DRILLING MORE WELLS OR DOING PUMPING TEST:
INVESTIGATING THE RELATIVE VALUE OF WATER HEAD
AND CONDUCTIVITY MEASUREMENTS IN REDUCING
INVERSE GROUNDWATER MODELING ERROR

A Thesis in
Civil Engineering

by

Nuan Sun

© 2016 Nuan Sun

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2016
The thesis of Nuan Sun was reviewed and approved* by the following:

**Chaopeng Shen**  
Assistant Professor of Civil Engineering-Water Resource  
Thesis Advisor

**Christopher J. Duffy**  
Professor of Civil Engineering-Water Resource

**Xiaofeng Liu**  
Assistant Professor of Civil Engineering-Water Resource

**Patrick J. Fox**  
Harriette K. Shaw Professor  
Head of the Department of Civil and Environmental Engineering

*Signatures are on file in the Graduate School
ABSTRACT

Given limited resources, an investigator interested in inversely estimating the groundwater conductivity field can either invest in hydraulic head (H) measurements (e.g., drill a well) or conductivity (K) estimates (e.g., conduct pumping tests). They result in different information content and cost. While advanced stochastic methods exist to estimate the worth of data, there is no first-order answer that enables fast decision making. Here we empirically determine the worth of H or K data, in terms of reducing calibrated K error, as a function of data density, recharge rates, and boundary conditions. We found that normalized K error can be well approximated by a smooth function of heterogeneity-normalized H and K data densities. Stepwise polynomial fitting suggests including only the first-order terms and a mild interaction term. The relative information content ratio of H to K for the next data point ($R_n$) suggests the worth of K is always more than that of H, and could be as higher than 10 when K data is sparse. $R_n$ is a function of data densities and K correlation length, thus the decision must consider the amount of presently available data and the heterogeneity of the field. We suggest investment decisions can be charted by following a straight line on which the information content per unit cost ratio is separated to be less- and greater-than-1 values.
# Table of Contents

List of Figures .................................................................................................................................................. vi
List of Tables ....................................................................................................................................................... viii
Acknowledgements ................................................................................................................................................ ix
Chapter 1 Introduction ......................................................................................................................................... 1
  1.1 Study Motivation ........................................................................................................................................ 1
  1.2 Overview on groundwater model calibration .............................................................................................. 4
  1.3 Overview on study of data worth to groundwater modeling ........................................................................ 7
  1.4 Research objectives .................................................................................................................................. 10
  1.5 Terminology ............................................................................................................................................... 11
Chapter 2 Methodology ...................................................................................................................................... 13
  2.1 Introduction ................................................................................................................................................ 13
  2.2 Methodology of parameter estimation in MODFLOW and PEST ............................................................. 14
    2.2.1 Conceptual model identification ........................................................................................................ 14
    2.2.2 Groundwater flow process ............................................................................................................... 15
    2.2.3 Parameter estimation ......................................................................................................................... 17
    2.2.4 Implementation of Gauss-Marquardt-Levenberg approach ............................................................. 20
    2.2.5 Regularization .................................................................................................................................. 21
  2.3 Process model development ....................................................................................................................... 23
    2.3.1 Synthetic model construction ........................................................................................................ 23
    2.3.2 Observation abstraction and parameterization ................................................................................. 25
    2.3.3 Termination criteria of inverse modeling iterations ........................................................................ 26
  2.4 Dimensionless control variables for data densities and calibration errors ................................................. 27
  2.5 Calibration errors to different recharge and boundary condition ............................................................ 31
2.6 Experimental design and multi-variable polynomial curve fitting ..................33
2.7 Relative data worth and economic analysis ..................................................34
Chapter 3 Result and Discussion .........................................................................36
  3.1 Verification of the dimensionless numbers .........................................................36
  3.2 Impact of recharge and boundary condition on model calibration errors ..........37
  3.3 Errors as a function of effective data densities ..................................................39
  3.4 Multi-variable polynomial curve fitting ............................................................41
  3.5 Economic worth of data acquisition against model calibration error reduction 46
Chapter 4 Conclusions and Limitations ..................................................................51
Appendix A: Random log(K) Fields and Evaluations with Variograms .................53
Appendix B: MATLAB Function to Plot $R_n$ and $R'_n$ ........................................56
References .............................................................................................................58
List of Figures

Figure 1. Water-level declines. Red regions indicate areas in excess of 500 square miles that have water-level in excess of 40 feet in at least one confined aquifer since predevelopment, or in excess of 25 feet of decline in unconfined aquifers since predevelopment. Blue dots are wells in the USGS National Water Information System database where the measured water-level difference over time is equal to or greater than 40 feet (Reilly et al. 2008). ................................................................. 2

Figure 2. (a) Sampling cost, remediation cost total cost, and number of remaining plume realizations versus number of samples collected in example sampling program for preselected plume. (b) Sensitivity of average optimum number of samples, prior remediation cost, average optimum remediation cost, and average optimum total cost to variance of log hydraulic conductivity (James and Gorelick, 1994). .................. 8

Figure 3. Discretization scheme used by MODLFOW to numerically solve groundwater process equation (Harbaugh et al. 2000). .............................................. 16

Figure 4. Locations of observation points (black) and pilot points (red). Boundaries of cells constituting domain grid are shown with black lines. ............................................ 25

Figure 5. Goodness of match between model computed and observed water head values at 256 hypothetical observation wells.............................................................. 31

Figure 6. Boundary condition transaction from initial (left) to the comparison (right). Red cells represent those of specific head and gray color is for no-flow boundaries. Blue means the cell is activated in numerical simulation and the black exhibits the procedure of transaction........................................................................................................ 32

