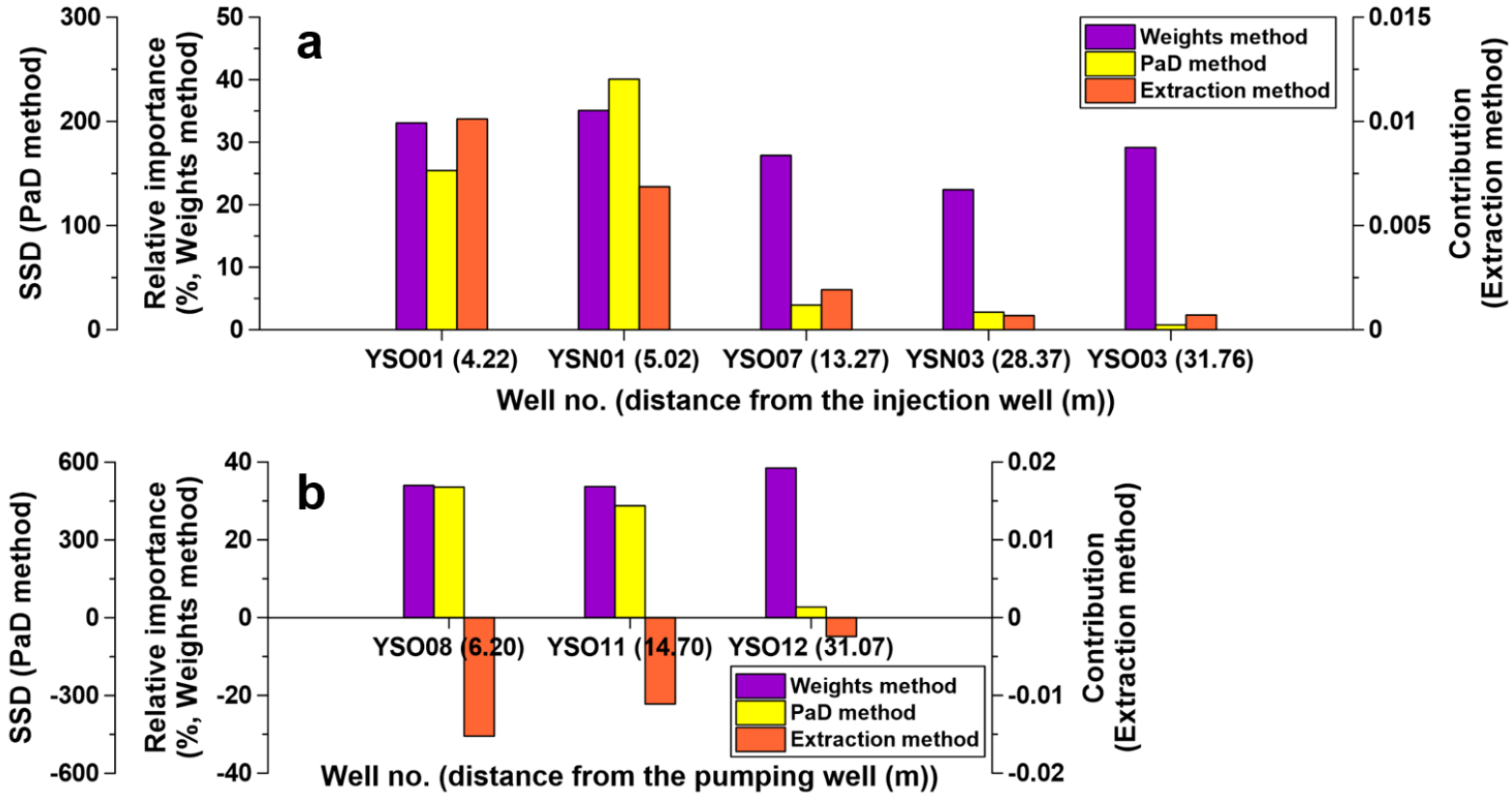

Fig. 4-23. Contribution or relative importance of the GWHPs obtained by three methods (a) for wells around the injection well; (b) for wells around the pumping well.

