Group Case Study: Ground Water Level Statistics Analysis in Beijing Plain

Feiran Wang:1066589

Weijia Luo: 1073885

Yining Zang:1068605

Zhechen Zhang:1074448

Zhuowei Quan:1073845

1. Introduction of Beijing Plain

The area we study is the Beijing Plain in China. Beijing is the capital of China, and the plain area accounts for a large proportion.

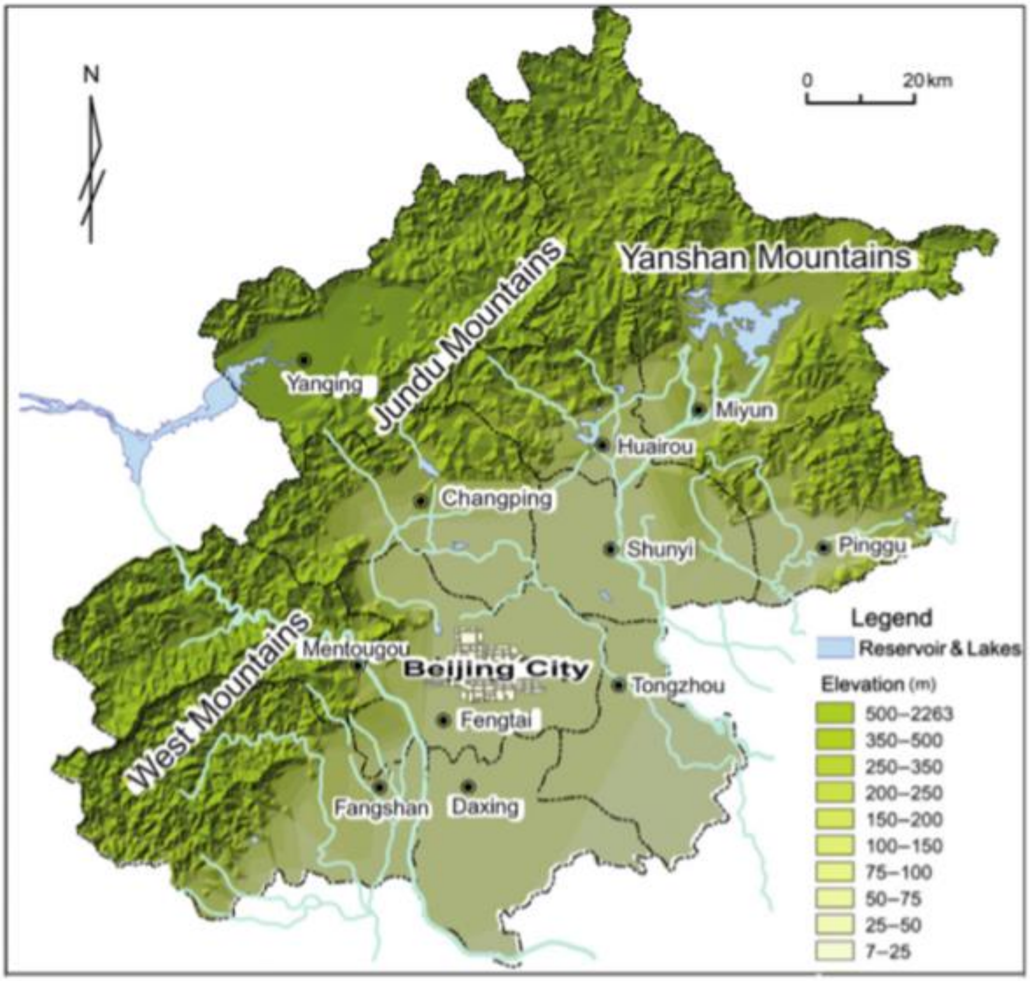


Figure 1. Topographic map of Beijing Plain

Figure 1 clearly shows the topography of Beijing, with the plain at the southeast corner.

1) Hydrogeology condition

The area of Beijing Plain is 6338 square kilometers, accounting for 38.6% of the total area of Beijing, in which elevation is low and sloping toward the Bohai Sea, with an altitude of 20 to 60 meters.

2) Climate conditions

Beijing is an inland city located in the north of China. So it is typically warm temperate semi-humid continental monsoon climate, with four distinct seasons. Spring and autumn are short, but the wind is relatively heavy, and the weather is dry. The summer is hot and rainy, up to 39 degrees, and the winter is colder, which can be below minus. The annual average temperature is 10 ~ 13 degrees.

It has an average annual rainfall of more than 590 millimeters, which is one of the areas with the highest rainfall in North China. However, the precipitation season is very unevenly distributed. 75% of the annual precipitation is concentrated in summer, and there are often heavy rains in July and August. From 1979 to 2006, the average annual evaporation of the Beijing Plain measured at 14 stations in Beijing was 1725mm, which is almost three times the annual average precipitation.

Overall, Beijing has a resident population of nearly 20 million, resulting in huge water demand and overexploitation of groundwater. And climate variability exerts extra pressure on the groundwater reserve.

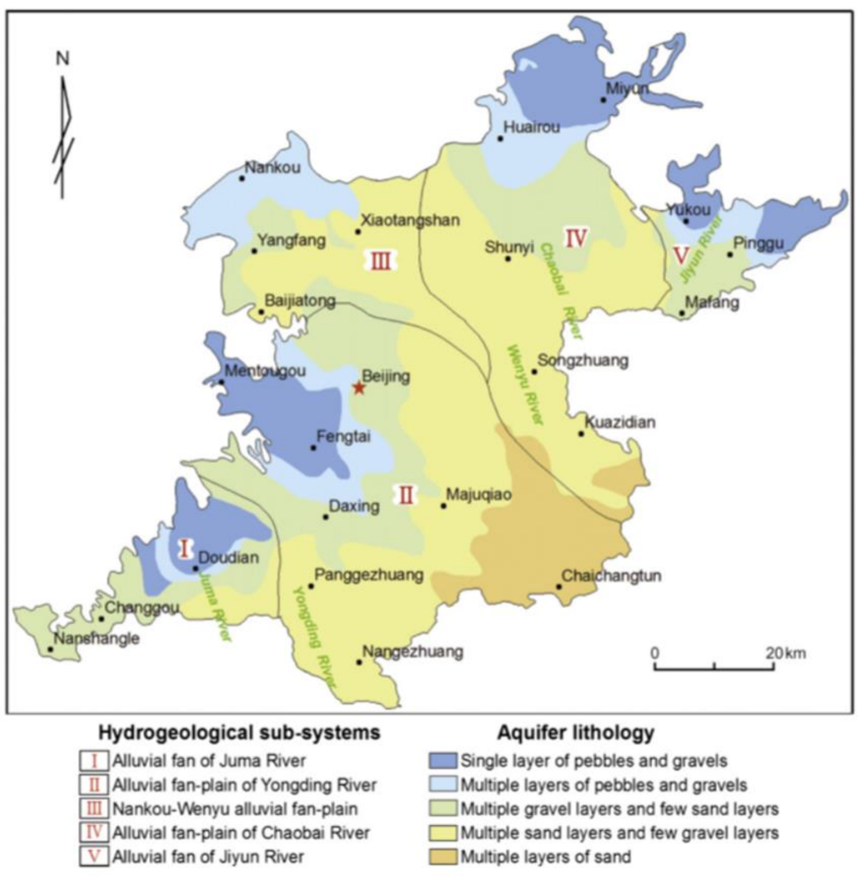


Figure 2. Hydrogeological zone map

The figure above shows the regional distribution of hydrogeology in Beijing.

