A Method of Simulating Regional Groundwater Level Distribution Based on Digital Elevation Model

ZHAO Zhenhua MA Zenghu HUANG Dan WANG Dongqing XIAO Wei
Northwest Institute of Nuclear Technology, Xi’an, Shaanxi 710024, China
starzzzh@163.com

Abstract. It is very important to use the measured data to obtain the regional groundwater level distribution in groundwater flow modeling, which can provide the initial flow field or boundary condition for modeling, and can be used as the verification basis for the simulation results. Traditional methods are usually interpolated using existing water level data, and the accuracy of the results is greatly influenced by the number of boreholes and the rationality of distribution, so it is not applicable in ungauged region of northwest China. Actually, the flow of groundwater in most areas is affected by the topography at the regional scale, and the terrain and groundwater flow path are correlated. In this paper, a method based on digital elevation model to simulate the regional groundwater level distribution is presented. The selection of the trend surface equation and parameter calibration are the key points and difficulties, because low order polynomial may lead the trend surface equation cannot fit the measured data well, and high order polynomial maybe lead the instability of the solution to the equation. Therefore, this paper established the high order polynomial as the initial trend surface equation, then the correlation between each parameter in the trend surface equation and the elevation of the measured groundwater level was analyzed, and the parameters with obvious correlation with the elevation of groundwater level compose the trend surface equation, and then using the measured data realized the optimal estimation of the average trend level of groundwater, to ensure the best trend surface equation, finally, according to Lanczos singular value decomposition theory, the trend surface equation solution were obtained, and an unbiased and optimal estimation of the residual using the method of universal kriging interpolation was presented. This method was applied in a certain area of northwest China, and the error of the calculation results was much less than that of the kriging interpolation method, which was more consistent with the measured values and could be used to extrapolate. This method has a good prospect in the simulation of groundwater flow field in ungauged region of northwest China.

1. Introduction
It is very important to use the measured data to obtain the regional groundwater level distribution in groundwater flow modeling, which can provide the initial flow field or boundary condition for modeling, and can be used as the verification basis for the simulation results, in which work, the simplest method is to directly interpolate using the groundwater level data, such as Kriging method, inverse distance weighted method, spine interpolation method, radial basis function method, and so on, but these methods are only limited to obtain unbiased optimal interpolation, and ignore the strong correlation between terrain elevation data and groundwater level. Neuman (1984) and Sun (1999) proposed and studied Residual Kriging method to simulate the regional groundwater level distribution[1,2], which provided a good idea for this research, but this method needs a large number of...
measured groundwater level data. Chen (2000) used Co-Universal Kriging method, and using elevation data as auxiliary information, to simulate the regional groundwater level distribution, which preliminarily solved the problem of insufficiency of measured data[3]. Desbarats (2002) directly used digital elevation model as trend surface to simulate the regional groundwater level distribution in a certain area of Canada, but this method ignored the use of measured data, not suitable for areas with poor correlation between topography and groundwater level. Zhu (2004) proposed a corrected Residual Kriging method which uses digital elevation data as auxiliary information to modify the trend surface of groundwater level [5,6], and it was used to simulate the regional groundwater level distribution in Fuyang River basin in the North China Plain, which achieved a good result. In this paper, the correlation between each parameter in the trend surface equation and the elevation of the measured groundwater level was analyzed, and the parameters with obvious correlation with the elevation of groundwater level compose the trend surface equation, then digital elevation model was used to realize the optimal estimation of the average trend level of groundwater and an unbiased and optimal estimation of the residual using the method of universal kriging interpolation was presented, and finally the regional groundwater level distribution was simulated. This method was applied in a certain area of northwest China, and the error of the calculation results was much less than that of the kriging interpolation method, which was more consistent with the measured values and could be used to extrapolate. This method has a good prospect in the simulation of groundwater flow field in ungauged region of northwest China.

2. Research Method and mathematical analysis

During the migration of the groundwater, the flow surface in the main flow direction of the same hydrologic unit has a certain degree of inclination [1]. In addition to the influence of the special geological structure, there is a certain correlation between the flow direction and the direction of the dipping topography [5]. Therefore, the trend surface of the regional groundwater level can be simulated with the feature of the terrain elevation data and groundwater level.