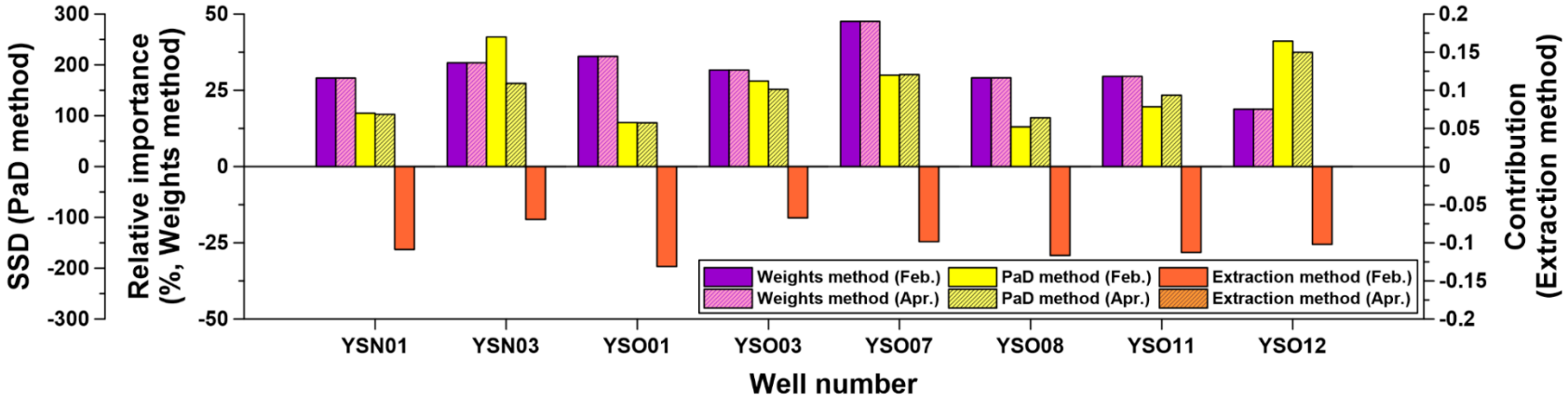

Fig. 4-24. Contribution or relative importance of the WCC on February and April in 2017 obtained by three methods.

5 CONCLUSION

In this study, prediction of the groundwater level and evaluation of influence factors on the fluctuation of the groundwater level in the riverside area were performed by using the artificial neural network (ANN) model. Various human activities such as surface water levels controlled by the dam operation, water curtain cultivation (WCC) and the groundwater heat pump system (GWHPs) around the study area have affected the groundwater level complexly in the study area, three influence factors (surface water level, WCC, GWHPs) were selected to be applied as input variables for artificial neural network models. Weights, biases, and three model parameters (number of hidden nodes, learning rate, momentum) which are components of architecture of an ANN model were determined in train period and test period for each well. Based on these determined components, the groundwater level forecasting was successfully performed in prediction stage with RMSE in a range of 0.03~0.06 m. Prediction of the groundwater level at several wells which are close to pumping/injection well for GWHPs and WCC site shows relatively higher error than the others, implying that these two variables are complicated to predict the groundwater level more accurately. Nevertheless, errors of predictions indicated that the results are reliable compared to previous studies of the ANN model.

To understand a contribution of influence factors affecting temporal changes of the groundwater level, weights method and PaD method were applied in this study. Both methods calculated the contribution or relative importance of each influence factors based on ANN models with its own limitations (i.e. weights method shows defect in a spatial comparison, PaD method is not suitable for a temporal comparison). To evaluate the contribution and the relative importance of influence

factors which are proper to spatial and temporal comparison, a concept of contribution was newly defined for data based on time series. The keywords of defined contribution in this study are ‘strength’ and ‘frequency’, meaning how much and often the input factors affect the groundwater level. By controlling the input data set, a response of the groundwater level to each input variable was obtained and used to calculate the contribution. The results denoted that effect of surface water level is the most dominant factor, effect of the GWHPs is the lowest one, and the magnitude of impacts are slightly different according to their features. These results also seem to be able to compare with other researches on this area studying the interaction between the surface water and the groundwater, or impact of surface water on the groundwater level using various analytical methods.

6 REFERENCES

Alley, W. M., Reilly, T. E., and Franke, O. L., 1999, Sustainability of Ground-Water Resources, United States Geological Survey Circular, 1186.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology I: preliminary concepts, Journal of Hydrologic Engineering, 5, 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology II: hydrologic applications, Journal of Hydrologic Engineering, 5, 124–137.