Hydrology condition

Because it is located inland, Beijing has no natural lakes and only more than 80 rivers. Since the 1970s, the groundwater level in the Beijing Plain has declined significantly. Groundwater levels have decreased to historically low levels during the eight consecutive dry years from 1999 to 2006. The total drop in groundwater levels amounts to more than 20 m since the 1970s. Also, groundwater depletion has been caused either by over-exploitation or reduction of groundwater recharge. The combination of these two causes has accelerated groundwater depletion in Beijing Plain, China.

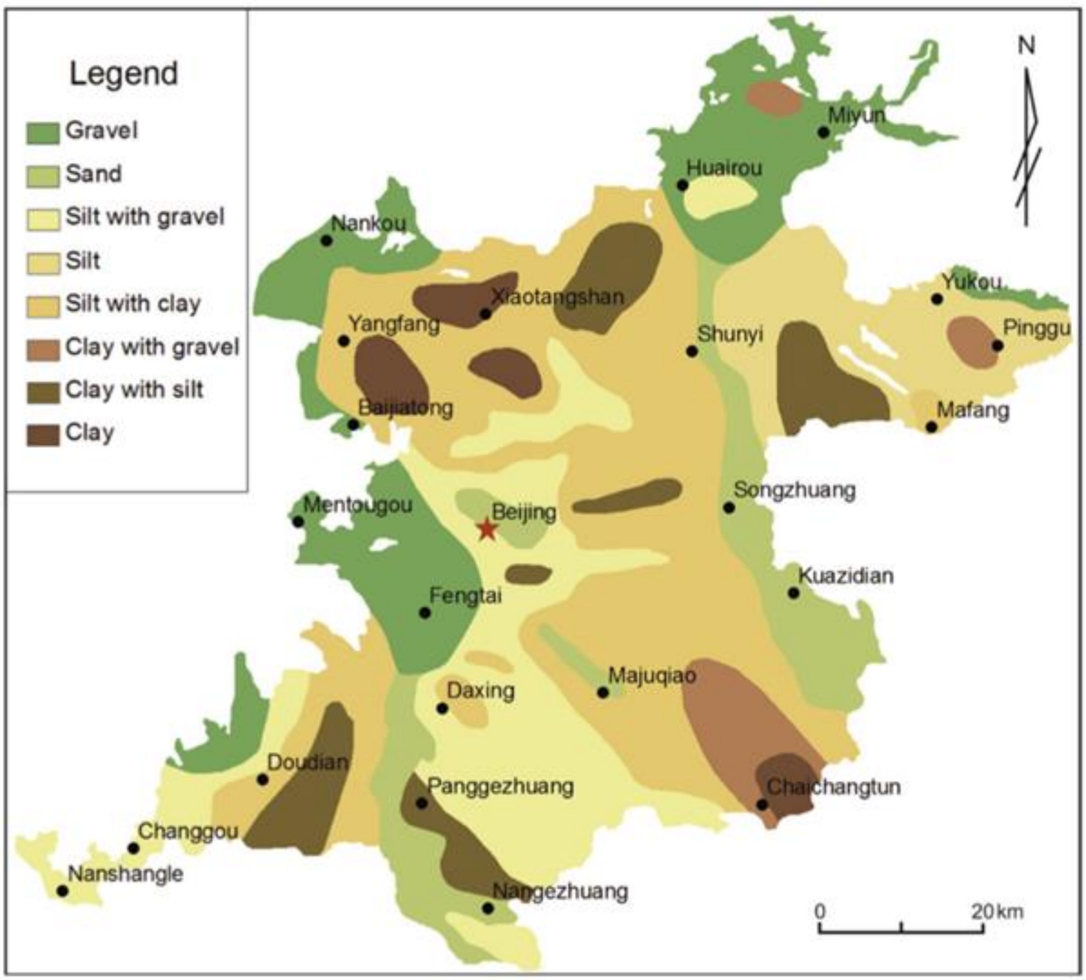


Figure 3. Unsaturated zone map

An unsaturated zone map was delineated using the water table depth map and borehole columns. The soil types of the Beijing Plain are divided into eight categories. Figure 3 shows the distribution of soil types. Gravels and sand are mainly distributed in alluvial fans. Clay dominated soils are distributed mainly in areas between alluvial fans. We can see that gravel and silt with clay are the main ones.

2. Statistical characteristics of data

In this part, the statistical analysis on the groundwater level of 138 observation wells, which are placed at Beijing Plain in 1995, was carried out to determine the regularity of the groundwater level distribution based on statistical parameters and the created frequency histogram.

Table 1: statistical parameters of the groundwater level

Descriptive Statistics			
Mean	32.85	Kurtosis	1.66
Median	30.45	Skewness	0.999
Standard Deviation	13.12		

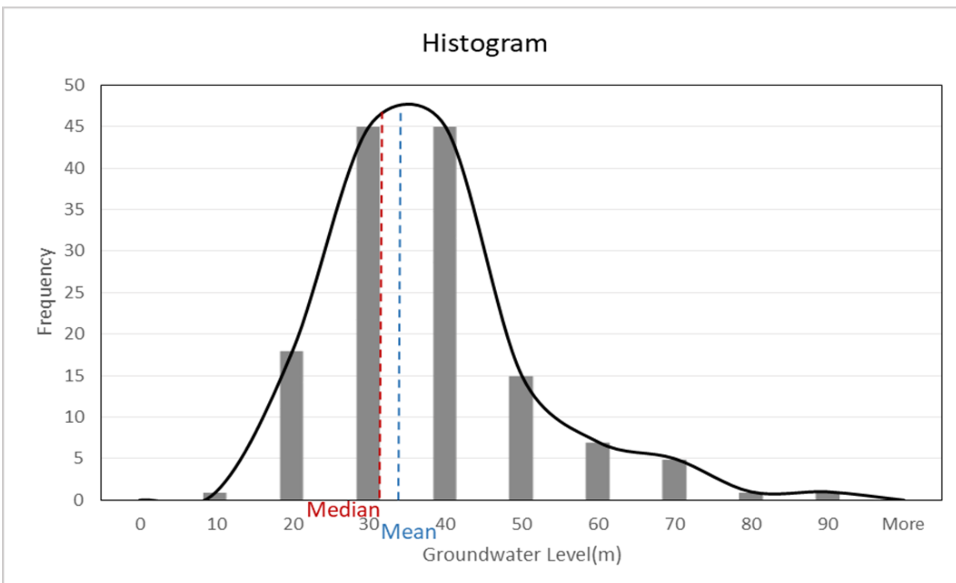


Figure 4: Relative frequency histogram of the groundwater in Beijing plain

1) Descriptors of asymmetry

From Figure 4, the shape of the histogram obviously shows asymmetric and skewed to the right due to the long tail occurring on the right side. On the other hand, from Table 1, the statistical parameter skewness coefficient is 0.99, which is greater than 0. Therefore it also presents the data set is positively skewness, indicating that most of the groundwater level gather around 20 to 40 meters, accounting for 90% more or less of the 138 samples. Based on this, it is known that the study area is situated in the alluvial fan with good recharge, but there are a few wells showing a high level at 60-90 meters because of that these wells were placed in the mountain areas with the high water level.

