



Originally published as:

Kühn, M., Schöne, T. (2017): Multivariate regression model from water level and production rate time series for the geothermal reservoir Waiwera (New Zealand). - *Energy Procedia*, 125, pp. 571—579.

DOI: <http://doi.org/10.1016/j.egypro.2017.08.196>



European Geosciences Union General Assembly 2017, EGU
Division Energy, Resources & Environment, ERE

Multivariate regression model from water level and production rate time series for the geothermal reservoir Waiwera (New Zealand)

Michael Kühn^{a,b,*} and Tim Schöne^{b,c}

^aGFZ German Research Centre for Geosciences, Fluid Systems Modelling, Potsdam, Germany

^bUniversity of Potsdam, Earth and Environmental Science, Potsdam, Germany

^cFree University Berlin, Department of Earth Sciences, Berlin, Germany

Abstract

Water management tools are necessary to guarantee the preservation of natural resources while ensuring optimum utilization. Linear regression models are a simple and quick solution for creating prognostic capabilities. Multivariate models show higher precision than univariate models. In the case of Waiwera, implementation of individual production rates is more accurate than applying just the total production rate. A maximum of approximately 1,075 m³/day can be pumped to ensure a water level of at least 0.5 m a.s.l. in the monitoring well. The model should be renewed annually to implement new data and current water level trends to keep the quality.

© 2017 The Authors. Published by Elsevier Ltd.

Peer-review under responsibility of the scientific committee of the European Geosciences Union (EGU) General Assembly 2017 – Division Energy, Resources and the Environment (ERE).

Keywords: geothermal reservoir; water management; data based model; multivariate regression; coefficient of determination; scenario analysis

1. Introduction

The geothermal water reservoir below the village of Waiwera is located about 40 kilometres north of Auckland on the Northern Island of New Zealand (Fig. 1A). Increased water temperatures are observed in an area of approximately 1 km² (Fig. 1B). Since 1863, commercial and private use supplies hotels and spas with water of

* Corresponding author. Tel.: +49 331 288-1594; fax: +49 331 288-1529.

E-mail address: michael.kuehn@gfz-potsdam.de

50 °C. Until the end of the 1960s, the warm water flow was artesian from all wells drilled. Due to overproduction, water needs to be pumped up nowadays and the hot springs on the beach ceased [1]. Consequently, in the early 1980s, the “Auckland Regional Water Board” (today Auckland Council) deployed a water allocation and management plan to enable a sustainable utilization of the resource [2]. Beside others, the management plan demands that the water level in the official and appropriate observation well of the council (no. 74 in Fig. 1C) is 0.5 m above sea level throughout the year in average. This guideline aims to preserve the resource, preventing intrusion of cold ground- and seawater into the reservoir.

