



Original Article

On the Influence of the Soil and Groundwater to the Subsidence of Houses in Van Quan, Hanoi

Dinh Xuan Vinh*

Hanoi University of Natural Resources and Environment, 41 Phu Dien, Tu Liem, Hanoi, Vietnam

Received 11 January 2020

Revised 14 April 2020; Accepted 22 August 2020

Abstract: The area of Van Quan, Hanoi before 2004 was the rice field. Nearby, Ha Dinh water plant has well-drilled underground water for residential activities. Van Quan's new urban area after being formed has detected many subsidences. The objective of this study is to assess the main causes of the subsidence of the houses, based on groundwater and soil. This paper applied the regression method to study the effect of soil and groundwater on the residential constructions in Van Quan urban area, Hanoi. Subsidence monitoring was carried out for 4 consecutive years, from 2005 to 2009, including over 500 subsidence monitoring points with high-precision Ni007 and INVAR gauges. A groundwater observation well is 30 meters deep at the site of the settlement. The results show a small effect of groundwater on subsidence. The characteristics of the young sediment area and the soil consolidation process are the main causes leading to serious subsidence in residential constructions in Van Quan urban area. This paper provides a different perspective on the impact of groundwater on the subsidence of residential structures within approximately 100 ha.

Keywords: monitoring, subsidence, residential houses, groundwater, soil.

1. Introduction

The situation of land subsidence in the region due to various subjective and objective causes that many scientists as Tuong The Toan, Tu Van Tran, Ty Van Tran [1-3] agreed as follows: Characteristics of sedimentary basins during consolidation, denudation or accretion of topographic surfaces, groundwater extraction

activities, and construction process. urban floor. In this paper, we want to explore the impact of groundwater on the upper floor and the consolidation process of soil on shallow foundation constructions, in particular, houses under 5 floors in Van Quan urban area, Hanoi. We have built a groundwater monitoring well with a depth of 30 meters in the survey area. Observation data of groundwater and subsidence

* Corresponding author.

E-mail address: dxvinh@hunre.edu.vn

<https://doi.org/10.25073/2588-1094/vnuees.4539>

of residential houses of Van Quan urban area were conducted regression analysis. Thereby we assess the influence of each cause to the settlement of the houses on the young sedimentary basin.

Some studies use the method of Terzaghi as Ty Van Tran, Hiep Van Huynh [3], or the Finite Element method as Tu Van Tran et al [2], based on groundwater monitoring data to forecast ground subsidence. In this study, we use the groundwater monitoring data in the subsidence area (about 100 hectares) and the subsidence monitoring data of the houses according to national Class II leveling Regulation. Conducting the regression analysis for each cause of subsidence. The first is groundwater. The second is the during consolidation subsidence of the soil because Van Quan urban area is located on a young sedimentary basin [2].

2. Research Methods and Data

The raw monitoring data including appropriate measurements is a very important part of the building safety data. Based on the monitoring data, one can recheck the design plan as well as the construction process and the operation of the building. The raw data provide valuable information that sheds light on the stability of the building. However, the raw data cannot reveal the shifting field or the deformation trend of the building. A comprehensive analysis is therefore needed to accurately and comprehensively identify various deformations from a large volume of raw data. Two types of dynamic models are formulated to analyze deformation monitoring test data, non-parametric models based on mathematical-statistical theory, and principles-based parametric models major of continuous mechanics.

Non-parametric model based on mathematical - statistical prediction algorithms. The first model is based on a functional relationship between the independent variables (the environment variables) and the dependent variables (are the deformations). Models of this

type can be interpreted as internal causes and results within the system. This format includes multiple regression (MR) model, stepwise regression (SR), principal component regression (PCR), partial least square regression (PLSR) and artificial neural network (ANN). The second model is based on the statistical rule of dependent variables ie using linear statistical models themselves, not by other environment variables. They do not establish a model between cause and effect. This type includes Time series (TS series), Gray system (GS). The deformation prediction model is based on information drawn from the deformation monitoring data series, these processes are performed in different ways.