2) Descriptors of central tendency

From Table 1, the mean and the median value is 32.85 m and 30.45 m, respectively. The mean value is a bit larger than the median, but not much different. Hence, the mean and median values are good indicators for describing the central tendency of the groundwater level in the study area. Totally it is obtained that there is an average groundwater level at approximately 30 meters in Beijing plain.

3) Descriptors of dispersion

From Table 1, the standard deviation is with a small value of 13.11. In the meanwhile, the relative frequency histogram is narrow from Figure 1, stating that the data set shows a small degree of dispersion from its mean value, which means that there are the majority of values are close to the average value. Therefore, in this case, it is inferred that most of the groundwater level data are around the average value, indicating they uniformly distribute spatially, does not show a huge change in space.

4) Descriptors of flatness

The Kurtosis is lower than 3, which illustrates that the distribution is flatter comparing normal distribution.

From Figure 5, the cumulative frequency distribution shows that the groundwater level smaller than 50 meters is about 90 % among 138 observation wells. Besides, all groundwater level of 138 wells is lower than 90 meters.

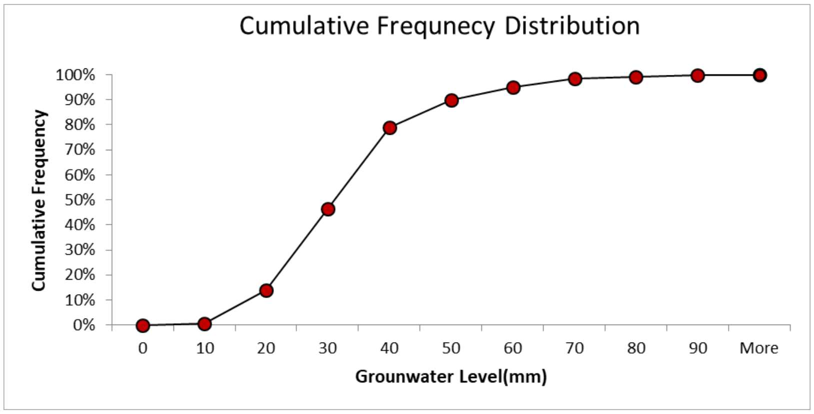


Figure 5: Cumulative frequency distribution of the groundwater level in Beijing Plain

Except for statistical parameters and histogram, the third method to describe the frequency distribution is fitting probability distribution functions to empirical distributions. Because of the positive skewness of Groundwater level data, so lognormal can be used to fit the empirical distribution.

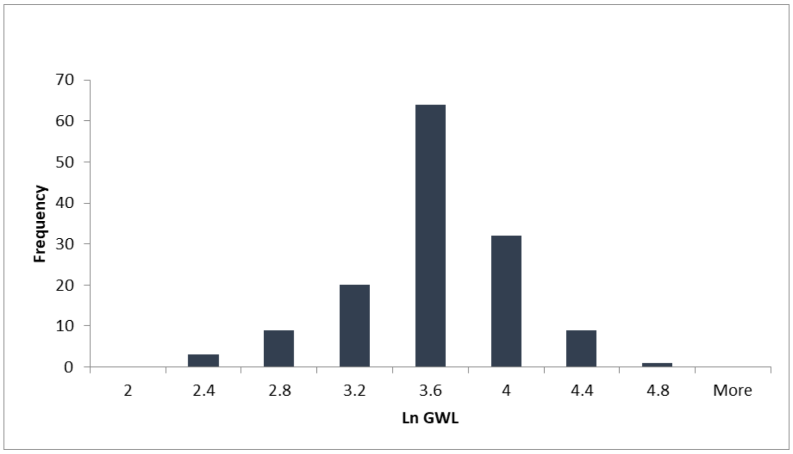


Figure 6: Histogram of logarithmic values of Groundwater level in Beijing Plain

The histogram of logarithmic values of the groundwater level shows that it is approximately symmetric about its midpoint. Therefore the Logarithmic groundwater level follows the Normal distribution. Based on this, we can plot the cumulative frequency of logarithmic values of the groundwater level in a logarithmic probability paper. It will show a straight line. Then the parameters like mean value and standard deviation can be found from the graph. Besides, with the known distribution function, it is easy to compute the probability of groundwater level exceeding any particular values.

3. Regression Analysis

Tongzhou district in Beijing Plain was chosen as the multiple regression analysis by using monthly groundwater level and precipitation in the same month, one month ago, two months ago, until five months ago due to the fact that groundwater level took five months to reach the peak after the peak precipitation. Figure 7 presents the monthly precipitation and groundwater level from 1989 to 2013. From the graph, there was an overall increase in groundwater level from 1989 to 1994 when totally yearly precipitation increasing during the same years. Similarly, when the precipitation overall decreased from 1995 to 2003, the groundwater gradually decreased along with this.

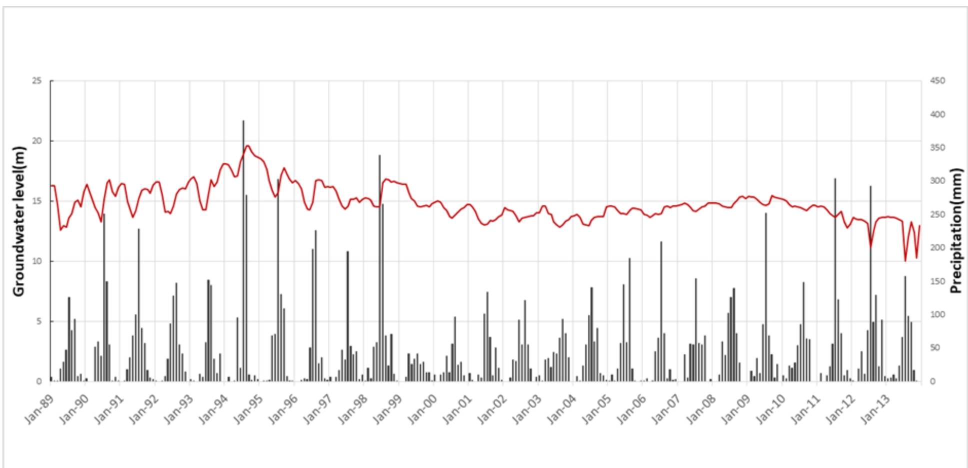


Figure 7: Groundwater level and precipitation in Tongzhou from 1989 to 2013.

From the correlation matrix (Table 2), it is known that the correlation coefficient is pretty small, but it is also clear that there is a time lag after precipitation. In addition, the groundwater level has a relatively higher correlation coefficient with the precipitation that occurred five months ago.

Table 2: the correlation matrix

	GWL	*Precipitation*	*Pt-1*	*Pt-2*	*Pt-3*	*Pt-4*	*Pt-5*
GWL	1						
Precipitation	-0.0584	1					
Pt-1	0.0964	0.4600	1				
Pt-2	0.1661	0.1445	0.4592	1			
Pt-3	0.1788	-0.1149	0.1441	0.4592	1		
Pt-4	0.1933	-0.2504	-0.1129	0.1467	0.4616	1	

Pt-5	0.2316	-0.3520	-0.2483	-0.1102	0.1494	0.4614	1

The regression equation was obtained from the regression.

$$GWL=-0.0006P+0.0025*P_{t-1}+0.0025*P_{t-2}+0.0011*P_{t-3}+0.0012*P_{t-4}+0.0051*P_{t-5}+14.305$$

From the ANOVA table (Table 3) and regression statistics table (Table 4), it is known that multiple correlation coefficient R is only 0.3 and the correlation determination value is only about 0.1, showing that the goodness of fit is 10%, which indicates that the model provides a not good fit to the observed data. From the ANOVA table, the calculated F is less than Fa, so the null hypothesis is accepted, meaning that the dependent variable is not related to any of the predictor variables.