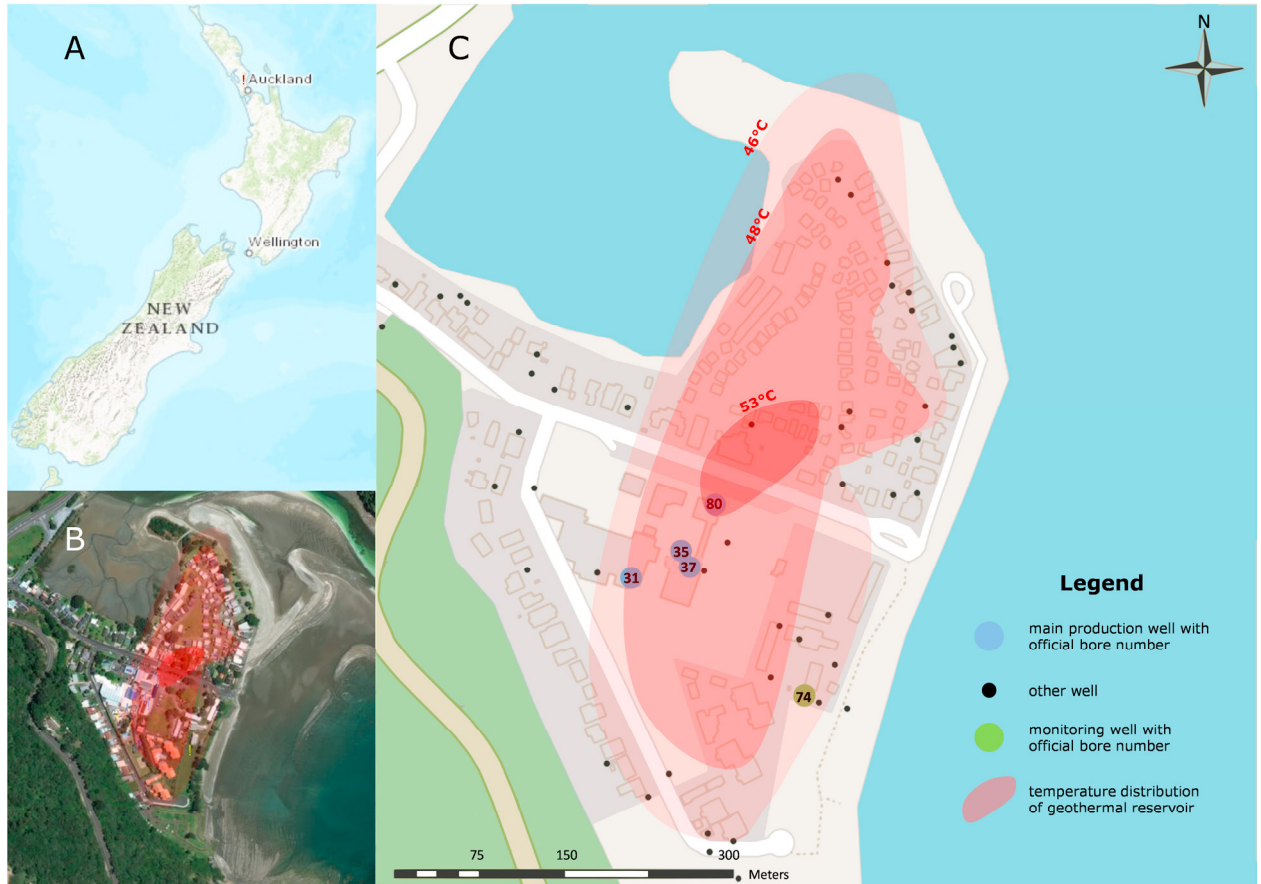


Fig. 1. Location of Waiwera around 40 km north of Auckland on the Northern Island of New Zealand (A). The satellite image shows the village on the south bank of the estuary of the Waiwera river on a flat peninsula (B). The street map indicates locations of the main production wells (blue dots), smaller wells of mainly private users (black dots), the monitoring well (green dot) and the approximate temperature distribution (C).

Water management tools are essential to ensure the conservation of natural resources. They are based on simple mathematical equations [2] or complex, computer based simulation models [1] and are used to determine water quantity and quality or to give insides about underlying processes governing the system. In order to use model outputs for regulation, they need to be scientifically accurate and robust [3]. For a sustainable water management, it is necessary to be able to forecast the water level as a function of the production rates in the production wells. A previous study showed that the best predictions for Waiwera were provided by a multivariate regression model of the water level and production rate time series taking into account the production rates of individual wells [4].

The work presented here is based on data sets of almost three decades (since 1986) of metred water production rates from the Waiwera geothermal reservoir and resulting water levels in the official observation well 74 [5,6]. We

developed and applied data driven regression models for production scenarios to determine the maximum production rate still in compliance with the guidelines of the “Auckland Council” [7].

2. Data based linear regression models as water management tools

2.1. Setting and data basis

The studied site of Waiwera is developed with around 100 wells (Fig. 1C) with depths ranging from around 10 m to 400 m. The geology of the actual reservoir can essentially be described as one massive block of a 400 m thick alternating sequence of folded, fractured and faulted Waitemata sandstones and siltstones. The impermeable base rock underneath is the Waiheke greywacke and the system is confined to the top by unconsolidated alluvial and marine sediments. The geothermal water enters the reservoir through a fault zone from below. However, for data-based models, precise knowledge of the geological background, coupled processes, and specific parameters is not necessary to quantify them. Therefore, the interested reader is referred to previous publications regarding more detailed information about the hydrogeology of the system [1-4].