Parameter model based on the analysis of monitoring data by continuous mechanical rules. First, determine the relationship between the dependent variables and the independent variables built on mechanical rules. Next, linear statistics are applied to correct the assumed calculation values or parameters throughout the calculation. This model type has a Kalman filter [4].

Regression analysis is a statistical method where the expected value of one or more random variables is predicted based on the condition of other (calculated) random variables. Regression analysis is not just about curve matching (choosing a curve but best matching a set of data points); it must also coincide with a model of deterministic and stochastic components. The defined component is called the predictor and the random component is called the error term. Regression analysis is both a mathematical-statistical method and a deformation physics explanation, so it can be used to predict deformation. Calculation of univariate or multivariate regressions is the solution a system of linear equations based on the least-squares principle the functional model is represented as a matrix.

$$Y = X\beta + \varepsilon \quad (1)$$

In this model, Y is a dependent variable, that is, the vector of deformation measurement, matrix representing the component of the

dependent variable is $Y^T = (y_1, y_2, \dots, y_n)$, n is the amount of measurement; Equation (1) has many variables x and each variable has a parameter β that needs to be estimated; The vector of random error ε is the deviation of measured value (RMS measured value), $\varepsilon^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$. Where the measurements are random components and follow the standard distribution rule $N(0, \sigma^2)$, we can apply the Gauss - Markov procedure. The random model is

$$\left. \begin{aligned} \sum \varepsilon \varepsilon^T &= E\{\varepsilon \varepsilon^T\} = \sigma^2 Q_{\varepsilon\varepsilon} \\ Q_{\varepsilon\varepsilon} &= I \end{aligned} \right\} \quad (2)$$

X is a matrix of the form

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \quad (3)$$

Matrix (3) shows m deformation-causing factors, each deformation-causing factor represents a measure of an independent variable or its function, they form the elements of the matrix X , similar for the dependent variable there are all n groups;

β is the regression coefficient vector, $\beta^T = (\beta_0, \beta_1, \dots, \beta_m)$. Where:

β_0 is the coordinate origin coefficient;

β_1 is the slope coefficient of Y according to the variable x_1 and keeping the variables x_2, x_3, \dots, x_m constant;

β_2 is the slope coefficient of Y according to the variable x_2 and keeping the variables x_1, x_3, \dots, x_m constant;

...

β_m is the slope coefficient of Y according to the variable x_m and keeping the variables x_1, x_2, \dots, x_{m-1} constant.

The slope coefficient β_1 represents the change in the mean of Y per unit of change of x_1 regardless of the change of x_2, x_3, \dots, x_m , so the β_j is also called partial regression coefficients.

For multivariate linear regression equations, we find the estimate $\hat{\beta}$ by the least-squares method so that

$$\sum_i (y_i - \hat{y}_i)^2 = \|Y - \hat{Y}\|^2 = \|e\|^2 = \min$$

We obtain vector

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

and posterior accuracy

$$\Sigma_{\hat{\beta}\hat{\beta}} = \sigma_0^2 Q_{\hat{\beta}\hat{\beta}} = \sigma_0^2 \cdot (X^T X)^{-1}$$

Elements on the diagonal of the covariance matrix $\Sigma_{\hat{\beta}\hat{\beta}}$ are the variances of the estimates β_j ie $q_{\hat{\beta}\hat{\beta}} = S_{\beta_j}^2$.

Post-regression values

$$\hat{Y} = Y + V = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY$$

The H -matrix is called the "hat" matrix [5].

The principles of a multivariate linear regression model and solutions are consistent with the indirect adjustment model and the common solution in surveying, but different in that: the number of causes of deformation influence in the multivariate linear regression model has not been predetermined, it is necessary to use a certain method to defined regression, making the optimal regression model.