Table 3: ANOVA table

	df	*SS*	*MS*	*F*	*Significance F*	*Fa(6,286)*
Regression	6	62.1339	10.3556	5.5198	1.9863E-05	2.1303
Residual	286	536.5617	1.8761			
Total	292	598.6957				

Table 4: Regression statistics

Multiple R	0.32215238
R Square	0.103782156
Adjusted R Square	0.084980383
Standard Error	1.369704366
Observations	293

On the other hand, by plotting predicated groundwater level values and actual measurement values (Figure 8), it is clear that there is not a good relationship between them.

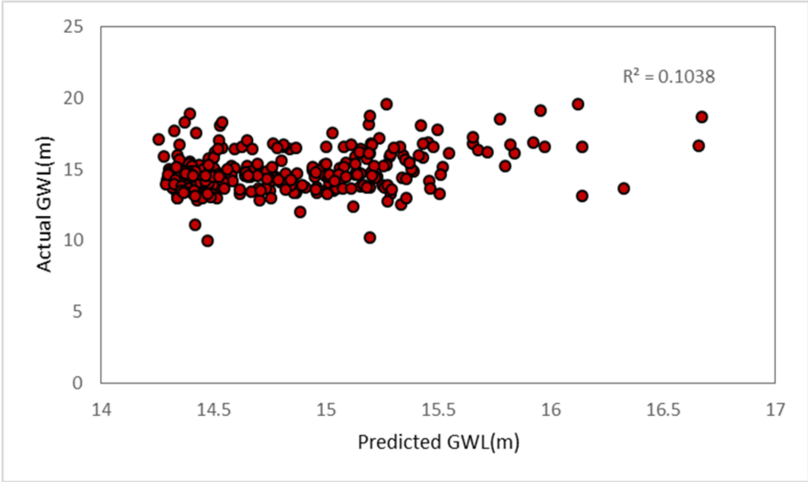


Figure 8: Predicted and actual groundwater level in Tongzhou.

As a result, the groundwater level is not directly related to the precipitation in the study area Tongzhou. There are three reasons to explain this situation. Firstly, the unsaturated zone is covered with silt and clay soil type in Tongzhou district, leading to the weak infiltration into the aquifer. Secondly, the area is near the cone depression, indicating intensive human activities such as extraction influence on the groundwater level. Lastly, the area may be accepted lateral recharge from other sources.

4. Time series analysis

The groundwater level time series in Tongzhou district that is selected for time series analysis is from an observation well 137. The observation well 137 is an unconfined aquifer observation well, and it includes 8-years time series from 1996 to 2004.

1) Trend analysis:

The parameters of linear +trend are calculated by FREQ as follows. The linear trend function is Ht = 38.0144 – 0.0276t and the standard deviation of series is 0.686. The trend exists at a confidence level of 95%, because the calculated t-value 160 is larger than the t-alpha 2, which is from the Student’s table. Figure 9 indicates a clearly sharp downward trend in Well 137 with a dropping level of 14 meters from 1996 to 2004.

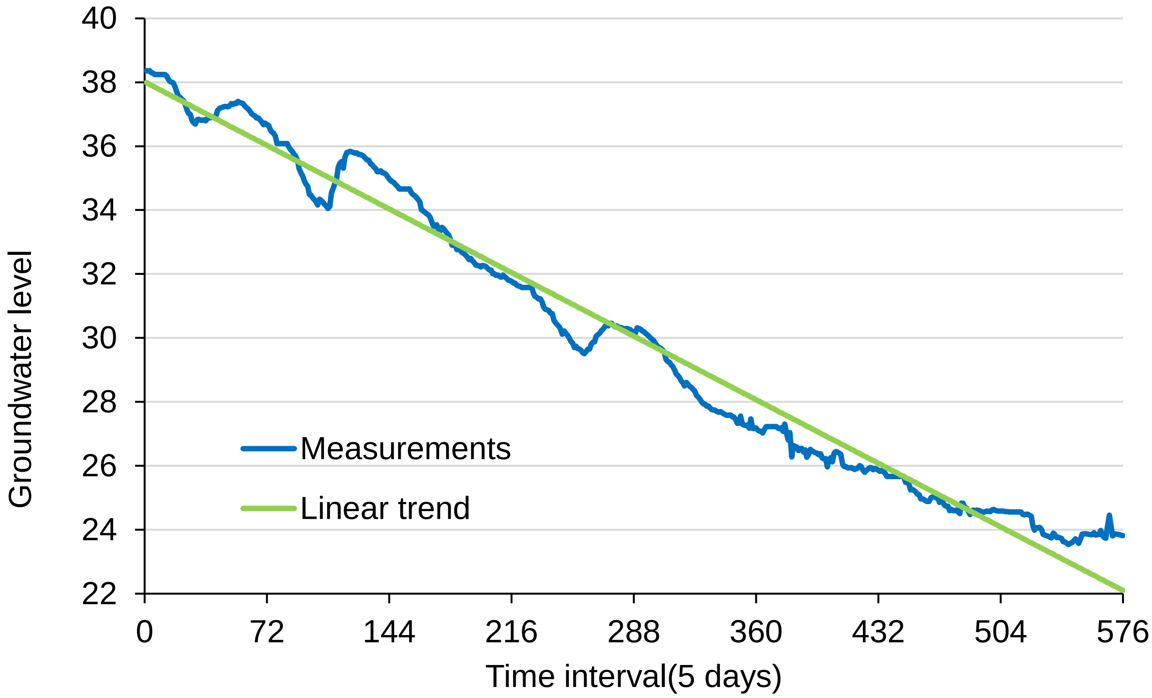


Figure 9. Groundwater level time series and fitted linear trend in Well 137

2) Periodic fluctuations

The differences between the groundwater time series and the trend line show the periodic fluctuation. The Harmonic series are used to fit the series period fluctuation. Table 5 lists the results. It can be observed the dominant periodic fluctuation is eight years period, which is the total time series period. And choosing the first five periods to estimate the period fluctuations, the estimated line compared with time series draws on Figure 10. The main five periods are larger than one year, which means a long periodic cycle exists in Beijing Plain and will be considered in the forward frequency optimization analysis. Another analysis from Figure 10 is that the groundwater table data series with the trend removed usually have seasonal changes. This is due to the influence of precipitation and extraction.