Suppose that \( h(x, y) \) is the measured groundwater level, \( z(x, y) \) is the terrain elevation, where \( (x, y) \) is the horizontal coordinate. Then \( h \) can be expressed as two parts, that is

\[
h(x, y) = \mu(x, y, z) + R(z)
\]

(1)

In the formula, \( \mu(x, y, z) \) is the trend component, indicating the trend level of the groundwater level at \( (x, y) \), and \( R \) indicates the difference between the actual water level and the predicted water level.

2.1. Analysis of the trend component (\( \mu \))

The trend component \( \mu \) represents the average value of the predicted water level expressed by the terrain elevation \( z \) at \( (x, y) \). Where \( z \) is a function of \( x \) and \( y \), because \( \mu \) is related to \( z \), \( \mu \) has a certain degree of correlation with \( x \) and \( y \).

In previous studies, the trend surface equation was adopted in the standard form of one or high order polynomial. In practical simulation of the groundwater level, low order polynomial make the trend surface equation not fit the measured data well, and high order polynomial lead to the extreme sensitivity of the trend surface equation and the large deviation between the predicted water level and the actual water level in some nodes with the increase of the power of the polynomial.

Zhu (2004) carried out some experiments on Fuyang River in the North China Plain, and concluded that the simulated error was small when one order relation was satisfied [5], but the conclusion was still doubtful whether all regions are applicable. In order to research generality, this paper chooses the high order polynomial. When the order of the polynomial is greater than 3, the function will become very complex. On the basis of the groundwater level and the terrain elevation having the continuous slow change and relatively simple relationship, \( m \) represents the order of the polynomial, which is not greater than 3, so the fitting polynomial can be expressed as
\[ \mu(z, x, y) = \sum_{k=0}^{m} (ax + by + cz)^k \] (2)

Where \( a, b \) and \( c \) are the weight coefficients, we can use the optimization to obtain \( m \) that satisfies the optimized residuals (the minimum variance of residuals), that is:

\[ m = \min \{ \text{sum} \{ \text{var}(h - \mu^m(z, x, y)) \} \} \] (3)

Where var is the variance symbol. After obtaining \( m \), the trend component of groundwater level is obtained:

\[ u = \mu^m(z) = \sum_{i=0}^{k} a_i f_i \] (4)

Where \( f_i \) is the monomial, \( k \) is the number of the monomial, \( a_i \) is the coefficient of the monomial, after obtaining the coefficient \( a_i \), it can obtain the optimal expression of the trend groundwater level \( \mu \).

2.2. Analysis of the residuals (R)

After obtaining the trend level \( \mu \), the residual value can be obtained by subtracting the average level from the actual water level, that is:

\[ R = h - \mu \] (5)

In geostatistics, the residual \( R \) is generally required to satisfy the unbiasedness in the region, that is:

\[ E(R_i, i = 1, 2, \cdots, n) = 0 \] (6)

Where \( E \) is the expectation symbol, and \( n \) is the number of compute nodes in the region.

Since we have added the known information of the terrain elevation, the trend component has been optimized expression based on the terrain elevation, but the residual of non-known points is still unknown. The residual part is interpolated in the condition of the unbiasedness and the minimum variance using the kriging interpolation method.

We know that the residuals of \( m \) sample points are \( R_i(i = 1, 2, \cdots, m) \), now use \( R_i \) to estimate \( R \) at any point in the field. \( R^* \) is the estimator of \( R \), which can be expressed as:

\[ R^*(z) = \sum_{i=1}^{n} \lambda_i R(z_i) \] (7)

Where, \( \lambda_i \) is the weight coefficient, and \( n \) is the number of nodes in the computed region. Therefore, the equation set can be obtained in the condition of the unbiasedness and minimum variance.

\[ \sum_{j=1}^{n} \lambda_j C(z_j, z_j) - \sum_{i=0}^{k} \gamma f_i(z_i) = C(z, z_i)(i = 1, 2, \cdots, n) \]

\[ \sum_{i=1}^{n} \lambda_i f_i(z_i) = f_i(z)(l = 0, 1, \cdots, k) \] (8)

Where \( C \) is the covariance, \( \gamma \) is the Lagrange multiplier. The above equation set contains \((n+k+1)\) unknown numbers and \((n+k+1)\) equations, so it can accurately solve \( n \) values of \( \gamma \) to obtain the unbiased and optimal estimator of the residual.

3. Application in a certain area of northwest China

3.1. Topography and geomorphology of study area

The area is low-lying from north to south, with a northwest to southeast tilt. The altitude ranges between 725 m and 2775 m above sea level. From north to south, landforms include low mountains,
denuded hills, an inclined plain and an alluvial lacustrine plain. The mountainous area is the recharge area of groundwater, and the plain area of the basin is the runoff and excretion zone of groundwater. The topographic and geomorphologic three dimensional diagram of the study area is shown in Figure 1.