In linear regression analysis, we include the following concept: Residual Sum of Square (Q), Total Sum of Square (S) and Explained Sum of Squares (U). We have

$$Y - \hat{Y} = (Y - \hat{Y}) \quad (4)$$

The concepts are defined as follows:

$$\left. \begin{aligned} S &= (Y - \bar{Y})^T (Y - \bar{Y}) = \sum_{i=1}^n (y_i - \bar{y})^2 \\ Q &= (Y - \hat{Y})^T (Y - \hat{Y}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = V^T V \\ U &= (\hat{Y} - \bar{Y})^T (\hat{Y} - \bar{Y}) = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \end{aligned} \right\}$$

Where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

\hat{y}_i is the regression value of the dependent variable.

Can prove that: $S = Q + U$

In regression, the correlation coefficient (R) is a statistical index that measures the degree of correlation between deformation-cause factors and measured deformation values [6]. The correlation coefficient is close to 0, meaning that

the deformation-cause factor and the measured deformation values are not related to each other. If the coefficient is close to -1 or +1, the deformation-cause factor and measured strain value have a great relationship. We have

$$R^2 = U/S$$

We have conducted a groundwater monitoring of a well built in the urban area of Van Quan (Figure 1). Simultaneously with monitoring the subsidence time of the houses (Figure 2), we conduct monitoring the groundwater level (Figure 3).



Figure 1. Groundwater monitoring well.



Figure 2. Cracks on Van Quan houses.

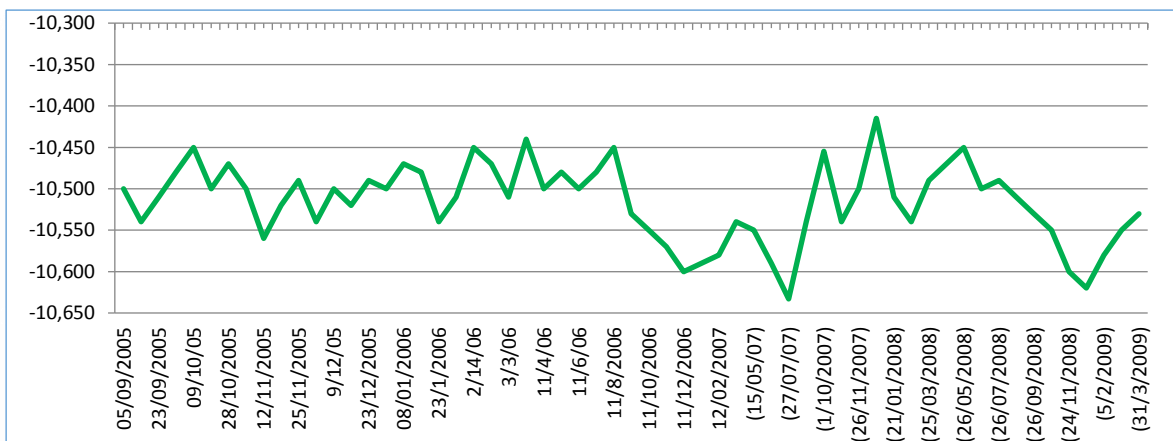


Figure 3. Groundwater level in Van Quan area during monitoring of subsidence.