Table 5 Results of periodic analysis of groundwater level series in Well 137

J	Periods	SIG	SIG/SUM	A(J)	B(J)
0	0			0	
1	576	0.2183	0.4655	0.6007	0.2753

2	72	0.0605	0.5944	0.3377	-0.083
3	82.2857	0.0368	0.6728	-0.0788	-0.2595
4	144	0.0368	0.7513	0.2712	-0.001
5	192	0.036	0.828	0.0075	-0.2682

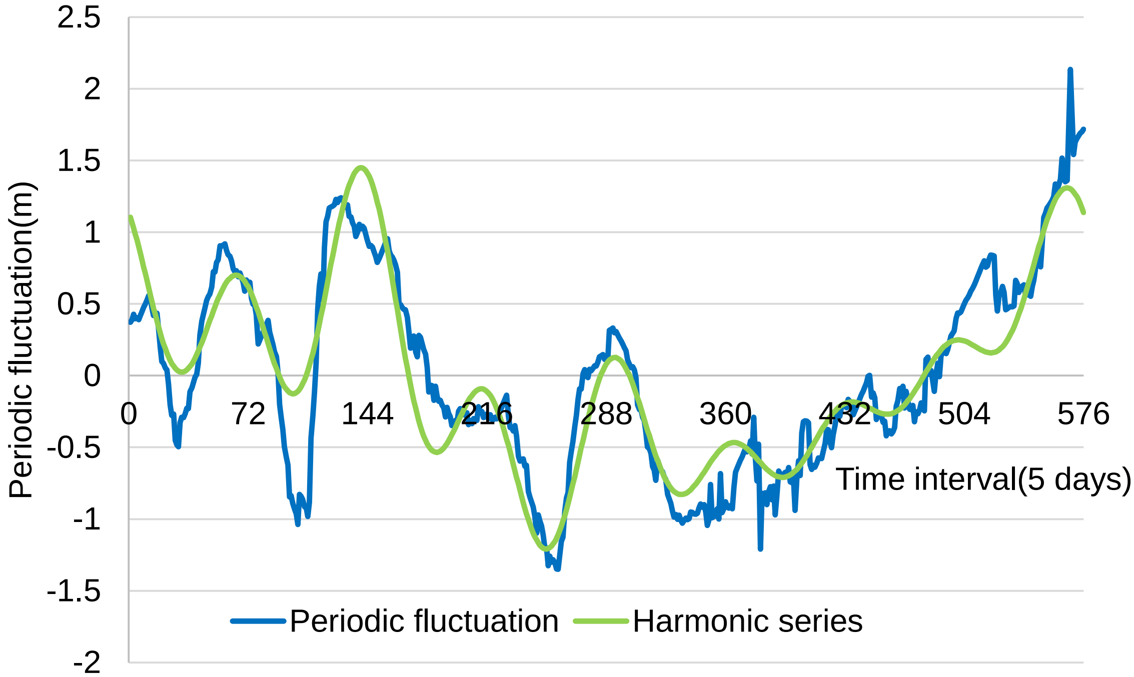


Figure 10. Fit for harmonics to groundwater levels in Well 137

3) Autoregressive analysis

After subtracting the trend and period of the groundwater time series data, the left components are stationary, and the autoregressive model is used to analyze the autocorrelation characteristics of the stationary series. The analysis shows that the stationary series can be simulated with the AR (1) model. The AR (1) model fits well with the actual sequence. The equation of AR(1) model is $Z_t=0.9415*Z_{t-1}$. The fit of the AR(1) model is shown in Figure 11 for observation well 137. The residuals are almost random and independent (Figure 12). The reason for some values of correlation is out of limitations may be the residual of the periodic component. The estimation of period fluctuation is not completely fit for the measured series; this may result in the remaining component is not totally stochastic.

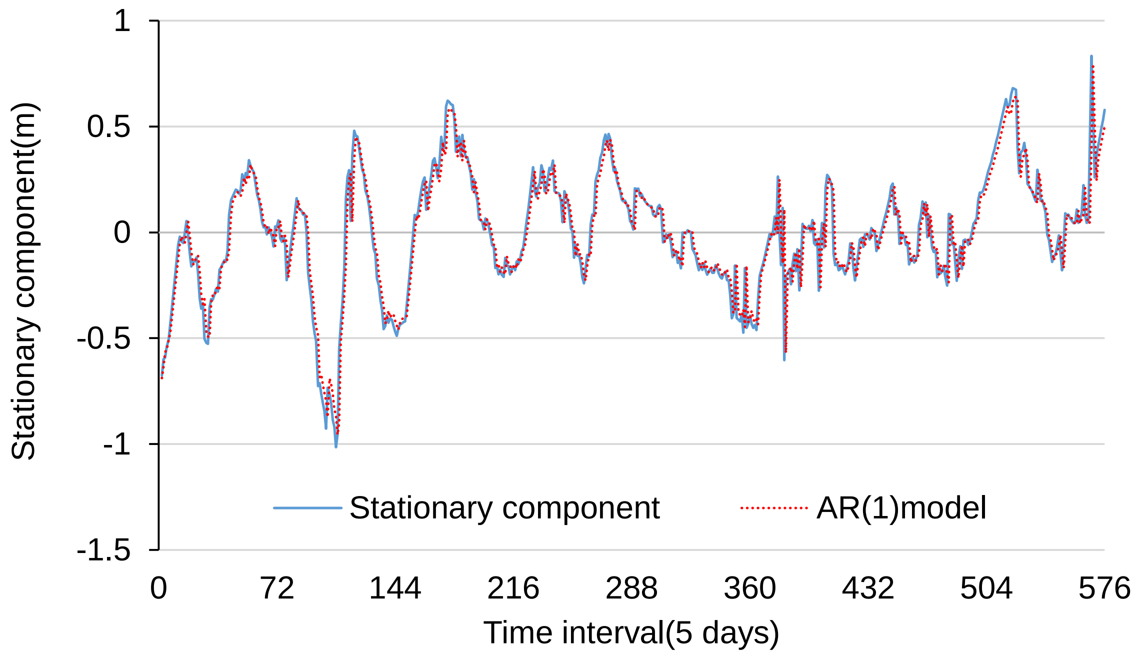


Figure 11. Fit of AR (1) model to the stationary component

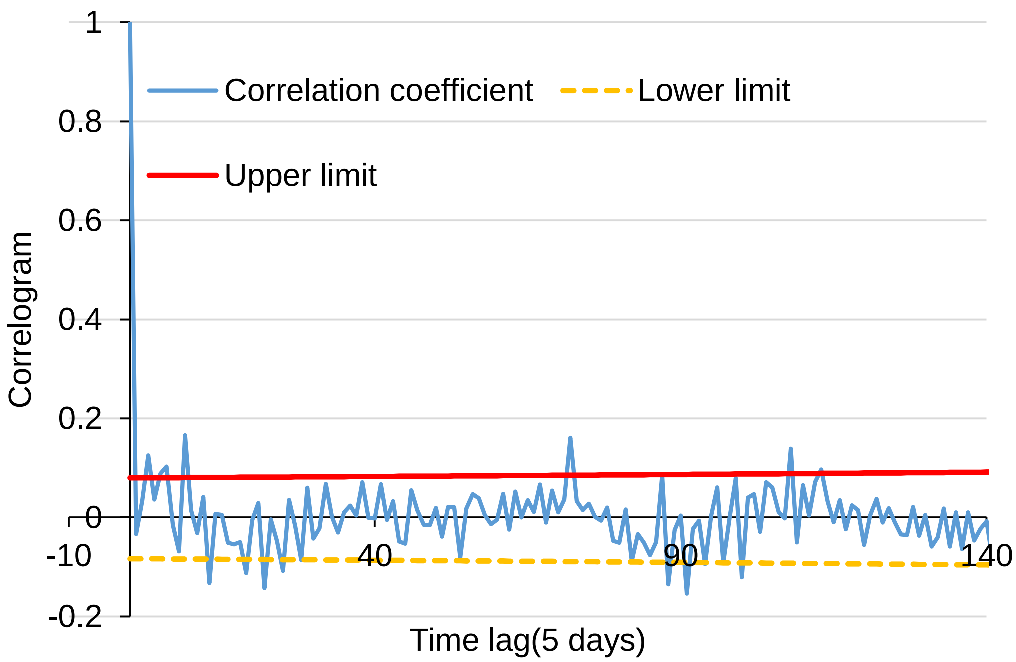


Figure 12. Correlogram of residuals

4) Additive model

The additive model for the groundwater level series of Well 137 is built by integrating trend, harmonic series, and AR (1) model. Figure 13 shows the fit of additive models to groundwater level series for observation Well 137. The model can simulate general variations of groundwater level series very well.