Following the water allocation and management plan [2], abstraction rates from the geothermal reservoir are metred since 1986. Measurements of the reservoir water level are dating back even longer. In that way, almost three decades of data are available [5,6] and are displayed in Fig. 2.

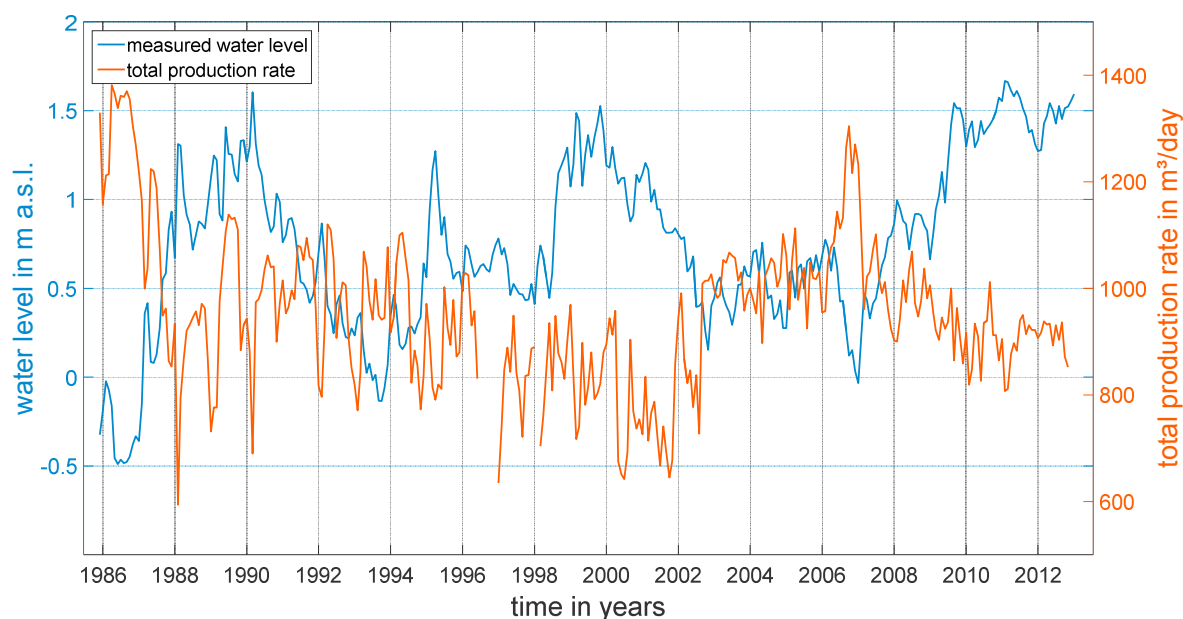


Fig. 2. Total production rates (orange) and resulting water levels (blue) for the period 1986 to 2012 [5,6]. The dashed line at 0.5 m a.s.l. indicates the water level to be reached in average during one year following the reservoir management regulations of the “Auckland Council” [2].

The minimum water level of -1.2 m was observed at the beginning of the 1980s and the maximum of +1.6 m recently. The higher the production rates from the field are, the lower the water level in the observation well is and vice versa. Highest abstraction rates reached almost 1,500 m³/day and lowest were just above 500 m³/day. The metred production rates available for the period 1986 to 2012 [5,6] disclose that most of the water has been pumped from the wells with numbers 31, 35, 37 and 80 (Fig. 1C). These are the main production wells and they are all located at a similar distance of approximately 150 m from the monitoring well number 74 but differ significantly in total drilling depth and length of the installed casing (Table 2).