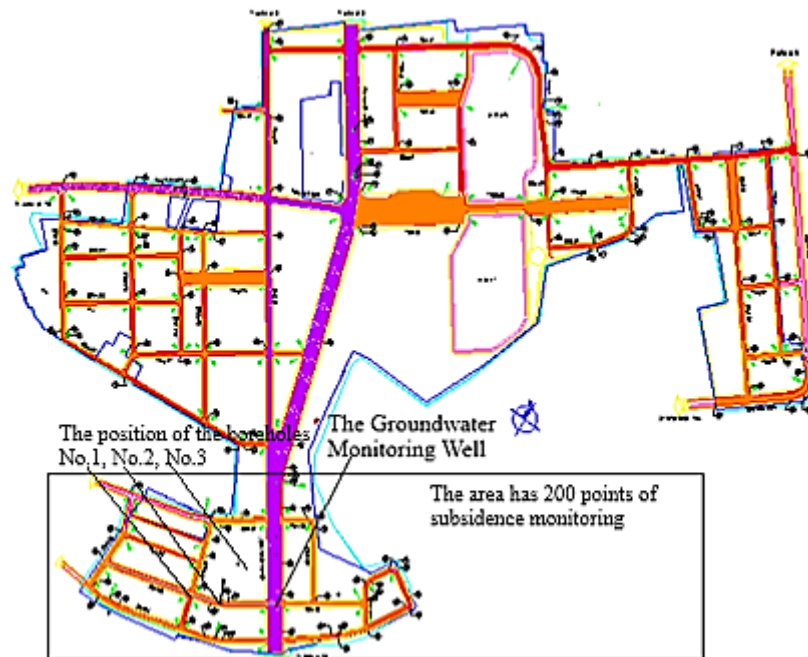


Figure 4. The groundwater monitoring well, the points of measurements and the boreholes.

Monitoring data from May 2005 to March 2009. Subsidence monitoring is done by high-precision leveling Ni007 and Invar gauges. The measurement technique complies with the national grade II standard. Monitoring the groundwater level with the Piezometer gauge. (Figure 4).

3. Theory and Calculation

Methods of assessing the conformity of the regression model according to mathematical statistics include: Calculating the correlation coefficient R , using statistical tests to evaluate the overall model, calculating standard errors of estimates, statistical tests list each individual independent variable. In geodesy, we are interested in testing the overall regression model and testing the dominance of each deformation effect factor (such as temperature, time, pressure,...) on the dependent variable (deformation values).

The regression model we build is based on a finite set of measurement data, so it may be affected by measurement errors ε . We have the following hypothesis

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_1: \text{Have at least one coefficient } \beta_j \neq 0$$

If the assumption H_0 is true, that is, all slope coefficients are zero, then the regression model built has no effect in predicting or describing the dependent variable. Formulation

$$F_{tt} = \frac{\frac{U}{m}}{\frac{Q}{(n-m-1)}} \quad (5)$$

In this formula, U and Q are known, n and m are sample size (number of measurements) and independent variable (number of factors affecting deformation into the model), respectively. The degree of freedom of the numerator $f_1 = m$, the degree of freedom of the denominator $f_2 = (n-m-1)$. Select the confidence level for the F statistic with 95%, that is, the

alpha level for the test is 5%. Look up distribution table F to find the limit value $F_{f_1, f_2, \alpha}$. If $F_{tt} > F_{\text{limited}}$, reject the H_0 hypothesis. The F statistic must be used in combination with the significance level value when you are deciding if your overall results are significant.

Test the dominance of each factor affecting deformation (such as temperature, pressure, time,...) to the dependent variable (is the measured deformation value). We have the following hypothesis

$$H_0: E(\hat{\beta}_j) = 0$$

$$H_A: E(\hat{\beta}_j) = \hat{\beta}_j \neq 0$$

Create the following statistics according to the T distribution

$$T = \frac{\frac{\hat{\beta}_j^2}{q_{\hat{\beta}_j \hat{\beta}_j}}}{\frac{Q}{(n-m-1)}} < T_{n-m-1, \frac{\alpha}{2}} \quad (6)$$

$q_{\hat{\beta}_j \hat{\beta}_j}$ is the j^{th} element on the main diagonal of the matrix $Q_{\hat{\beta} \hat{\beta}}$, where $q_{\hat{\beta}_j \hat{\beta}_j}$ is the variance of the regression coefficient estimates ($S_{\hat{\beta}_j}^2$); Q is the residual sum of square. Look at the distribution table of T, get significance level of 5%, dominance of deformation influence coefficient $\hat{\beta}_j$ is 95% respectively. If $T < T_{n-m-1, \frac{\alpha}{2}}$, then the corresponding deformation-cause factor x_j has a very small effect on deformation, which can be removed from the regression equation.