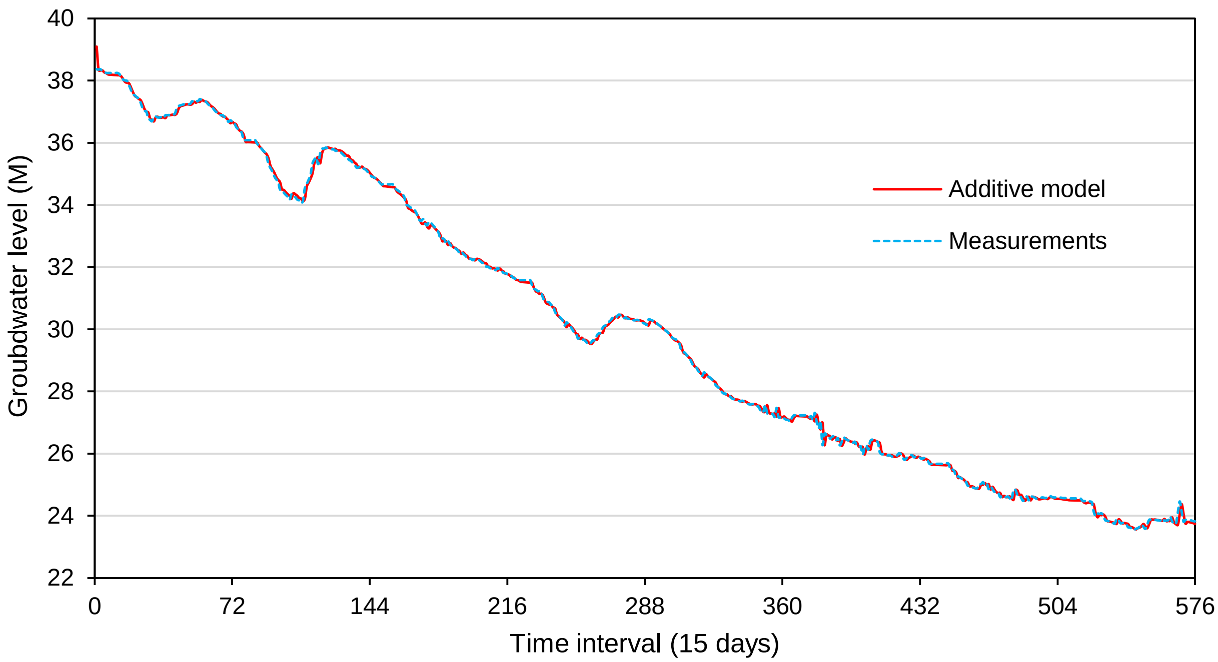


Figure 13. Fit of additive model to groundwater level series

5. Spatial analysis

In the spatial analysis of Beijing Plain, the water level data of all the unconfined aquifer observation wells in one day are used in the Kriging method. Figure 14 shows the spatial distribution of observation wells of unconfined aquifer.

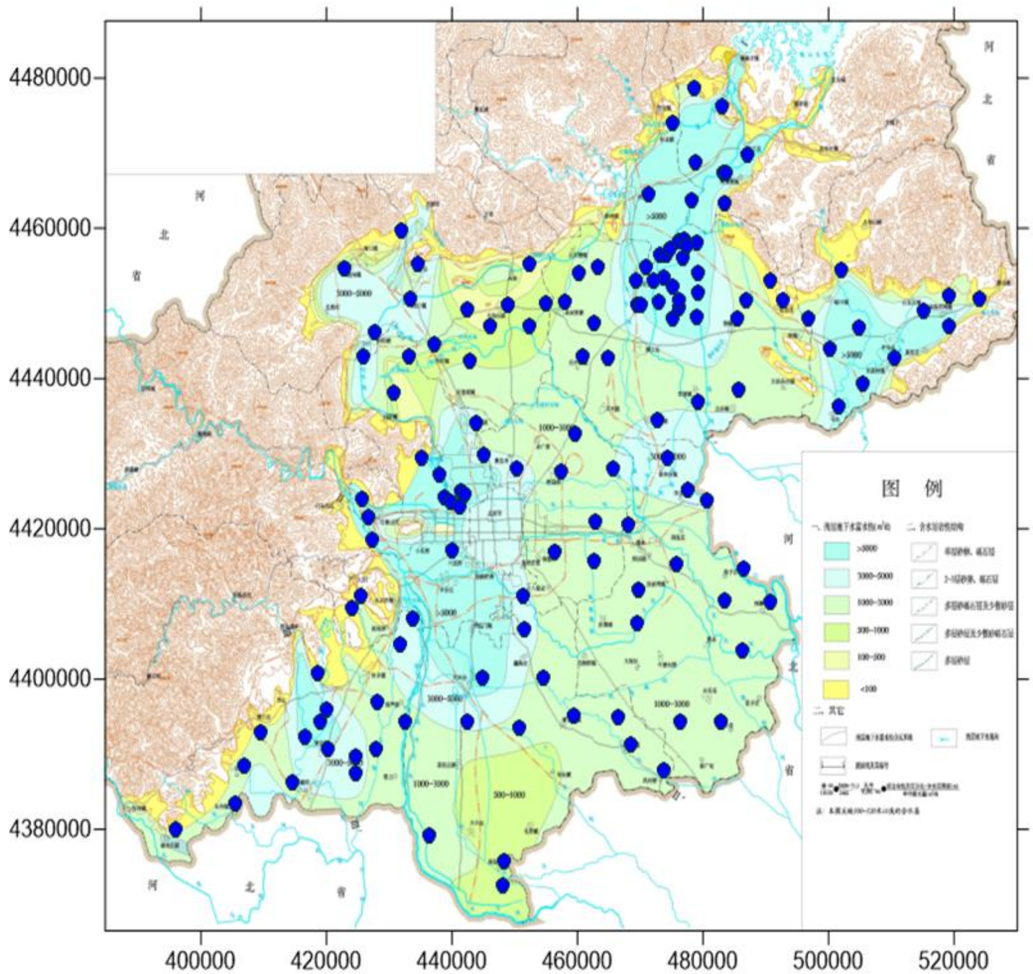
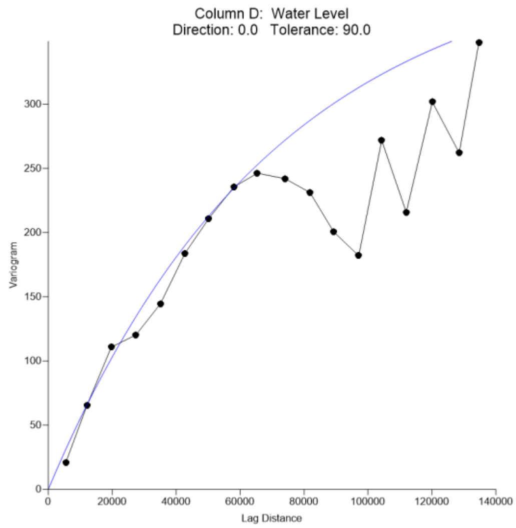