...Table 1: Characteristics of wells 31, 35, 37 and 80.

parameter / well number	well 31	well 35	well 37	well 80
distance to monitoring well 74 [m]	148	136	123	150
total drilling depth [m]	204	311	107	387
length of installed casing [m]	66	-	38	115

A major change occurred in 1998 when the wells 35 and 37 were closed and 80 newly drilled and taken into operation as major production well. At the same time, usage of well 31 has been reduced significantly. This is the reason why the models, described in the following, were applied either to the time period before or after 1998.

2.2. Data check and handling

Setup of the data-based models was subject to the working order as follows. At first, the availability, quantity, quality and type of data were checked. After that, conceptual considerations led to the decision as to which univariate or multivariate model approaches better suit the existing data. The calibration was then carried out. After validation and assessment of the prognostic capability of the models, it was decided if and which model to apply for scenario calculations [8].

Consistent and complete data sets are the fundament for creating regression models. While the average water levels were available for each day within the period 1986 to 2012, the production rates were not complete. They were missing for the comparably short time span July 1996 to December 1996. Hence, this interval could not be used to develop a regression model.

The data check was done based on the Pearsonian correlation coefficient (PCC) which describes strength and direction of the linear relationship between two variables. In that way, the observation pairs of total production rates and resulting water levels were quantified between -1 (strong negative linear correlation) and +1 (strong positive linear correlation) for each year and used to define the best data basis for creation and calibration of the linear regression models. In most years the data show the expected negative correlation with a PCC between -0.6 to -0.9. However, in the years 2000 to 2002, there is a weakly positive linear correlation between production rates and water level with PCC ranging from 0.2 to 0.4 [7]. Due to that anomaly, the data before 2003 were excluded from the derivation of the regression models in order to obtain a better fit. The implementation of the period 2000 to 2002 would otherwise lead to a significant reduction of accuracy of the models with regard to their prognosis capabilities.

2.3. Derivation of the linear regression models

The linear regression of water level and production data finally was done with the function *LinearModel.fit* (X, y) in the software package MATLAB[®]† to create different models. The matrix X contains the production rates x_i to which a corresponding water level y_i from y is assigned at each time $i = 1, 2, \dots, n$. The function *LinearModel.fit* outputs the regression coefficients, standard errors, t-statistic and p-value as results. These values were used to assess the quality of the regression models.

Regression equations were derived for the time periods before and after 1998 and applied afterwards to recalculate observed water levels. For the quality assessment of the regression models, the underlying data sets were divided into calibration and validation parts. For the model before 1998, the period 1986 to 1993 was used for calibration and 1994 to 1996 for validation. For the models after 1998, the period 2003 to 2010 was used to calibrate the model and the time span 2011 to 2012 for validation.

The quality of the models was determined with r^2 , the coefficient of determination. It is a statistical measure of the accuracy of the linear regression and can take values between 0 and 1. In the first case there is no linear relationship between the data, in the latter a perfect one. The threshold value for r^2 which characterizes a satisfactory

† MATLAB[®] Version R2015b, The MathWorks, Inc.: de.mathworks.com (accessed 27.07.2016)

linear correlation has to be defined and discussed individually for each case. However, a value of 0.5 is the minimum limit for an acceptable measure of determination [9].

Additionally, the Root Mean Squared Error (RMSE) was calculated as measure of the prognosis accuracy of the models. It is the square root of the mean prediction error and is always positive and greater than zero. Because the unit of the RMSE corresponds to that of the underlying data, average deviations of the calculated water levels from the observed ones are obtained.

3. Water level calculations with univariate and multivariate regression models

Since the well constellation in Waiwera significantly changed in 1998, the following water level calculations are presented for the time periods before and after that year. For comparison with the existing results of Chapman [10], we reproduce his successful model for the earlier time frame first. Afterwards, we investigate regression models newly developed for the time period until recently. Finally, production scenarios are studied to determine the maximum abstraction rate of geothermal water from the reservoir in agreement with the management regulations.