Brunner, P., and Simmons, C. T., 2012, HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model, Groundwater, 50(2), 170-176.

Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A., and Lek, S., 1995, Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece), Ecological Modelling, 120(2-3), 157-165.

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

Davis, J. C., 1973, Statistics and data analysis in geology, John Wiely & Sons, Inc., New York.

Eriksson, E., 1970, Groundwater time series, Hydrology Research, 1(3), 181-205.

Foster, S., and Chilton, P., 2003, Groundwater: the processes and global significance of aquifer degradation, Philosophical Transactions of the Royal Society of London B: Biological Sciences, 358(1440), 1957-1972.

Gevrey, M., Dimopoulos, I., and Lek, S., 2003, Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecological Modelling, 160(3), 249-264.

Gleeson, T., Alley, W. M., Allen, D. M., Sophocleous, M. A., Zhou, Y., Taniguchi, M., and VanderSteen, J., 2012, Towards sustainable groundwater use: setting long-term goals, backcasting, and managing adaptively, Groundwater, 50(1), 19-26.

Goh, A. T. C., 1995, Back-propagation neural networks for modeling complex systems, Artificial Intelligence in Engineering, 9(3), 143-151.

Jan, C.-D., Chen, T.-H., and Lo, W.-C., 2007, Effect of rainfall intensity and distribution on groundwater level fluctuations, Journal of Hydrology, 332(3), 348-360.

Jeong, J., Park, Y., Jo, Y., and Lee, J., 2010, Time series analysis of groundwater level fluctuation data in Cheonjeonri, Chuncheon, Gangwon-do, Journal of the Geological Society of Korea, 46(2), 171-176.

Konikow, L. F., and Kendy, E., 2005, Groundwater depletion: A global problem, Hydrogeology Journal, 13(1), 317-320.

Korea Meteorological Administration (KMA), 2017, Home page. http://www.kma.go.kr.

Ministry of Land, Infrastructure and Transport, 2017, Reprot of development of a technology for use of geological underground water convergence energy in the water front.

Ministry of Land, Infrastructure and Maritime Affairs, 2008, Report of the basic investigation research of groundwater in Yangpyeong.

Mohanty, S., Jha, M. K., Kumar, A., and Sudheer, K. P., 2010, Artificial Neural Network Modeling for Groundwater Level Forecasting in a River Island of Eastern India, Water Resources Management, 24(9), 1845-1865.

Moon, S.-H., Ha, K., Kim, Y., and Yoon, P., 2012, Analysis of Groundwater Use and Discharge in Water Curtain Cultivation Areas: Study of the Cheongweon and Chungju Areas, Journal of Engineering Geology, 22(4), 387-398.

Nam, Y., and Ooka, R., 2012, Numerical simulation of ground heat and water transfer for groundwater heat pump system based on real-scale experiment, Energy and Buildings, 42(1), 69-75.

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

Olden, J. D., and Jackson, D. A., 2002, Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks, Ecological Modelling, 154(1-2), 135-150.

Padilla, A., and Pulido-Bosch, A., 1995, Study of hydrographs of karstic aquifers by means of correlation and cross-spectral analysis, Journal of Hydrology, 168(1-4), 73-89.

Park, B.-H., Bae, G.-O., and Lee, K.-K., 2015, Importance of thermal dispersivity in designing groundwater heat pump (GWHP) system: Field and numerical study, Renewable Energy, 83, 270-279.

Rumelhart, D. E., McClelland, J. L., The PDP Research Group, 1986, Parallel distributed processing: Explorations in the microstructure of cognition, MIT Press, Cambridge, Massachusetts, USA, pp516.

Sung, A. H., 1998, Ranking importance of input parameters of neural networks, Expert Systems with Applications, 15(3-4), 405-411.

Taormina, R., Chau, K., and 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.

Trefry, M. G., and Muffels, C., 2007, FEFLOW: A Finite-Element Ground Water Flow and Transport Modeling Tool, Groundwater, 45(5), 525-528.

Wang, S., Shao, J., Song, X., Zhang, Y., Huo, Z., and Zhou, X., 2008, Application of MODFLOW and geographic information system to groundwater flow simulation in North China Plain, China, Environmental Geology, 55(7), 1449-1462.