In the regression model, we must put the deformation-cause factors into the regression equation. In the process of testing their dominance, if any factors do not pass the test, they will be removed, and other factors must be included in the evaluation model. Assume a following multivariate linear regression equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m$$

The residual sum of squares and the explained sum of squares is Q_{m+1}, U_{m+1} , now we have

$$\begin{cases} \Delta Q = Q_m - Q_{m+1} \\ \Delta U = U_m - U_{m+1} \\ \Delta Q = \Delta U \end{cases}$$

Thus, the residual sum of squares increases by the reduction of the explained sum of squares after increasing the deformation-cause factor x_{m+1} , through which the regression equation also reflects the contribution of the additional increase factor with the regression effect. The predominance test for the added deformation-cause factor is as follows

$$H_0: E(\hat{\beta}'_{m+1}) = 0$$

$$H_A: E(\hat{\beta}'_{m+1}) = \hat{\beta}'_{m+1} \neq 0$$

Forming the F statistical distribution

$$F = \frac{\Delta Q}{Q_{m+1}/(n-m-2)}$$

$$= \frac{\Delta Q/(n-m-2)}{Q_{m+1}} \sim F_{1, n-m-2} \quad (7)$$

Taking the significance level of 5%, when $F > F_{1, n-m-2, \alpha}$, the original hypothesis is accepted, that is, the increased deformation-cause factor has a significant effect on the house's deformation, in contrast. it should not be added. In the regression equation, the influence factors of deformation often correlate with each other, that is, there is some relation to each other. The close correlation between the variables in the regression model created a multicollinearity phenomenon, making the variance of the regression coefficient estimates big valuable. The multicollinearity phenomenon also reverses the regression coefficient, instead of positive coefficients, that is, the high water level causes the deformation of the dam to be large, resulting in negative results, the high water level makes the dam less deformed.

Based on the above test steps, it is possible to induce the following step regression:

a) Prequalification of independent variables affecting the deformation

b) Determine the first univariate linear regression equation. Assuming that m

independent variables affect deformation, each of these independent variables creates a univariate linear regression equation, for a total of m equations. Calculate the residual sum of squares Q of each equation. If the regression equation with $Q_k = \min\{Q_i\}$, $i = \overline{1, m}$, then the regression equation with Q_k is collected after testing its according to equations (6) and (7).

c) Determine the best two-variable regression equation based on the univariate linear regression equation, in turn increasing the independent variables affect deformation, and have $(m-1)$ two linear regression equations. Calculate $(m-1)$ the residual sum of squares ΔQ , consider the difference $\Delta Q_j = \max\{\Delta Q_i\}$, $i = \overline{1, m}$.

The j^{th} incremental independent variable is the “waiting” independent variable, conducting its test, if adopted, it will be included in the equation. It is the best two-variable linear regression equation. If not, then stop at the univariate regression equation.

d) If two independent variables affecting deformation are dominant for dependent variable Y (amount of deformation), then according to the above method, continue to select independent variables to affect the third and fourth deformation,... So on until it is impossible to increase the new independent variable and can not remove any independent variables selected, then stop. As a result, we have the best regression model.

The independent variable affecting deformation is groundwater and time. The observation time characterizes the deformation of the test point over time, so its first-order differential is the subsidence rate, its second-degree differential is the subsidence acceleration. Simultaneous time represents the level of consolidation of the soil under the construction. It can be said that: the consolidation subsidence time lasts correspondingly the soil belongs to young sediments.

Develop a regression equation for groundwater variable γ and for time variable θ . We have a linear regression equation for groundwater

$$\hat{Y} = \beta_0 + \beta_1 x_\gamma$$

The linear regression equation for time

$$\hat{Y} = \beta_0 + \beta_2 x_\theta + \beta_3 x_{2\theta}$$

Based on the observed data series we have the following regression equation

- For the effect of groundwater on the subsidence of houses

$$\hat{Y} = 9876.1124 + 309.3856 x_\gamma + 56.5974$$

The correlation coefficient $R^2 = 0.0628 = 6.28\%$, that is, the water table affects only 6.28% to the subsidence of the structure. The posterior error of regression is 56.5974 mm. The posterior error of the estimated coefficient β_1 is $S_{\beta_1} = 158.29$. The test value according to (6) for β_1 is $T = -1.9545$, corresponding to the significance level of 5.55%. The correlation coefficient is too low and the post-estimation error S_{β_1} is too high, so we remove the groundwater element from the regression model.