3.1. Time period before 1998 with old well constellation

The need for a hydrogeological model to forecast the water level of the Waiwera reservoir became immanent with the thermal groundwater allocation and management plan [2]. Within his master thesis of 1998, Chapman [10] created a multivariate regression model. He considered specific production rates from separate wells as independent variables. In this way, positions and well depths are indirectly taken into account, because from a hydrogeological point of view, each well influences the water level in the monitoring well 74 differently. Exactly as in the original model, the data from 1986 to 1993 are used for calibration and from 1994 to 1996 for validation. As independent variables, the production rates of the major production wells to that time, number 31, 35 and 37 (Fig. 1C) are included in the regression to determine the dependent variable, the water level in the monitoring well 74 (Fig. 1C and Fig. 2).

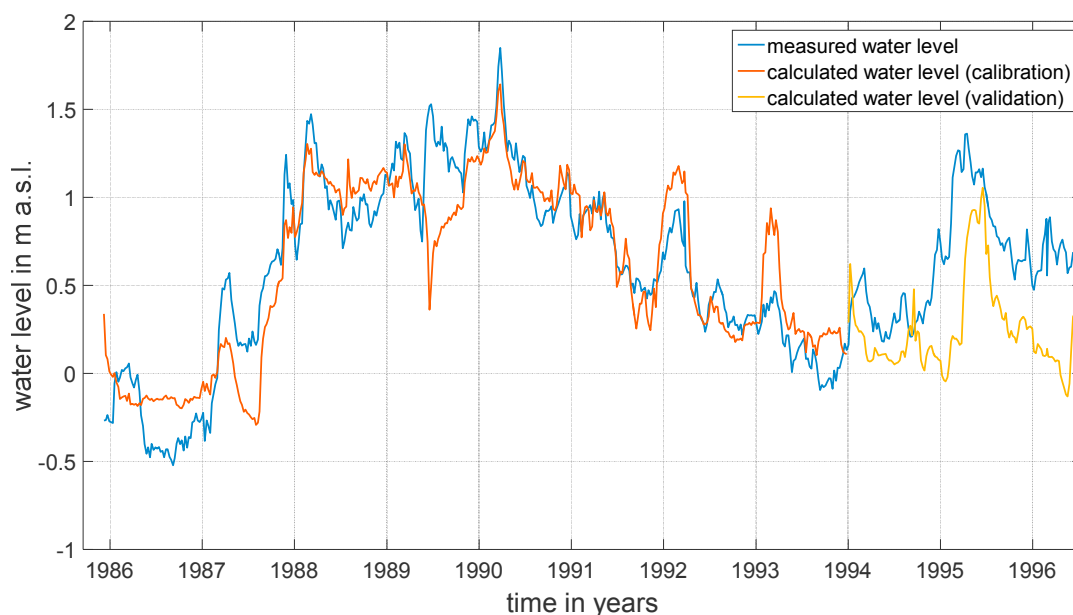


Fig. 3. Water level measured in well 74 and calculated values following the multivariate regression model of Chapman [2] with production wells 31, 35 and 37. The time period 1986 to 1993 was used for calibration and 1994 to 1996 for validation. The coefficient of determination is $r^2 = 0.72$, the Root Mean Squared Error is $RMSE = 0.29$ m.