Figure 7. Dimensional verification of effective data density of both H and K to normalized conductivity errors. .......................................................... 37
Figure 8. Normalized water head error (a) and conductivity error (b) changing with recharge values under two different water head data amount.................................................38

Figure 9. Normalized water head error (a) and conductivity error (b) changing with recharge values under various amount of data collection pertaining to both water head and conductivity. BC1 means domain with one specific water head boundary and BC2 stands for the domain with two specific water head boundaries ..............................................39

Figure 10. Normalized K error (a) and H error (b) as well as absolute K error (c) and absolute H error (d) relative to effective data densities.................................................................41

Figure 11. 3D visualization of adjusted polynomial curve fitting of effective data densities to normalized K (a) and H (b) errors with original scatters of ($\mu_H$, $\mu_K$), which is represented by black spots, used in the fitting process...........................................43

Figure 12. Contour representation of adjusted polynomial curve fitting results of normalized water head error (a) and normalized conductivity error (b) changing with data densities..................................................................................................................44

Figure 13. $R_\mu$ of K error (a) and H error (b) with effective data densities $\mu_H$ and $\mu_K$. Bold lines highlight the less-than-one values among the contour.........................46

Figure 14. $R_\mu$ when $\mu_H$ change with K data measurements (a, c) and $R'_\mu$ when $\mu_H$ is independent (b, d) based on two different domains regarding $\lambda$. .........................48

Figure 15. $R_c$ based on assumption that field investigation of K is often associated with an additional H data (Scenario A) within two different domains regarding $\lambda$. Bold lines highlight less-than-one values among the contour. .................................................50
List of Tables

Table 1 Values of calibration termination control factors ..............................................27

Table 2 Ranges of hydraulic water head with each of the recharges under two H data densities regarding $\mu_H$ ......................................................................................................................................33

Table 3 Multi-polynomial curve fitting result and P-values of T-test ..............................42
Acknowledgements

I would like to express my faithful gratitude to Dr. Chaopeng Shen for his professional instruction and supervision on this interesting research project, which is of great educational value to me. His support not only helped me to gain an excellent experience during the last two years of academic career, but will positively enlighten my future development as well. I also deeply appreciate the other two committee members, Dr. Xiaofeng Liu and Dr. Christopher Duffy for their distinguished professionalism in the field of my study. My work has been dramatically promoted and improved according to their suggestions. Moreover, a great thank you goes to the two PhD candidates, Kuai Fang and Xinye Ji, who have given me significant help throughout the endeavor. Finally, I would like to thank my mother and father for their unconditional love and continuous support during my life, this academic pursuit would not be possible without their efforts and spiritual encouragements.
Chapter 1 Introduction

1.1 Study Motivation

As a vital component of the global hydrological cycle, groundwater plays a crucial role in maintaining life and developing human society. Excluding the water locked in polar ice, about 89% of fresh water is stored in subsurface, and more than 1.5 billion people are living on groundwater supply, globally (Alley, 2002; Koundouri, 2004).

Groundwater resource is being increasingly used in agriculture, industry and commercial facilities as a subsequence of the rapid blossom of pumping technologies (Konikow and Kendy, 2005). During the year of 2000, a rate of worldwide groundwater abstraction of 734 (±82) km$^3$ a$^{-1}$ was estimated (Wada et al. 2010). While concentrating on the United States, the USGS conducted a survey to nation-wide water use in 2010: about 355 billion gallons of water in total are customized per day, 22.33% of which are from fresh and saline groundwater resource (Maupin et al., 2010). McGuire et al. (2000) reported a typical example of High Plain where groundwater storage has reduced by 6% of predevelopment cumulation during the 20$^{th}$ century. Moreover, Wada et al. (2010) drawn the conclusion that groundwater depletion has been increased to 283 (±40) km$^3$ a$^{-1}$ in 2000 from 126 (±32) km$^3$ a$^{-1}$ in 1960 based on a numerical model.

Excessive withdraw of groundwater can result in extensive damages like degradation in water quality, ground subsidence, higher cost to maintain wells yields and devastating impact to relevant ecosystem etc. (Konikow and Kendy, 2005; Wada et al., 2010), therefore, the management of groundwater has attracted much concern of people during the past decades; as depletion has extended dramatically all around the world (Alley, 2006). However, estimation of large-scale groundwater condition is an issue with challenges due to complexities of groundwater property, various investigation system with different spatiotemporal scales and personal research objectives (Reilly et al., 2008). Furthermore, there is no such regular arrangement to integrate groundwater
information on the federal level, hence the researchers mostly rely on the local resources. Despite the varying of temporal and spatial scales inherent in the defective condition above and comparing materials available all across the country, still, people have drawn a sketch picture (Figure 1) of present national groundwater condition regarding water level declines and pointed out further direction to better assessment and management of groundwater resource (Dennehy et al., 2015; Reilly et al., 2008).

As a response to the explosion of groundwater withdraws in the last half century, methods of studying groundwater have made significant progress. Based on field investigations, USGS has established a database of groundwater-related data throughout 50 states for more than 100 years, This database is the primary resource for nationwide groundwater managers and policymakers (USGS FS-058-95). With the basis of geophysical principles, electronic and magnetic methods are commonly used to study aquifer structure and subsurface flow regarding containments (Karlik and Kaya, 2001; Wattanasen and Elming, 2008).