- For the effect of soil consolidation time on the subsidence of the houses

$$\hat{Y} = 6694.9641 - 1.4108 x_\theta + 0.0024 x_{2\theta} + 6.7862$$

The correlation coefficient $R^2 = 0.9809 = 98.09\%$, ie the time of soil consolidation affects 98% of the settlement of the building. The slope coefficient β_2 indicates the settlement rate and β_3 indicates the settlement acceleration is 0.0024 mm²/week. The posterior error of the regression is 6.7862 mm. The posterior error of the estimated coefficient β_2 is $S_{\beta_2} = 0.0422$, the coefficient β_3 is $S_{\beta_3} = 0.0002$. The test value according to (6) for β_2 is $T = -33,4372$, corresponding to the significance level of 6.6.10-74%, and β_3 is $T = 9.8588$, corresponding to the significance level of 2.3.10-16%, the value This is very small by our standards (5%).

4. Results and Discussion

Based on the results of regression analysis of the causes of subsidence of residential houses, the groundwater level and the time of consolidation of the soil from 2005 to 2009, we can draw a regression line of subsidence according to the consolidation time of the soil background.

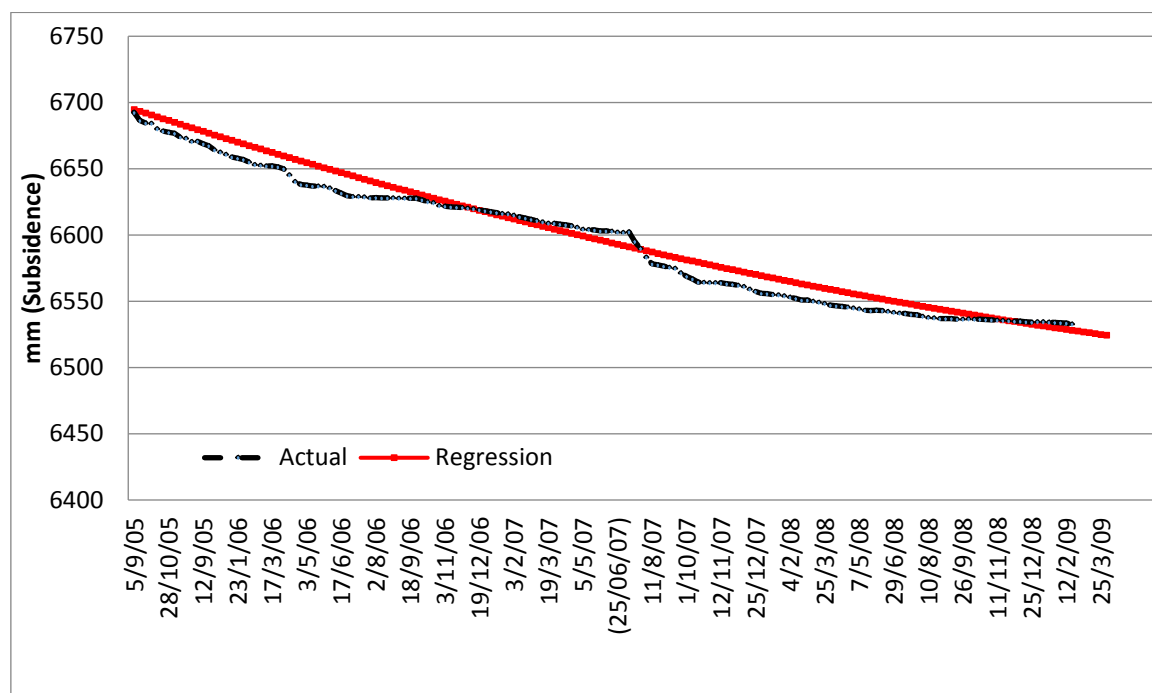


Figure 5. Soil consolidation plays a major role in subsidence of the Van Quan houses.