Wong, H., Ip, W.-c., Zhang, R., and Xia, J., 2007, Non-parametric time series models for hydrological forecasting, Journal of Hydrology, 332(3-4), 337-347.

Yang, Z. P., Lu, W. X., Long, Y. Q., and Li, P., 2009, Application and comparison of two prediction models for groundwater levels: A case study in Western Jilin Province, China, Journal of Arid Environments, 73(4-5), 487-492.

Yao, J., Teng, N., Poh, H.-L., and Tan, C. L., 1998, Forecasting and Analysis of Marketing Data Using Neural Networks, Journal of Information Science and Engineering, 14, 843-862.

Yi, M., Kim, G., Sohn, Y., Lee, J., and Lee, K., 2004, Time series analysis of groundwater level data obtained from national groundwater monitoring stations, Journal of the geological society of Korea, 40, 305-329.

Yoon, H., Hyun, Y., Ha, K., Lee, K.-K., and Kim, G.-B., 2016, A method to improve the stability and accuracy of ANN- and SVM-based time series models for long-term groundwater level predictions, Computers & Geosciences, 90, 144-155.

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., and Lee, K.-K., 2011, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, Journal of Hydrology, 396(1-2), 128-138.

Yoon, H., Park, E., Kim, G.-B., Ha, K., Yoon, P., and Lee, S.-H., 2015, A Method to Filter Out the Effect of River Stage Fluctuations using Time Series Model for Forecasting Groundwater Level and its Application to Groundwater Recharge Estimation, Journal of Soil and Groundwater Environment, 20(3), 74-82.

국문 초록

최근들어 지하수의 이용이 지속적으로 증가함에 따라서 지속가능한 지하수자원의 이용은 중요한 이슈로 떠오르고 있다. 특히 본 연구지역인 양평 강변지역은 지하수를 이용한 지열 냉난방 시스템의 운영, 수막재배와 같은 인위적인 활동이 집중되어 있어 지하수위 변화 예측이 수자원 관리 및 보호를 위해 필수적이다. 본 연구에서는 지하수위 예측을 위한 방법으로 인공신경망 기법을 적용하였다. 먼저, 예측을 위해 지하수위의 영향요인에 대한 분석 결과 강 수위와 수막재배, 그리고 지열 냉난방 시스템의 운영이 지하수위 변동과 높은 상관관계를 보였고, 반면에 강수는 낮은 상관관계를 보였다. 위에서 언급한 세가지 영향요인을 입력자료로 활용하여 학습기간과 시험기간을 통해 연결강도와 편중값, 인공신경망 모델의 파라미터들을 결정하는 등, 인공신경망의 구조를 결정하였고, 이를 적용하여 8 개 관정에서의 지하수위 예측을 실시하였다. 지하수위 예측결과, 관측기록과 비교하여 RMSE 가 0.03 m 에서 0.06 m 사이의 낮은 값을 보였다. 추가적으로 지하수위 변동으로부터 영향요인들의 영향력을 각각 파악하기 위하여 각 영향요인의 기여도 및 상대적 중요도를 계산하였다. 먼저 인공신경망 모델에서 입력자료들을 정량화하는 기존방법인 weights method 와 PaD method 를 이용하여 기여도 또는 상대적 중요도를 계산했다. 하지만 두 방법들은 각각 공간적인 비교와 시기별 비교에서 한계를 드러내었고 이를 극복하기 위하여 본 연구에서는 extraction method 를 개발하였고 공간, 시간적 비교가 가능함을 확인하였다. 결과적으로 지표수가 지하수위의 변화에 가장 우세한 영향 (64.09~83.30 %)을 미쳤고 수막재배가 다음으로 높은 영향 (16.04~26.76 %)을 보였으며 지열 냉난방 시스템은 매우 적은 영향 (0.29~6.26 %)을 나타냈다. 특히 수막재배의 영향은 시기에 따라서 매우 급격하게 변화하는데 겨울철

수막재배가 내내 일어나는 기간에 대해서는 오히려 지표수보다 더 높은 기여도를 나타내었다. 인공신경망 기법을 이용한 지하수위의 영향요인 평가는 매우 복잡하고 비선형적인 특징을 갖고 있는 대수층에서의 지하수위 변동에 대한 원동력을 이해하는데 큰 도움을 줄 것이다.

주요어 : 지하수위 예측, 인공신경망 모델, 영향요인, 기여도, 상대적 중요도