Figure 1. Water-level declines. Red regions indicate areas in excess of 500 square miles that have water-level in excess of 40 feet in at least one confined aquifer since predevelopment, or in excess of 25 feet of decline in unconfined aquifers since predevelopment. Blue dots are wells in the USGS National Water Information System database where the measured water-level difference over time is equal to or greater than 40 feet (Reilly et al. 2008).
During the last several decades, computer-based numerical models have been increasingly used in groundwater studies. As stated by Gallagher and Doherty (2007), “computer models are often used to predict the behavior of environmental systems”. Computer-based simulation and modeling have been demonstrated to be effective approaches to solving groundwater management problems. Like most other hydrological models, groundwater models are typically operated both conceptually and numerically. However, conceptual models are highly simplified against realistic systems, in addition that the phase of parameterization is hardly determined to sufficiently accounts the heterogeneity of domain aquifer. Mathematical algorithm solving the governing equations may not accurately infer the values of parameter either. All these matters can subsequently introduce errors to the models thus weakening the confidence with which one can rely on them (Christensen and Doherty, 2008; Muleta and Nicklow, 2005; Tonkin and Doherty, 2009). In addition to such a discrepancy between model outputs against historical field records, uncertainty is another intrinsic feature of models. Quantitative knowledge of uncertainty in a spatial perspective can help modelers assess the areas where the existing models are more to deliver inaccurate results.

Error and uncertainty pertaining to both parameter and prediction are commonly regarded as essential criteria when evaluating the performance of a computer model. Data collection is the foundation of a model with deterministic impact on error and uncertainty quantification. Incomplete or inaccurate data can cause large misfit between model outputs and data records. Fallacious locations where data being collected can leave areas with uncertainty to be uninformed where predictions will be biased. On the other hand, it is never practicable to ceaselessly investigate the site with limited resource. Overly considered data collection can needlessly waste millions of dollars given the fact that a single water sample for contamination test may require thousands of dollars (James and Gorelick, 1994). In one word, tradeoff decision must
be made to improve the model performance by appropriately adding data information and to ensure such investigation are financially feasible.

The earliest research on (additional) data collection worth can be ascended to 1970s (Davis et al., 1972; Gates and Kisiel, 1974; James and Gorelick, 1994) followed by remarkable amount of others’ work generally focusing on either numerically exposing the uncertainty associated in model itself or relating uncertainty analysis to groundwater management problems. The latter always involves monetary consideration of measurement costs with various site investigation scenarios. The conclusion of cost-effective data collection scheme is drawn with respect to the reliability of the model, which is determined by uncertainty in model calibration and prediction. However, previous work in literature rarely considered model calibration errors in terms of not only measurements, but parameters as well. It lacks an integrated economic concept relating data collection cost to both error and uncertainty quantification. Moreover, early work generally makes determinations of optimal measurement scenario among various alternatives within the aquifer under study. On the other hand, studies such as how well particular field investigation strategy can serve simulations of various domains can further complete people understanding on data sufficiency.

1.2 Overview on groundwater model calibration

In the following section, we will briefly introduce some previous work on groundwater modeling as well as associated error and uncertainty analysis. Thereafter we will display the current stage of research on data worth in improving the modeling quality and simultaneously to control data measuring budget. Before continuing, one is strongly recommended to glance over the “Terminology” at the end of present chapter in order to get familiarized with the terms that are particularly used in groundwater modeling research.
A number of computer-based models have been released and demonstrated to be “satisfying” in representing realistic aquifer system. One of the most popular applications of modeling is to simulate subsurface flow with chemical contamination. For instance, GFLOW (Yager and Neville, 2002) and STANMOD (Feinstein and Guo, 2004) uses analytical methods to simulate subsurface flow; Model Viewer (Zhen, 2004) and Environmental Insite (Tonkin and Becker, 2005) visualize the model-and-user interface; other tools like GIS and Microsoft Excel are representatively used in ModTech (Pint and Li, 2006) and Jiao and Leung's work (2003) as supplement to provide better match with modeler’s needs, just to name a few. Another computer program suite, MODFLOW (firstly named as Modular Model), was originally developed by U.S. Geologic Survey (USGS) in the 1980s together with the first documentation of McDonald and Harbaugh (1983) and has become one of the most widely used groundwater flow models since 1990s. MODFLOW-2000 (Harbaugh et al., 2000) was then developed to integrate groundwater flow transport and parameter estimation to transition the program from package-based to process-based (Mcdonald and Harbaugh, 2003). The applications of MODFLOW (-2000) during the last two decades are rarely exclusive, but always associated with other programs like PEST, specific cases will be introduced in later reviews.

As stated earlier, data source is the basis of a computer model and can crucially but also partially determine model error and uncertainty. Often, parameters constituting aquifer properties cannot be readily measured, let alone an extensive relevant network. This is primarily due to the vast financial and technical requirements of such measurements. Inverse modeling is a more advanced approach to curve fitting of observations that is used by van Genuchten et al. (1991) under the assistance of nonlinear least-squares optimization scheme to estimate the parameters all over the domain (Vrugt et al, 2001). Abundant work has been done to expose and reduce error and uncertainty within parameter estimation process.
Spatial characterization of hydrogeological condition within the model domain can provide more prior information to calibration especially when using zones, each of which contains uniform hydraulic property and parameter value, to parameterize the model. Zonation is implemented before actual parameter estimation process based on modeler’s knowledge on the area of study, which can hardly embrace the complexity of study field at adequate or even just satisfactory level (Doherty 2003a). Pilot point method has been widely used as an effective alternative to parameterize the model with capability to generate a smooth variation of the parameter while also reflect the heterogeneity of studied domain (Doherty, 2003; Lavenue and de Marsily, 2001; Ramarao et al, 1995).