Although some scientific studies suggest that the groundwater level strongly affects the background subsidence. But to consider specific residential constructions, when the soil background is loaded with the houses under 5 floors with the foundation structure without reinforced concrete piles. This case has shown that the cohesive subsidence factor of the soil is the main cause of the subsidence of the houses.

The underground water observation well in Van Quan urban area is made of Tien Phong plastic pipe with a diameter of 90 mm, a depth of 30 m from the protective steel pipe mouth on the ground, the bottom of the tube is in direct contact with the soil and is not prevented way. Due to insufficient funds, we could not build a deeper groundwater monitoring well, or have a higher standard. This aquifer is at the top of the aquifers, not surface water or affected by surface water. Monitoring data of groundwater level directly at the well did not notice much change

in the period 2005-2009. The fluctuations are mainly recorded during the rainy and dry seasons. Because of the relatively stable groundwater level in Van Quan, it cannot cause the subsidence of residential houses.

For the young sedimentary areas, the consolidation element subsided over time, constructions from three floors should have reinforced concrete foundation piles, constructed by the method of pressing piles. The depth of reinforced concrete piles should exceed the fill and soft soil layers, for Van Quan area is about 15 m depth, based on the geological survey drilling boreholes (Figure 6).

In fact, after 2008, most of the residential houses in VanQuan's new urban area have to reinforce their foundations with piles, increasing construction costs, but ensuring stable and safe houses for a long time. This is also an experience for civil engineering designers in delta areas with weak soil.