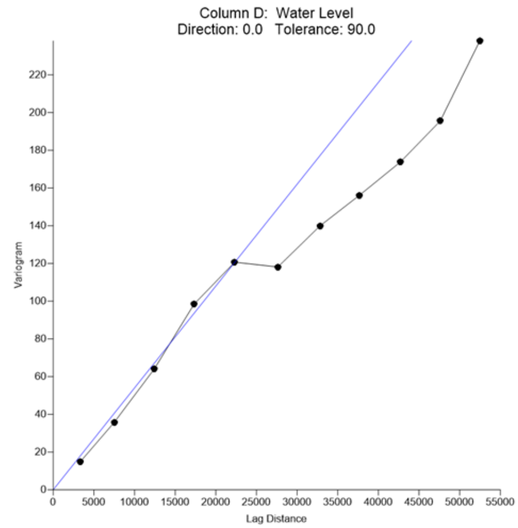
Figure 14. Locations of observation wells of unconfined aquifer in Beijing Plain

1) Fit of variograms

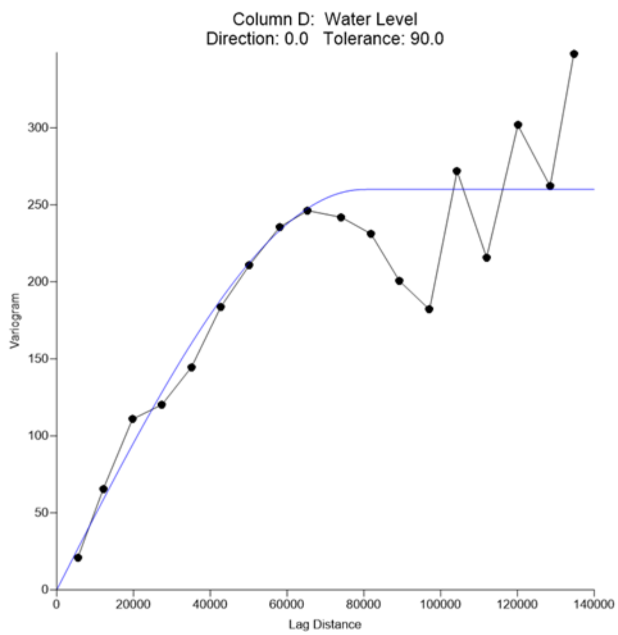
In this group case study, we use three different variograms model, exponential model, linear model, and spherical model to fit. The fits of three variograms models are shown in Figure 15, and the three histograms of residuals are shown in Figure 16 as follows.



a) Exponential model



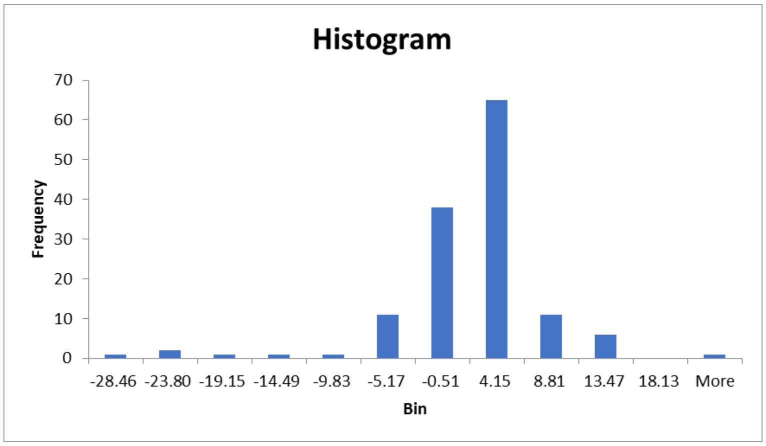
b) Linear model



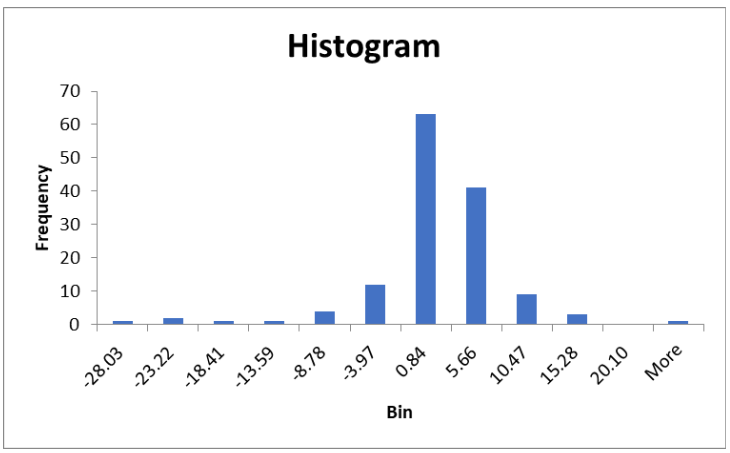
c) Spherical model

Figure 15. The fits of three variograms for water level values

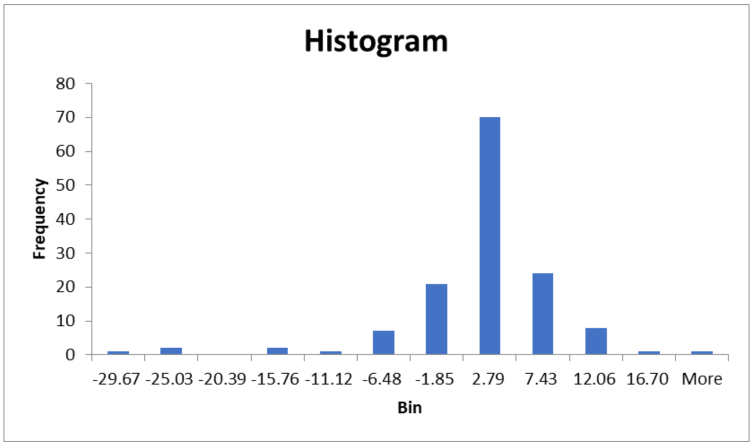
The fitness of three variograms in Figure 16 are similar. For all three variograms, the fit at the beginning part of the lag distance is very good, while at the part of the larger lag distance, the performance of the spherical model is better than the others, which we can also get by comparing the histograms of residuals. The spherical model is the only model in which the histogram of residuals is normal distribution. The cross validation (not shown in this report) also indicates the good performance of three variograms models by calculating the mean error is near zero. However, the ratio of estimation error variance to Kriging variance is much larger than, which is the indication of errors in the larger lag distance.