As attempts are being made to sufficiently reflect the spatial heterogeneity of some aquifer properties like hydraulic conductivity, one can rapidly increase the number of pilot points as parameters. Such a matter can lead to substantial non-uniqueness in parameter estimation, where a variety of parameter sets can fit the model with similar model-to-measurement discrepancies, especially when the number of parameters exceeds the number of observations. Regularization methodology has occupied a commonplace in treating such a drawback of pilot point by introducing “modeler preference” of parameter value distribution in inverse modeling (Doherty, 2003; Tonkin and Doherty, 2005). Its advanced variant, namely, adaptive regularization is demonstrated to be more efficient in stabilizing parameter values in (Doherty and Skahill, 2006).

Among literature model calibration has also been implemented using tools other than PEST. For example, to study optimal data network, Thompson et al. (2013) integrate an established surface and groundwater model with a subsurface model to create a unified MODFLOW-SUEFACT model, which is later to be calibrated with some data that is not always emphasized by modelers. Another method named, Meta-Heuristic, is used to calibrate groundwater models. It is demonstrated to be valid with the illustration of a
case study on Ghaen plain, Iran by criticizing sum of squared deviation (SSD) and the absolute value of deviation (SAN) between simulated and observed hydraulic head values (Haddad et al., 2013).

1.3 Overview on study of data worth to groundwater modeling

Previous work has been done seeking optimal measurement strategy to serve groundwater management decisions and mostly focus on subsurface environment remediation. James and Gorelick (1994) studied the cost-effectiveness of sampling based on Bayesian data worth framework to solve a hypothetical contamination problem where the pollutant plume is uncertain in both location and extent. The cost-effectiveness here is defined as the cost of potential remediation program reduced by data information is greater than that to obtain them. Under multiple assumptions, say, two-dimensional steady state, the number of 6 is deemed to be optimum for the studied case, such the number is also particularly sensitive to the variance of hydraulic conductivity of studied aquifer (see Figure 2). Similar work can be found in (Freeze, et al, 1992), Risk-cost-benefit objective function (Freeze et al., 1990) is used to facilitate tradeoff between the cost of data collection and expected value of risk reduction brought by the data. A synthetic landfill leachate issue is illustrated, where the risk depends on the uncertainty associated with aquitard continuity as well as hydraulic conductivity. Adjusting the objective to Bayesian theory, data worth pertaining to reducing the uncertainty of both aspects above is calculated, which is thus the tolerant cost to gain a data sample.

Data collection strategy has also been studied under other various approaches help to solve groundwater problems. Tucciarelli and Pinder (1991) employed a chance-constrained stochastic technique to find best number and locations of additional measurements that can result in minimum summation cost of data acquisition and groundwater reclamation. One similar chance-constrained model is
Figure 2. (a) Sampling cost, remediation cost total cost, and number of remaining plume realizations versus number of samples collected in example sampling program for preselected plume. (b) Sensitivity of average optimum number of samples, prior remediation cost, average optimum remediation cost, and average optimum total cost to variance of log hydraulic conductivity (James and Gorelick, 1994).

coupled with an integer-programing sampling network design model in Wagner (1999) to optimize pumping and sampling strategy. In a synthetic aquifer remedial case design, the former model identifies least cost of possible pumping plan while the latter model identifies optimal sampling network in reducing simulation model uncertainty.
What is worth raising in Wagner’s work is that he compared the contribution of multiple data types to reduce model uncertainty, which will thus save groundwater remediation cost. Conclusively, the worth of measuring hydraulic conductivity is much greater than that of aquifer state investigation (i.e. to measure hydraulic head and pollutant concentration).

The hydraulic conductivity (K) of groundwater aquifers is an important parameter for either contaminate fate and transport studies or evaluations of groundwater resources (Freeze and Cherry 1979). While large-scale estimates of groundwater conductivities become increasingly available, e.g., (Gleeson et al. 2014), there are still many districts in the world, e.g., California desert in the US, where high-quality K data is too scarce for groundwater modeling purposes. K can be inferred from pumping tests (Marsily 1986) or lithological estimates (Hördt et al. 2007; Hyndman and Gorelick 1996). Typically, K values are not randomly distributed spatially, but are auto-correlated in space (Rehfeldt et al. 1992). Therefore, one K measurement not only informs us about the K values at the site, but also gives us some information about the adjacent region. On the other hand, K can also be estimated inversely through calibrating a suitable groundwater flow model to observed hydraulic head (H) using estimated recharge. Often they make use of available K data points and geostatistical model to constrain the inversion process. This inversion is possible because, given an accurate estimate of recharge, H carries information about K.

More efforts can be found in literature studying data worth on groundwater model development. For example, Vrugt et al (2001) developed a Parameter Identification Method based on Localization of Information (PIMLI) to serve uniqueness of parameter estimation. With a synthetic model where there are artificially known values of parameters, a conclusive figure is obtained where specific subsets of data are separated out containing the most data information on one of the parameters, which means, each value of parameters is sensitive most to one of the data subsets.
The worth of different types of data to characterize aquifer heterogeneity and to reduce uncertainty in parameter estimation is focused by Fu and Jaime Gómez-Hernández (2009) with a new blocking Markov chain Monte Carlo algorithm. The conclusion was outlined as log-conductivity data can reduce the spatial distribution uncertainty of conductivity more than that of the piezometric head, whereas measurement of water head works reversely. And the combination of both data types can generally act better in model uncertainty reduction. To name a few of other works that focus on data worth to groundwater model calibration, say, Christiansen et al. (2011); Thompson et al. (2013); Lubczynski and Gurwin (2005); Bakr and Butler (2004).