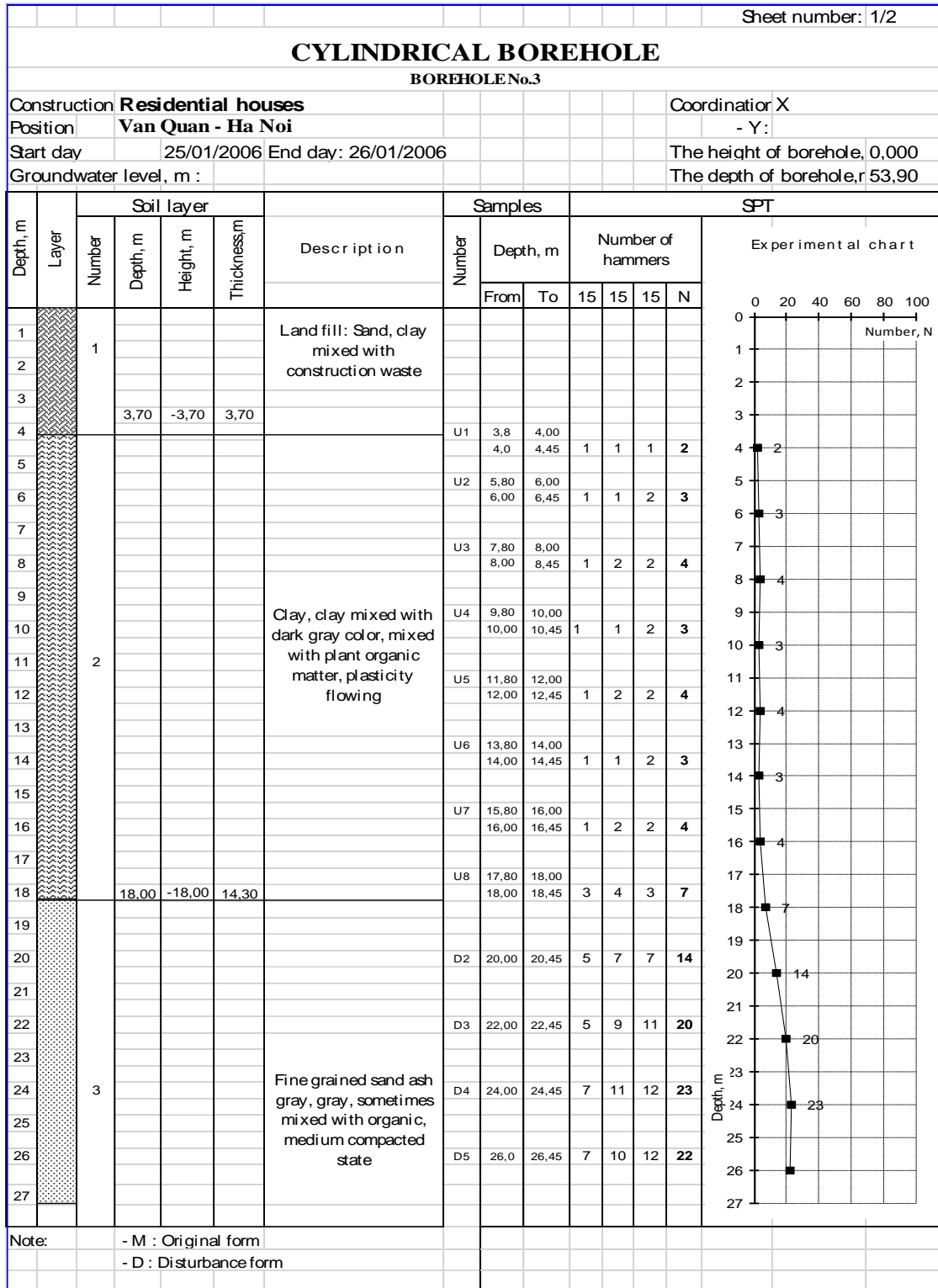


Figure 6. Cylindrical of Borehole No. 3 at the Van Quan residential houses.

5. Conclusions

Regression model is a traditional analytical method to evaluate the impact of independent causes on measured values. Groundwater level and soil consolidation process over time are factors to consider when designing a building. The study showed that the groundwater level in the upper floor fluctuated very small and 98% of subsidence of residential houses in VanQuan's new urban area was due to the weak soil.

This study case is only for residential buildings from 3 to 5 floors with non-reinforced concrete foundation and only consider the top aquifer. For buildings under 3 floors are not covered by this study. Buildings above 5 floors often have foundations made of reinforced concrete piles up to a depth of 20 to 60 meters, so they may be affected by deeper aquifers. More comprehensive studies are needed on this issue to be clear about the impact of groundwater on the subsidence of buildings.

Acknowledgments

The author thanks the support for monitoring data of Van Quan of HUDCIC Consulting Investment and Construction Joint Stock Company. The author also thanks the comments of reviewers who helped improve the content of this article.

References

- [1] T.T. Toan. Forecast of land surface subsidence due to lowering of groundwater level. Final report of the project RD 9505, Hanoi. 1999. (in Vietnamese).
- [2] T.V. Tu, H.N. Anh, D.D. Minh, N.M. Tung, Forecast of ground deformation in Ha Dong area due to urbanization and groundwater extraction. *Journal of Earth Sciences* 35(1)(2013) 29-35. (in Vietnamese).
- [3] T.V. Ty, H.V. Hiep. Current status of groundwater extraction and correlation between water level lowering and land subsidence: Research in Tra Vinh and Can Tho city. *Can Tho University Journal of Science. Topics: Environment and Climate Change* 1 (2017) 128-136. (in Vietnamese).
- [4] D.X. Vinh, N.T. Nhung, N.V. Quang. Determination of Deformation of Construction Using Parametric Modeling-Kalman Filter Application and NonParametric Modeling-Time Series Application. *VNU Journal of Science: Earth and Environmental Sciences* 34(3) (2018) 1-3. <https://doi.org/10.25073/2588-1094/vnuees.4274>. (in Vietnamese).
- [5] P.J. Huber, E.M. Ronchetti. *Robust Statistics*. Second Edition. Published by John Wiley & Sons, Inc. Canada. 1981.
- [6] R.A. Maronna, R.D. Martin, V.J. Yohai. *Robust Statistics: Theory and Methods*, John Wiley & Sons, Ltd. England. 2006.