3.2. The characteristics of groundwater flow field in study area

The study area is large, from the macroscopic scale, it can be considered that the groundwater in the whole area has the unified hydraulic connection, and the difference of different types of groundwater is mainly reflected in the amount of water. The groundwater flow field in the study area is controlled by the topography, and the general flow is from north to southeast, and finally drains into the depression in the southeast. Simulating the groundwater flow field in the study area is mainly based on two types of data, one is the drilling data, and the second is the geophysical exploration results, among them, the boreholes is mainly located at the north of the KLK mountain, and there are few boreholes south of KLK mountain. Therefore, it is impossible to obtain the groundwater level of the study area which is located to south of KLK mountain with traditional interpolation methods. But, from landform, strata structure and geological structure, the south area and the north area of KLK are similar and symmetry in a certain range, as a result, distribution of groundwater level should also have a certain similarity, which can be analogy. Based on this premise, the research design is using boreholes data and digital elevation model to establish the trend surface equation of groundwater level distribution in the north area of mountains with the method proposed in this paper, and then simulate the study area groundwater flow field by the trend surface equation.

3.3. Determination of the surface equation of groundwater level distribution
Thirty four groups measured groundwater level data in the study area can be divided into two groups, thirty groups of which were used to fit to the trend surface, four groups of which were used to test and verify the reliability of trend surface equation, and the boreholes distribution was shown in Figure 2.

3.3.1. Calculating the trend component. According to the above research, when \( m = 3 \), equation (2) is expanded to obtain the general expression of the fitting function of the trend component.

\[
\begin{align*}
\mu &= a_0 + a_1 x + a_2 y + a_3 z + a_4 x^2 + a_5 y^2 + a_6 y z + a_7 x y + a_8 y z + a_9 z^2 + \\
&\quad + a_{10} x^2 z + a_{11} x z^2 + a_{12} y^2 z + a_{13} y z^2 + a_{14} x y z + a_{15} z^3
\end{align*}
\]  (9)

In the formula, \( a_0 \sim a_{15} \) is the weight coefficient of the monomial. The correlation between the groundwater level and DEM data of 30 known measured points in the research area is analyzed, we can see that most of the correlation coefficients are large (Figure 3), in particular, it is strongly associated with the monomial containing \( z \). Therefore, it is reasonable to simulate the trend surface equation using the DEM data as auxiliary information.

According to equation (3), the optimized order of the polynomial can be obtained, namely \( m \) is equal to 2. When the order is 2, the equation (9) is expressed as:

\[
\mu = a_0 + a_1 x + a_2 y + a_3 z + a_4 x^2 + a_5 y^2 + a_6 x y + a_7 x z + a_8 y z + a_9 z^2
\]  (10)

Figure 3. The correlation coefficient between each monomial and the measured water level

The equation (10) has 10 unknown numbers and 30 equations. This is an overdetermined equation set. \( d_1, \ d_2, \ d_3 \cdots d_{30} \) there are 30 sets of measured data to substitute into the trend surface equation, the following linear equations can be obtained.

\[
\begin{align*}
\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{30} \end{pmatrix} &= \begin{pmatrix} a_0 \\ a_1 x_1 \\ a_2 y_1 \\ a_3 z_1 \\ \vdots \\ a_9 z_1^2 \\ a_0 \\ a_1 x_2 \\ a_2 y_2 \\ a_3 z_2 \\ \vdots \\ a_9 z_2^2 \\ \vdots \\ a_0 \\ a_1 x_{30} \\ a_2 y_{30} \\ a_3 z_{30} \\ \vdots \\ a_9 z_{30}^2 \end{pmatrix} \\
&= a_0 \begin{pmatrix} 1 \\ x_1 \\ y_1 \\ z_1 \\ \vdots \\ z_1^2 \\ 1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ z_2^2 \\ \vdots \\ 1 \\ x_{30} \\ y_{30} \\ z_{30} \\ \vdots \\ z_{30}^2 \end{pmatrix} + a_1 \begin{pmatrix} 1 \\ x_1 \\ y_1 \\ z_1 \\ \vdots \\ z_1^2 \\ 1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ z_2^2 \\ \vdots \\ 1 \\ x_{30} \\ y_{30} \\ z_{30} \\ \vdots \\ z_{30}^2 \end{pmatrix} + a_2 \begin{pmatrix} 1 \\ x_1 \\ y_1 \\ z_1 \\ \vdots \\ z_1^2 \\ 1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ z_2^2 \\ \vdots \\ 1 \\ x_{30} \\ y_{30} \\ z_{30} \\ \vdots \\ z_{30}^2 \end{pmatrix} + \cdots + a_9 \begin{pmatrix} 1 \\ x_1 \\ y_1 \\ z_1 \\ \vdots \\ z_1^2 \\ 1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ z_2^2 \\ \vdots \\ 1 \\ x_{30} \\ y_{30} \\ z_{30} \\ \vdots \\ z_{30}^2 \end{pmatrix}
\end{align*}
\]  (11)