1.4 Research objectives

Either H and K data helps reduce the uncertainty about the K field, i.e., they both have mutual information (Cover and Thomas 1991) with the K field. However, they vary in the information content they carry and their cost. Field-collected data can be very expensive, and the cost scenarios can vary greatly from place to place and from time to time. Traditionally, H data points can be read from existing wells or new wells. A K data point, if estimated using pumping test, requires more efforts and time. Therefore, we assume that a K data point costs more than an H data point. More recently, data collection effort can take the form of securing and interpreting existing well records and pumping records. Even in this scenario, a K data point can take more effort to obtain, and it indeed tends to be sparser. Therefore, given limited resources, which could be time and budget, an investigator needs to decide on the type of data is the most worthwhile to pursue. Thus, from a practical point of view, there is a need for understanding the relative worth of H and K data, condensed into the question “should we obtain an H or K data point next?”. However, searching through literature we found that there are no clear, simple answers to this inquiry.

The goal of this paper specifically includes:
i) To produce the first-order estimate of the relative worth of H and K data points on calibrated K error reduction, and to identify its main controlling factors.

ii) To examine if calibrated K error can be described by a function of recharge, boundary conditions, and relative data densities of H and K.

iii) To seek methods reducing the dimensionality of the problem by non-dimensionalization, and verify the effectiveness of the dimensionless variables.

iv) To comment on the best strategy in directing investment in H and K data collection honoring the economic costs of data collections.

1.5 Terminology

A number of technical terms are frequently used in the field of groundwater modeling. To avoid confusion, the specific definition of each is qualitatively introduced presently. Mathematical specifications of several can be found in Chapter 3.

i. Conceptual model vs. numerical model: Conceptual model is a simplified, conceptual representation of characteristics of groundwater system or aquifer, and always concerns all available information related, like geology, boundary and initial conditions, discharge and recharge, hydrological source, etc.

   Numerical model is to mathematically solve the physical governing equations with numerical methods such as discretization or finite difference approach.

   Parameter: In groundwater system there is a set of physical properties configuring general condition and behavior of aquifer that are relatively consistent with time but always spatially distributed, those properties are called parameters. Popularly concerned parameters in groundwater modeling include conductivity, recharge, thickness of aquifer and storage coefficient etc. The
process using limited known parameter values to determine overall model inputs throughout temporal and spatial scale of modeling is called parameterization.

iii. Objective function: A mathematical function used to quantitatively calculate misfit between measurement and model prediction.

iv. Calibration: The process that modelers refine the simulation to better represent realistic circumstance and obtain the best agreement of measured data and model output (i.e. the minimum value of objective function). The term is equally referred as parameter estimation.

v. Insensitivity: Inclusion of sensitivity in model calibration is used to derive model generation from parameter. Insensitivity occurs when field observation does not contain sufficient information to promote successful parameter estimation.

vi. Non-uniqueness: More than often in hydrologic modeling, multiple combinations of parameter values can provide modeler with similar calibration results. This is always led by the heterogeneity of simulated system and simplification of the model.

vii. Regularization: The effort made to achieve uniqueness of the solution, which is exact parameter value obtained from calibration.

viii. Pilot points: A set of scattering points each of which is assigned with hydraulic property values, which are further used as parameters in model calibration.
Chapter 2 Methodology

2.1 Introduction

In this chapter, we mathematically present study on data worth in reducing model calibration errors. The worth of data, in the context of this paper, is defined as the reduction of root-mean-squared-error of log-transformed K due to the inclusion of the data. Following conventional principle, the error is firstly obtained from model-to-measurement misfit of water head, one of the most popular observations used to calibrate groundwater model. Beyond that, the innovation of present work lies in the exploration of “parameter error” facilitated by synthetic “true” values of parameter. Thus, the discrepancy between estimated parameter field and synthetic parameter field constitutes the second error to expose the reliability of our model.

The worth of data measurements is also studied with a two-pronged strategy. Site observation of hydraulic water head has been verified on its capability to represent the complexity of model domain. However, such representation is characterized by indeterminate limitation, which can be partially relieved with abundant data collection under the idealized premise that all the measurements are obtained precisely without noise. As the criteria of calibration process to tell model when to cease iterative calculations, the misfit between field observations and their correspondences in model outputs is always reduced until reaching an acceptable threshold specified by modeler. Direct measurement to parameter itself has been demonstrated in efficiency to improve model calibration performance (Wagner, 1999; Dausman et al., 2010; Moore et al., 2011). Part of our task is to quantify the contribution brought by measured parameter values at certain locations, namely, parameter sampling, of the domain to reduce calibration errors. Conclusively, errors regarding both parameter and observation sets derived in model optimization with various data content pertaining to H and K are comprehensively compared.
The assumptions of this study include:

i. Steady State

ii. Two-dimensional simulation

iii. Uncertainty-free conceptual model, that is, domain state variables like geometry, boundary conditions, recharges are known for sure

iv. Noise-free measurement, that is, we did not consider the uncertainties in the synthetic K and H data, which is again based on our overall objective. We will leave to future study the influence of data uncertainty on analysis results.

2.2 Methodology of parameter estimation in MODFLOW and PEST

In current section, we represent mathematical algorithm employed by MODFLOW and PEST to facilitate parameter estimation with inverse modeling. Those who are familiar with such methodological procedures may skip this section. We used MODFLOW and PEST as tools to serve our overall research objectives, both of which are highly developed to realize various requirements of groundwater modelers, the details inside them are very complicated and are massively beyond current research scope. Therefore, we only briefly introduce the segments of the two computer-based suites which are relevant to our study. For other details, one is recommended to go to their documentations as posted in the following.