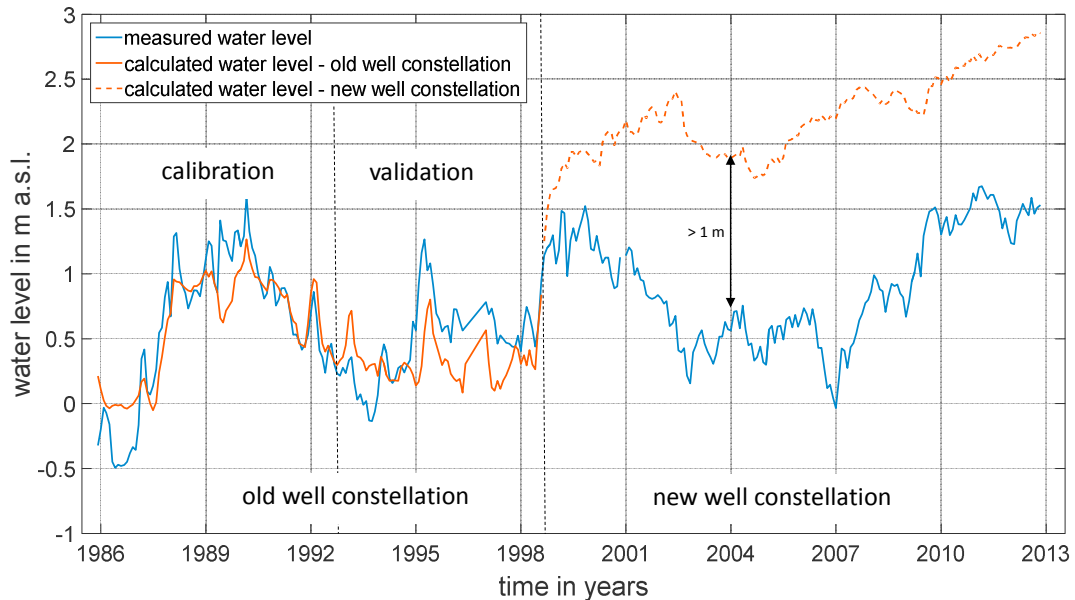


Fig. 4. Water level measured in well 74 and calculated values following the multivariate regression model of Chapman [2] for the old and new well constellation before and after 1998, respectively. The model fails with deviations of more than 1 m between measurements and calculations.

Fig. 3 shows the results of water level calculations with the multivariate regression model compared to the measurements. The coefficient of determination in this case is $r^2 = 0.72$ and the Root Mean Squared Error is $RMSE = 0.29$ m. From Fig. 4 it gets obvious that the same model applied to the time period after 1998 fails because of the significantly changed well constellation at Waiwera.

Time period after 1998 with new well constellation

The previous chapter and former publications [3,4] outline that data-based multivariate regression models might be applicable to manage the geothermal reservoir of Waiwera. However, as shown in Fig. 4, due to the changed well constellation, a new model has to be derived. As of August 1998, there are still two large, active production wells: the old well 31 operated with a lower production rate and the newly drilled well 80, which accounts for approximately 70% of the total production from the Waiwera geothermal field. The data check revealed that the period 1998 to 2002 does not provide suitable values for derivation of a regression model [7]. This is why the production rates and water levels from 2003 to 2010 are used for calibration. Fig. 5 displays the calculated water levels in comparison to the measured ones. The years 2011 to 2012 served as validation period. The coefficient of determination for the new multivariate regression model is 0.75 and the averaged deviation between calculated and observed water levels expressed by the RMSE is 0.19 m.

The aim is to provide the “Auckland Council” with a simple and fast-to-use model. Therefore, we additionally tested the quality of a univariate regression model. A model was derived from the same data basis for the period 2003 to 2010 with only one independent variable for describing the water level. An earlier version of a univariate regression model, just dependent on the total production rate of all wells together, was one of the first models to be built for the geothermal reservoir Waiwera [2,3,4]. Although this model represents long- and short-term trends as well (Fig. 5), it turns out to be of minor quality with a coefficient of determination of 0.58 and a RMSE of 0.43 m. For this model, the years 2011 to 2012 served as validation period, too.