The matrix form is:

\[
d = G \cdot n
\]  (12)
Where $d = [d_1, d_2, d_3, ..., d_{30}]$ is a 30 by 1 vector, $G = \begin{bmatrix} 1 & x_1 & \cdots & z_1^2 \\ 1 & x_2 & \cdots & z_2^2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{30} & \cdots & z_{30}^2 \end{bmatrix}$ is a 30 by 10 matrix, and $n = [a_0, a_1, ..., a_9]^T$ is a 10 by 1 vector.

Therefore, solving the trend surface equation is solving $n$ of equation (12), as the random measured data and the large difference between each item in the matrix $G$ lead to a singular matrix and a pathological equation. Thus equation (12) cannot be solved linearly according to the ordinary method.

The singular value decomposition is used to solve the equation (12). According to Lanczos singular value decomposition theory, any $M$ by $N$ matrix $G$ can be decomposed into the following form:

$$G = U_r S_r V^T_r$$

(13)

Where $r$ is the matrix rank, $S_r$ is a diagonal matrix composed of $r$ non-zero characteristic roots of $G^T G$ or $G G^T$, $U_r$ is a $M$ by $r$ eigenvector matrix of $G G^T$, $V_r$ is a $N$ by $r$ eigenvector matrix of $G^T G$. So we can obtain the inverse matrix $G_L$ of the matrix $G$ as long as you obtain the above quantity.

$$G_L = V_r S_r^{-1} U_r^T$$

(14)

Finally, the solution of equation (12) can be obtained:

$$n = G_L d = V_r S_r^{-1} U_r^T d$$

(15)

Using the singular value decomposition method the fitting coefficient ($a_0, a_1, ..., a_9$) can be obtained (Table 1), and the water level of the fitting point is calculated using the obtained trend equation. The calculated value and the measured value are projected onto the scatter diagram, as shown in Figure 4, there is a good consistency. The average value of the trend water level at any point in the region is predicted using the coefficient in Table 1 (Figure 5).

| Component | Value  | Component | Value   |
|-----------|--------|-----------|---------|
| $a_0$     | -3.18e+2 | $a_5$     | -1.52e-8 |
| $a_1$     | -3.99e-3 | $a_6$     | 5.50e-8  |
| $a_2$     | 3.13e-3  | $a_7$     | 1.95e-6  |
| $a_3$     | 1.62     | $a_8$     | -4.27e-6 |
| $a_4$     | 4.96e-9  | $a_9$     | -2.55e-4 |

(16)

It is seen from Figure 5, this set of coefficients is adopted to estimate the sag, plain and mountainous region in the research area. The average value of the groundwater level in the entire region is 1232m (The average value of the known measured point is 1276 m), the minimum value of the groundwater level is 674 m, and the maximum value is 2205 m (The minimum and maximum of the terrain elevation are 697m and 2764m respectively), comparing the known measured point with the terrain elevation, this trend level is acceptable.
3.3.2. Simulating the residuals. The trend surface is reasonable to estimate the average value of the regional groundwater level, and the residual part can further modify the accuracy of the trend surface. For 30 known points in the research area, the residual of the known point is the measured water level subtracting the trend level. The residuals are roughly distributed between -36m and 24m, the residuals close to 0 are the highest frequency (Figure 6). It is known from equation (1) that the groundwater level is the trend level plus the corresponding residual. In order to simulate the groundwater level in the entire research area, the residual error in this research area need to be simulated. Because the terrain elevation data is used as the known information to simulate the trend component, so it doesn't make sure that the residual part must conform to normal distribution, and the mean trend of the residuals in the whole space is unknown. For the interpolation problem of this kind of data, the universal kriging interpolation method gives a very good solution, which is using equation (8) to limit the spatial interpolation points, thus the unbiased and optimal residual interpolation in this region is obtained. The groundwater level distribution in the entire research area can be obtained by summing the value of the previous trend component with the corresponding residual (Figure 7).
3.4. Reliability test