2.2.1 Conceptual model identification

We select MODFLOW-2000 in the sake of its integrated function of solving equations of groundwater flow and transport as well as parameter estimation (Harbaugh et al., 2000). The construction of conceptual model, which contains fundamental data of simulated basin as well as the initial condition underling further parameter estimation,
is the portal of our series of numerical experiments. A suite of coverages that separates various hydrogeological features (i.e. boundary, river, well, etc.) comprise the conceptual model. A fitted rectangular frame is thus surrounded over conceptual model referring to the site’s boundaries, which we will use to launch discretization and serve as media of conceptual and numerical model.

### 2.2.2 Groundwater flow process

The most fundamental governing law in MODFLOW is the partial differential equation (PDE) of groundwater flow as in Equation 2.1

\[
\frac{\partial}{\partial x} \left( K_{xx} \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( K_{yy} \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left( K_{zz} \frac{\partial h}{\partial z} \right) + W = S_s \frac{\partial h}{\partial t} \tag{Equation 2.1}
\]

where \( K_{xx} \), \( K_{yy} \), and \( K_{zz} \) are hydraulic conductivities along three directions (i.e. x, y and z) in 3-D coordinate system, \( h \) is potentiometric water head, \( W \) stands for source/sink term that is volumetric flux out of per unit volume of soil, negative \( W \) represents outflow while positive \( W \) represents inflow, and \( S_s \) is specific storage of aquifer (Harbaugh, et al., 2000). When both sides equal to zero, steady state is shown. Heterogeneity can be represented by inequality of \( K \) along with three directions. The partial differential equation (PDE) should be projected into spatially distributed grid system generated by discretization when one applies it to flow in aquifer seeking its numerical solution, in which condition appropriate initial and boundary conditions are required to be identified.

The discretization of groundwater domain is processed both vertically and horizontally. The cells’ identification is assigned with \( i, j, k \) to notate numbers of a particular row, column and layer respectively. Central finite-difference approach with half spatial step is used in MODFLOW to estimate both hydraulic head and conductivity in cell grids. The formula arrangement of the discretization can be shown in Equation 2.2 and the corresponding qualitative illustration is given in Figure 3:
where $m$ stands for time step and $\Delta x$, $\Delta y$, $\Delta z$ are spatial steps along with each direction in the 3-D coordinate system. Theoretically in Figure 3 the width of each column (spatial step in X-axis direction) and the height of each row (spatial step in Y-axis direction) can be different from each other (Harbaugh, et al., 2000) based on geometry, geology and topography of basin. Yet, we choose to use the uniform-sized quadrates of the grid cell in this exercise to promote interactions among multiple computer programs and release computation burden, with the addition that we are concentrating on a large site that contains tens of thousands of grid cell which makes “personalized” grid size infeasible.

**Figure 3.** Discretization scheme used by MODLFOw to numerically solve groundwater process equation (Harbaugh et al. 2000).
2.2.3 Parameter estimation

The computer-based program PEST, which stands for Parameter ESTimation, was firstly released by John Doherty (1994) and has been extensively used in hydrologic modeling. Besides its powerful capability on model calibration with the assistance of highly parameterized inverse approach and mathematical regularizations embracing Gauss-Marquardt-Levenberg method, the PEST software suite’s efficiency is enhanced by its complementary functionalities of predictive error and uncertainty analysis as well as stochastic simulation utilities (Doherty, 2010; Doherty and Hunt, 2010).

The current study respects advantage of PEST suite on groundwater modeling and analysis in steady state. The equational algorithm installed in PEST is briefly introduced in the following.

Firstly we assume the linear Equation 2.3 as following,

\[ Xb = c \]  

Equation 2.3

where vector \( b \) contains the parameters of the simulated system with the order of \( n \), the m-dimensional array \( c \) is the system response that will further be compared to observations. Thus, the matrix \( X \) should be with a dimension of \( m \) by \( n \) and is interpreted as a model operator on parameters to generate outputs describing the forward action of the model. Essentially \( X \) contains sensitivity of involved parameters to corresponding model generated observations and can be equivalent to Jacobian matrix in linear problems (Tokin and Doherty, 2005; 2009).

As what has been stated, a model can never perfectly simulate realistic hydrologic system due to its infinite complexity, that is, there is always a discrepancy between model output \( c \) and the corresponding field measurements. Here however, we temporary presume the model is perfect to match outputs strictly with measurements,
in which condition the only potential error of model is that the measurements can be noisy. Equation 2.3 can be modified as following in such an assumption to remain valid,

\[ c = Xb + \varepsilon \]  

Equation 2.4

where \( \varepsilon \) is the term of innate error associated with field investigation itself.

On the other hand, some criteria need to be defined to evaluate model-to-measure misfit based on which “goodness” of values of current parameter set can be drawn. The quantitative existence of such criteria is objective function as in Equation 2.5

\[ \Phi = (c - Xb)^T(c - Xb) \]  

Equation 2.5

where \( c \) is real values from measurements, superscript \( t \) is matrix transpose operator; \( \Phi \) indicates the discrepancy of simulation against realistic values.