a) Exponential model



b) Linear model

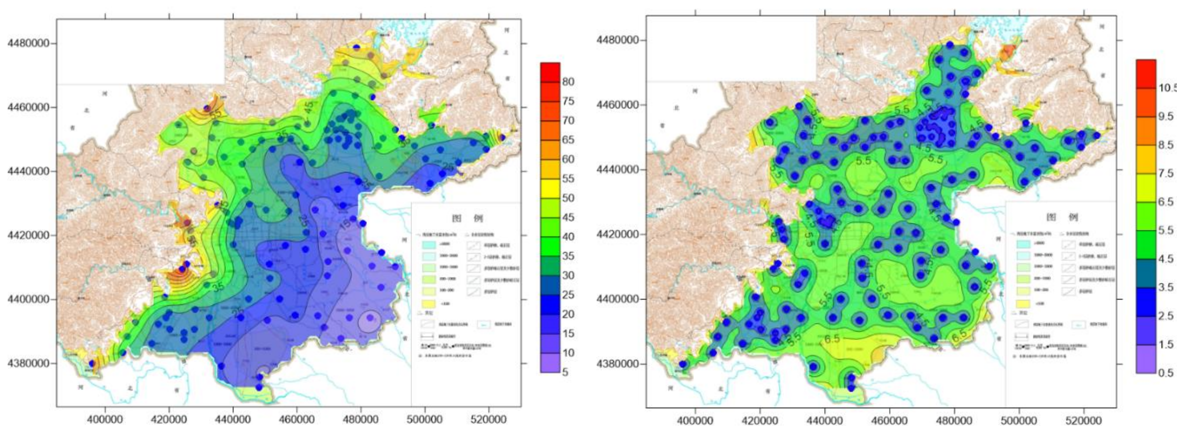


c) Spherical model

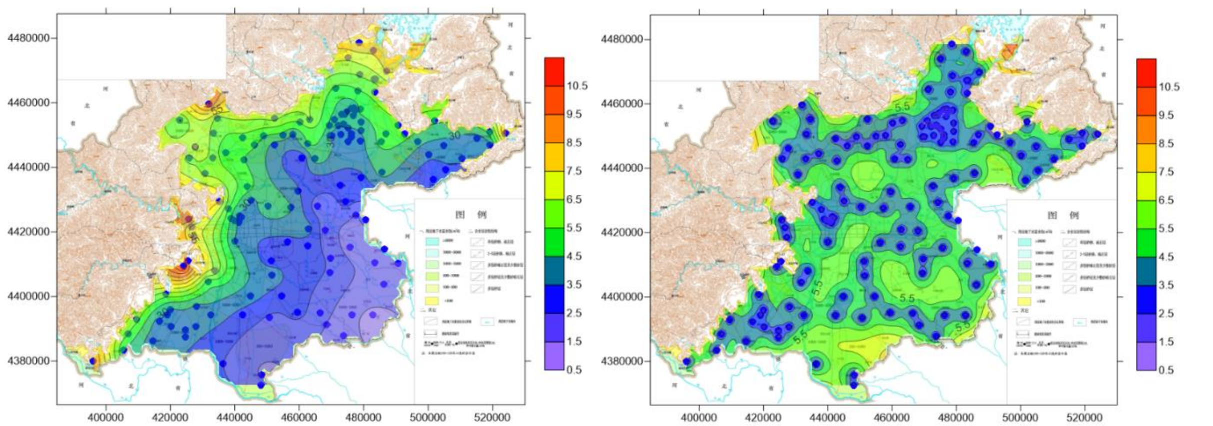
Figure 16. Histogram of residuals

2) Kriging estimation

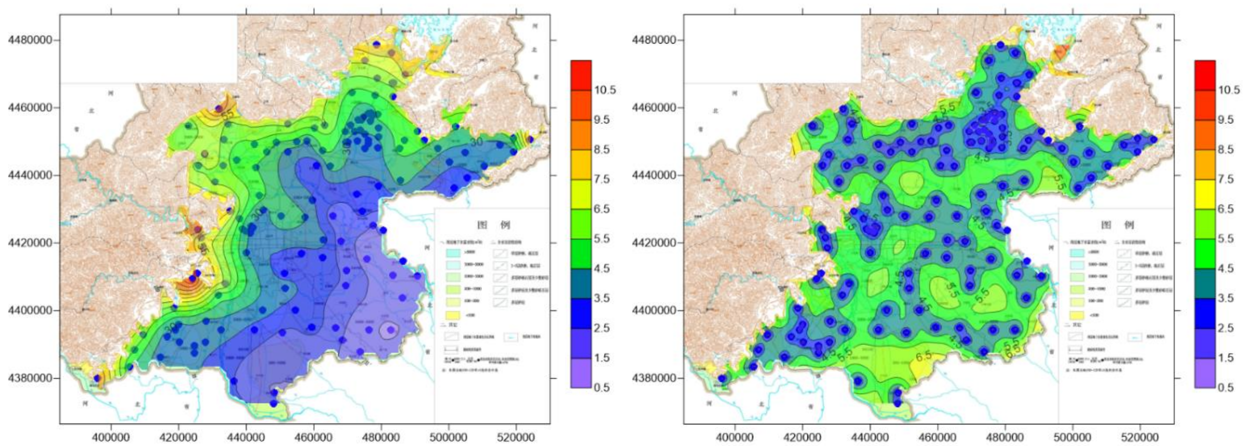
Because the goodness of fit for three variograms is similar, we choose to use all the three variograms to estimate the water level values at the locations where there are no measurements. Figure 17 shows the contour map of water levels and Kriging Standard Deviation in Beijing Plain. The results of the three different models are similar. This is consistent with the results of the fit of variograms and the histogram of residuals.



a) Exponential model



b) Linear model



c) Spherical model

Figure 17. Contour maps of estimates of water level and estimation KSD for water level

The contour maps of estimates of water level can indicate that the water level is higher in the northwest of Beijing Plain and lower in the southeast, which is consistent with the topological condition. And we can indicate that there is some area in the middle of Beijing over abstracting groundwater. One of the advantages od Kriging estimation is that it also can provide the accuracy of the estimation. The Kriging Standard Deviation is also shown in Figure 18. It is clear that estimation error is low in the areas with more measurements, and it is higher in the areas with fewer measurements. From the contour maps of Kriging Standard Deviation, the locations which are lack in observation wells are determined.

6. Conclusions

This group case study focuses on the statistical analysis for the groundwater level data in Beijing Plain, using the methodology, including the regression analysis, time series analysis (additive model simulating), and the spatial analysis (Kriging estimation). Regression analysis indicates the groundwater level in Beijing Plain does not have a good correlation with precipitation. This indicates this area is the discharge part of the aquifer, which is consistent with the topologic information. Time series analysis indicates the obviously decreasing trend and periodic of the groundwater level. And the additive model established in the report can simulate the long term series very well. It gives an available and confident model for forecasting the groundwater level change in Beijing Plain. Kriging estimation gives the results of estimating the groundwater level and the error of the estimation at the locations which do not have the measurements. The Kriging estimation results can be used in much forward work, including the monitoring system optimization.

7. References

Zhou, Y., et al. (2013). "Upgrading a regional groundwater level monitoring network for Beijing Plain, China." Geoscience Frontiers 4(1): 127-138.

Zhou, Y (2009). Hydrogeostatistics. 3rd ed. Delft: UNESCO-IHE, Institute for Water Education.

Zhou, Y (2009). Groundwater Monitoring. 3rd ed. Delft: UNESCO-IHE, Institute for Water Education.