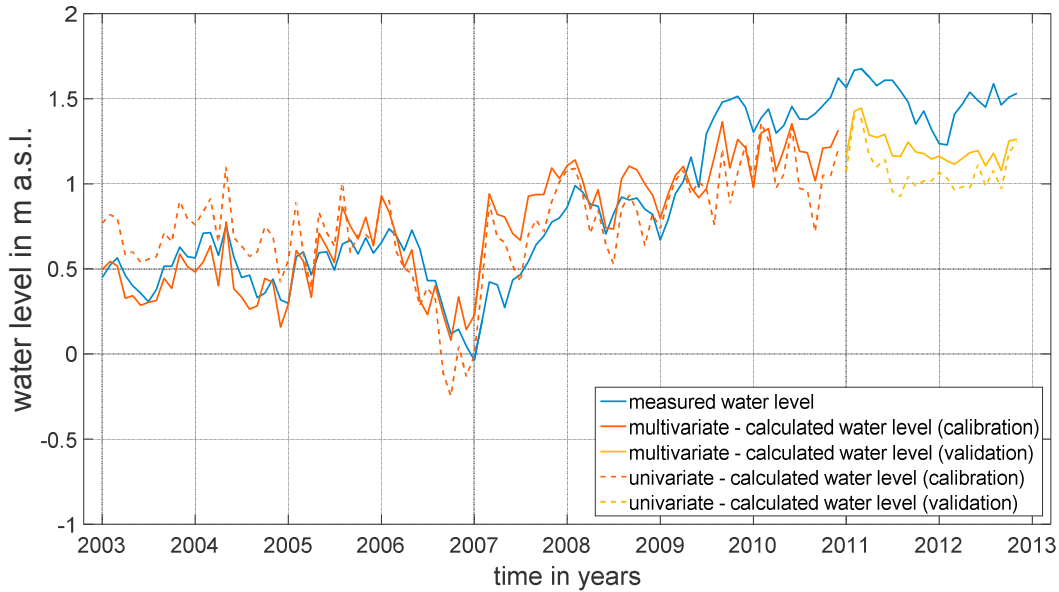


Fig. 5. Water level measured in well 74 and calculated values for a univariate (dashed) and a multivariate (solid) regression model based on either the total abstraction rate from the reservoir or separate from production wells 31 and 80. The time period 2003 to 2010 was used for calibration and 2011 to 2012 for validation. The coefficients of determination r^2 are 0.58 and 0.75, the RMSE are 0.43 (m) and 0.19 (m) for the univariate and multivariate model, respectively.

3.2. Production scenarios to determine the maximum abstraction rate

Purpose of any water management tool is to give accurate and precise prognoses of the water level to be measured in the monitoring well 74 (Fig. 1C) dependent on the production scheme to be applied for the geothermal field. Based on the presented results and previous studies [3,4], it is recommended to employ a multivariate regression model, if the mere question is the resulting water level in the reservoir from a certain production scenario without any change of the well constellation itself.

With the recommended model, different production scenarios were tested to determine the maximum abstraction rate in compliance with the water management regulations of the “Auckland Council”. On the one hand, Table 2 lists the resulting water level in the monitoring well 74 for the long-term average abstraction rate of 1,000 m³/day from either well 31 or well 80. On the other hand, it lists the determined production rates for well 31 and well 80 to meet exactly the regulation limit of 0.5 m a.s.l. in monitoring well 74.

Table 2. Production scenarios based on the recommended multivariate regression model.

production rates in m ³ /day			water level in m a.s.l.
well 31	well 80	total	monitoring well 74
1,000	0	1,000	0.2
0	1,000	1,000	0.7
0	1,075	1,075	0.5
895	0	895	0.5

From the results it is obvious that the shallower production well 31 (total depth 200 m, casing depth 66 m) has a larger impact on the water level in monitoring well 74 compared to the deeper production well 80 (total depth 385 m, casing depth 115 m). The same production rate of 1,000 m³/day either pumped entirely from well 31 or well 80 results in water levels of 0.2 m a.s.l. and 0.7 m a.s.l., respectively. Determination of the maximum production rate in compliance with the regulation limit of 0.5 m a.s.l. in monitoring well 74 reveals 895 m³/day for well 31 and 1,075 m³/day for well 80.

These results lead to the question, what the difference is from the hydrogeological point of view, between positions and depths of both production wells in regard to the monitoring well. Both do have the same distance of around 150 m. As mentioned, total depth and casing depth differ significantly. However, it could have been expected that well 80, which is of larger depth and has the deeper casing, shows a larger impact on the monitoring well 74 because this is a deep well itself (total depth 400 m, casing depth 150 m). Because this is not the case, the geological structure seems to conceal the reason behind it.