Under most circumstance where there are more than one observation categories in the model, for instance, groundwater hydraulic head in a number of wells and stream flux at some gauges, some of them may lean more to measurement noise. Furthermore, considering the issue of physical units, some observations can be significantly greater than others in terms of numerical values. In such cases, observations with greater values impact more to objective function result defined by Equation 2.5, in regardless of their reliability. To overcome this unfairness, the objective function is developed to be weighted as,

\[ \Phi = (c - Xb)^TQ(c - Xb) \]  

Equation 2.6

where \( Q \) is weight matrix used to define the greater contribution of a certain pair of observation and model computation to objective function with greater scalar in the matrix.
So far the algorithm has been described is typically for linear systems, which is not the phase of most models, but still stabilizes the foundation of further discussions. The models comprising integrated hydrologic processes are always characterized by high nonlinearity, even when studying groundwater hydrology only. In nonlinear systems, with minimization of $\Phi$ in Equation 2.6 parameter vector $b$ cannot be derived straightly from measurements. Subsequently, Equation 2.7 is derived to solve the problem in assistance with Taylor’s theorem as following:

$$c = c_0 + J(b - b_0) \quad \text{Equation 2.7}$$

where $b_0$ is the set of parameters that generates computed observations $c_0$. $b$ and $c$ are another pair of vectors slightly differs from $b_0$ and $c_0$; $J$ is Jacobian matrix of the function that maps n-dimensional parameter space to m-dimensional observation space. The elements in a certain row of $J$ are actually comprised of derivatives of one particular observation relative to each of the parameters. Thereupon the objective function is alerted to be

$$\Phi = (c - c_0 - J(b - b_0))^TQ(c - c_0 - J(b - b_0)) \quad \text{Equation 2.8}$$

where $c$ is now observation values from field measurement. Equation 2.8 is essentially an operation of linearization of the problem. It can be noticed that when the objective function is minimized and $b - b_0$ is replaced by $u$, named as upgrade vector, the following relationship can be drawn:

$$u = (J^TQJ)^{-1}J^TQ(c - c_0) \quad \text{Equation 2.9}$$

Unlike linear models, nonlinear relationship between parameters and observations, which characterizes most model conditions, cannot be optimized within one step. Linearization facilitated by Equation 2.7 is essentially a forward estimation with Taylors expansion, where the error is inevitably involved and so as to the upgrade vector $u$. Initial values comprising $b_0$ should be defined to launch iterative process
by calculating initial model output \( c_0 \) according to Equation 2.3. Upgrade vector \( u \) is then procurable to be added to current parameter vector. Updated vector derived in such way is then used as start to determine further upgrade vectors. Such iterative process continues until certain termination criteria being reached, at which point global objective function minimum may or may not be achieved.

The criteria used to cease execution by PEST guarantee no more than a reasonable termination, which is either optimal parameter has been calibrated or further execution will not persist in diminishing objective function \( \Phi \). To realize such a performance, a number of different criteria are compiled in PEST and will be specified later with our synthetic model design. As what has been stated, PEST denominates termination criteria from multi aspects. Foremost, value of objective function is argued to halt execution with three factors, which are Max Number of Relative Convergence Iteration (NPHISTP), Relative Convergence Limit (PHIREDSSTP) and Max Number of Iterations with No Improvement (NPHINORED). If there have been NPHISTP iterations within all up to date differ no larger than PHIREDSSTP, or the execution fails to lower the objective function over NPHINORED successive iterations, the optimization process will be ceased. Furthermore, parameter adjustment is another indicator that PEST desires no more iteration. If the maximum parameter change during the last Max Number of Parameter Change Iterations (NRELPAR) is less or equal to Relative Parameter Change Criterion (RELPARSTP), the PEST commands unlikelihood to achieve better result. The next factor can be assigned to PEST is Max Number of Iterations (NOPTMAX), when which is completed, execution is terminated in regardless of objective function and parameter set.

2.2.4 Implementation of Gauss-Marquardt-Levenberg approach

At current stage the interest lies in minimizing the value of the objective function, \( \Phi \),
which raises the gradient of an objective function. The derivative of $\Phi$ regarding parameter vector is given by

$$g_i = \frac{\partial \Phi}{\partial b_i} \quad \text{Equation 2.10}$$

where $g_i$ is the $i^{th}$ element in the vector of the gradient of the objective function, $b_i$ is the $i$th parameter. It can be told that Equation 2.10 provides us with the steepest direction of objective function descent, which is the opposite against $g$. However, as stated by Doherty (2010), $u$ given by Equation 2.9 is “normally a far better parameter upgrade direction than $-g$”. Consequently, effort is done to deduct the difference between two directions of $u$ and $-g$ by introducing “Marquardt parameter” (Marquardt, 1963) and the following equation can be derived from Equation 2.9:

$$u = (J^TQJ + \alpha I)^{-1}J^TQr \quad \text{Equation 2.11}$$

where $\alpha$ is Marquardt parameter, $I$ is an identity matrix and $r$ is residual vector for parameter set equivalent to $c-c_0$.

### 2.2.5 Regularization

With the objective function aiming to minimize model-to-measurement misfit, estimated parameter set may encounter with non-uniqueness, that is, various solutions of parameters can relatively lead objective function to the same stage. Worse still, if one makes too much to reduce the value of $\Phi$, the subsequent parameter set probably comprises unrealistic elements (Tokin and Doherty, 2005). To solve such undesired happiness, PEST employs the functionality of regularization that is expressed with regularization objective function as

$$\Phi_r = (d - Rb)^TQ_r(d - Rb) \quad \text{Equation 2.12}$$

where $Q_r$ is a matrix comprising squared weights assigned to each regularization
observation diagonally, \( R \) encapsulates operations to map parameter vector \( b \) to “regularization observation” vector \( d \), which is quoted because it actually contains preferred conditions pertaining each term in parameter set \( b \). For instance, \( R \) dedicates to generate the sensitivity differences of two parameters in \( b \). From modeler’s expert knowledge or laboratory measurements (which is normally referred as prior information) the two parameters should have the same impact on the simulated domain, i.e., the counterpart in \( d \) is deem to be zero. In present condition the corresponding term in \( Rb \) has to incline to zero as much to minimize \( \Phi_r \) by managing parameter vector \( b \), which otherwise brings penalty to the \( \Phi_r \).