The four measured points in the research area are selected for testing and compared with the kriging interpolation with no terrain elevation data (Table 2). By contrast, it is considered that the error based on DEM is lower than that of the kriging interpolation method in this paper, with some advantage in accuracy. In addition, the greatest advantage of DEM is that, after establishing the trend surface equation of groundwater level, this equation can be used to extrapolate the groundwater level distribution in the surrounding region. It is a very good way and method to obtain the regional groundwater level distribution without the borehole, which is impossible for the kriging interpolation method.

Table 2. Comparing the elevation of water level based on DEM with that of the kriging interpolation

| Drill point number | The measured water level elevation /m | The water level elevation based on DEM /m | The water level elevation of the kriging interpolation /m | The error elevation based on DEM /m | The error of the kriging interpolation /m |
|--------------------|-------------------------------------|-----------------------------------------|-----------------------------------------------------------|-----------------------------------|------------------------------------------|
| Y1                 | 1347                                | 1348                                    | 1338                                                      | -1                                | 9                                        |
| Y2                 | 1399                                | 1400                                    | 1391                                                      | -1                                | 8                                        |
| Y3                 | 1387                                | 1377                                    | 1373                                                      | 10                                | 14                                       |
| Y4                 | 1433                                | 1427                                    | 1423                                                      | 6                                 | 10                                       |

3.5. Simulation of groundwater level distribution in study area

Through the above research work, obtained the trend surface equation of regional groundwater level distribution, after verification, it was considered that the equation can be used for simulating groundwater flow field of study area. However, since the trend surface equation was obtained based on the boreholes data north of KLK mountain, it was only applicable to areas whose geological condition was similar to north of KLK mountain. Therefore, according to field investigation and remote sensing image analysis of the geological conditions, it was considered that the region between the KLK ridge line and the X fault had the similar geological conditions with the area where the boreholes were located. On this basis, the Global mapper software was used to extract the elevation grid data from digital elevation model of study area, then the elevation data were plugged into the trend surface equation, and the calculated groundwater level values of every grid were obtained. Using Surfer software to map the grid data, the groundwater flow field map of study area was finally obtained, as shown in Figure 8.

![Figure 8. The groundwater flow field map of study area](image_url)
4. Conclusions

(1) When there is a certain correlation between the terrain and the groundwater level, it is feasible to use the digital elevation model as an auxiliary means to simulate the regional groundwater flow field. In this work, the selection of the trend surface equation and parameter calibration are the key points and difficulties, which can be solved by the analysis of the correlation between each parameter in the trend surface equation and the elevation of the measured groundwater level, and then the best trend surface equation can be obtained. For the residual part of the equation, an unbiased and optimal estimation can be carried out using the method of universal kriging interpolation.

(2) In this paper, the groundwater flow field in a certain area of northwest China was simulated based on the trend surface equation of regional groundwater level distribution, by comparison, the accuracy was better than that of kriging interpolation.

(3) In practical application, the groundwater level trend surface equation can be used to extrapolate to surrounding areas with similar topography and geological conditions. This method has a good prospect in the simulation of groundwater flow field in ungauged regions of northwest China.

References

[1] Neuman S P, Jacobson E A. Analysis of nonintrinsic spatial variability by residual kriging with application to groundwater levels. Math. Geol., 1984, 16(5): 499-521.

[2] Sun, Ne-Zheng. Inverse problems in groundwater modeling. Kluwer Academic Publishers, the Netherlands, 1999, pp: 161-210.

[3] Chen J J, Guo Q Y, Wang H Q, Wang J S. The application of Co-Universal Kriging method for estimation of groundwater table. Journal of Hydraulic Engineering, 2000, 31(7): 7-14.

[4] Desbarats A J, Logan C E, Hinton, M J, Sharpe D R. On the kriging of water table elevations using collateral information from a digital elevation model. J. Hydrol., 2002, 265(1-4): 25-38.

[5] Zhu K, Zhang X W, Xia J et al. Simulation of groundwater level distribution in a large area using digital elevation model as auxiliary information. Journal of Hydraulic Engineering, 2004, 35(11): 15-21.

[6] Zhu K. Groundwater numerical simulation within a large-scale domain — A case study for the Fuyang River basin in north China. Wuhan University, 2004.