4. Conclusions

Linear regression models based on the abstraction rates from the main production wells and the resulting water level in the monitoring well provide a feasible way to derive a water management tool for the Waiwera geothermal reservoir with prognostic capabilities [3,4]. Our modelling results underline that the multivariate regression is better suited than the univariate model because it is more accurate predicting the water level. The integration of individual production wells instead of the total production rate into the regression model is a simple opportunity in order to improve the model quality. In this context, it is important for the data handling to check at first the correlation between production rates and water level using the Pearsonian coefficient. Such an assessment is necessary especially after changes in the general well constellation as well as large changes in the production rates even from existing wells. If the correlation declines significantly, a new model should be created. It is recommended that the multivariate regression model is renewed annually to take into account more recent fluctuations and trends in the water level and to always allow precise predictions.

The investigation of production scenarios based on the multivariate regression model revealed that a maximum of approximately 1,075 m³/day can be pumped to ensure a water level of at least 0.5 m a.s.l. in the monitoring well if only production well 80 would be used. Further, the results show that the current limit of use with an allocation of 1,500 m³/day [2] geothermal water is too high. If 1,500 m³/day would be produced from the reservoir the water level would probably decline to -0.5 m a.s.l. [7]. The water level reacts very sensitively to changes in the water consumption.

Data-based models are a simple and precise way to create water management tools. However, they disregard entirely the geological background and hydrogeological situation. Therefore, to quantify the impact of different potential locations or drilling depths of new wells on the water level in the geothermal reservoir of Waiwera in advance, process based models are required as well. To set up those models [1] detailed knowledge of the subsurface and of the processes involved is necessary. Beside the wealth of existing data from various surveys [2] further field work and monitoring is recommended.

Acknowledgements

The authors would like to thank the Auckland Council for provision of information and data.

References

- [1] Kühn, Michael, and Heinke Stöfen H. (2005) "A reactive flow model of the geothermal reservoir Waiwera, New Zealand." *Hydrogeology Journal* 13 (2005): 606-626, doi: 10.1007/s10040-004-0377-6.
- [2] Auckland Regional Water Board. (1987) "Waiwera thermal groundwater allocation and management plan 1986." *Auckland Regional Water Board*, Technical Publication No. 39, Auckland, New Zealand.
- [3] Kühn, Michael, and Charlotte Altmannsberger. (2016) „Assessment of data driven and process based water management tools for the geothermal reservoir Waiwera (New Zealand).” *Energy Procedia* 97 (2016): 403-410, doi: 10.1016/j.egypro.2016.10.034.

- [4] Kühn, Michael, Charlotte Altmannsberger, and Carmen Hens. (2016) “Waiwera’s warm water reservoir – how significant are models?” *Grundwasser* 21 (2016): 107-117, doi: 10.1007/s00767-016-0323-2 (in German).
- [5] Rose, JL, and G Zemansky. (2013) “Updated Hydrogeological Evaluation of the Waiwera Geothermal Aquifer.” GNS Scientific Consultancy Report 2013/67, *GNS Institute of Geological & Nuclear Science*, Wellington, New Zealand (unpublished).
- [6] Zemansky, G. (2005) “Hydrogeological Evaluation of the Waiwera Geothermal Aquifer.” Client Report 2005/131, *GNS Institute of Geological & Nuclear Science*, Wellington, New Zealand.
- [7] Schöne, Tim. (2016) “Hydrogeological analysis of water level and production data time series for the creation of a water management tool for the geothermal reservoir Waiwera (New Zealand).” B.Sc. Thesis, University of Potsdam, Earth and Environmental Sciences (in German, unpublished).
- [8] Araghinejad, Shahab. (2013) “Data Driven Modeling: Using MATLAB in Water Resources and Environmental Engineering.” Water Science and Technology Library, Teheran, Iran, *Springer Publishers*.
- [9] Moriasi, D, J Arnold, M Van Liew, R Bingner, R Harmel, and T Veith. (2007) “Model evaluation guidelines for systematic quantification of accuracy in watershed simulations.” *American Society of Agricultural and Biological Engineers ASABE*.
- [10] Chapman, MG. (1998) “Investigation of the dynamics of the Waiwera geothermal groundwater system, New Zealand.” Master Thesis, The University of Waikato, Hamilton, New Zealand (unpublished).