Now since we have two criteria (i.e. the two objective functions) to justify whether model misfit reaches tolerable or not, global objective function \( \Phi_g \) is given as in Equation 2.13 to synthesize the two. Despite that intention to decrease \( \Phi_r \) always leads to \( \Phi \) increasing (Doherty, 2010), PEST emphasizes more on \( \Phi_r \) while ensuring that \( \Phi \) is “suitably low” resulting in that \( \Phi \) is generally a bit greater than it could possibly be calibrated to. The multiplier \( \mu \) defines the proportion of contribution to calibration imperfection brought by regularization objective function, in other words, the departure of estimated parameters against the preferred condition. The suitable value of \( \mu \) can by evaluated through studying uncertainty in various determinations of \( \Phi_r \), which is however always not accessible (Doherty 2003b). Alternatively, an upper bound of \( \Phi \) can be defined as \( \Phi^l \) such that calibration can be satisfactory without forcing \( \Phi \) to be possibly minimum. \( \Phi^l \) limits the maximum tolerable model-to-measurement misfit below which calibration is judged to be done. As the iteration of parameter estimation being undertaken seeking the minimization of global objective function, with the assumption that \( \Phi_r \) has reached minimization, the multiplier then has to be functionalized to meet the requirement, which in practice is normally equality of measurement objective function value \( \Phi \) to its upper bound of tolerance.

\[
\Phi_g = \Phi + \mu \Phi_r
\]

Equation 2.13
2.3 Process model development

2.3.1 Synthetic model construction

We used a rectangular domain for the numerical experiments. The fact that only one layer is simulated in two-dimensional domain indicates the synthetic aquifer is unconfined. Although the geometry of model is highly simplified, it features major hydrologic processes of a realistic aquifer and is dedicated to making the following numerical experiments simplified and reproducible. Top and bottom elevations are 150 m and -100 m representatively, and the length of side is 8000 m. Please note that since our analysis will in the dimensionless form, the domain size, etc., does not need to be attached to a certain unit, as long as it is consistent with the rest of the variables, e.g., K ([L/T]). The complexity of the domain is characterized by spatial heterogeneity of hydraulic conductivity without any source or sink exists in the model. The domain is bounded by a specific head (Dirichlet) boundary conditions, which include a linear river with water head of 130m located at the east most column of domain grid, and other three no-flow boundaries. Both recharge and boundary condition are consistent in the phase of studying data worth to calibration errors, and the two hydrogeological features will be altered when examining their impacts to model calibration.

Six random K fields distinguished by different spatial data correlation lengths, \( \lambda \), were generated with FIELDGEN, a supplemental utility for the model Water Flow and Balance Simulation Model (WaSiM) (Schulla, 2015) (see Appendix I). A grid, each cell of which will contain a random value, was defined to match domain dimensions (cell size and cell number) and coordinates (origin point). A set of normally distributed values is then randomly generated with of user-specified mean, standard deviation, and most importantly, data correlation length indicator (the values specified in FIELDGEN is not actual data correlation length). Such a set of values is thus interpreted to become exponents to base 10, thus, the values in each of six fields are all subject to log-normal distribution. In order to make calibrations comparable, the ranges of random K values
are adjusted to be 0.5 to 200 in all generations by slightly modifying the aforementioned mean and standard deviation in execution, the six K fields will thereafter be used in model calibration. The correlation lengths of the normal-distributed \( \log(K) \) fields (Rehfeldt et al. 1992) are depicted using empirical variogram (Curran, 1988; Eggleston et. al., 1996; Gringarten and Deutsch, 2001) to characterize the domain heterogeneity. However, experimental model plotted on distributed data can be fluctuated thus bring indeterminacy to the evaluation. Various model variograms normally including Gaussian, Spherical, Linear and Exponential are available to trend the experimental variogram with identical property values of variogram, say, sill, range and contribution by weight least squares (Robinson and Metternicht 2006; Webster 1986). Here we selected the spherical model to portray the trend of experimental semi-variogram (Neuman, 2012) with pre-determination of variogram lag, \( h \), and number of lags, \( n \) according to domain size. The derivations of empirical semi-variogram and spherical model variogram can be described as

\[
\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (x_i - y_i) \quad \text{Equation 2.14}
\]

where \( N(h) \) is the number the pairs of points the distance of which fall within lag \( h \); \( x_i \) and \( y_i \) are values of two points of pair \( i \); \( \gamma(h) \) is the semivariogram value specified by lag distance \( h \) (Deutsch and Journel, 1992, P. 39-43),

\[
\gamma(h) = \begin{cases} 
1.5 \frac{h}{\lambda} - 0.5 \left( \frac{h}{\lambda} \right)^3, & \text{if } h \leq \lambda \\
c, & \text{if } h > \lambda 
\end{cases} \quad \text{Equation 2.15}
\]

where \( \lambda \) and \( c \) are range and contribution of the model variogram.

To numerically solve groundwater flow Equation 2.1, discretization with Equation 2.2 is implemented with 80x80 regular grid cells resulting cell size being 100m×100m. The grid is constructed based on a two-dimensional coordinate system with the left-lower
corner of the southwestern cell being the origin.

2.3.2 Observation abstraction and parameterization

Given K value at each of the grid cells, forward calculation of Equation 2.1 can provide water head all over the domain. The estimation of a K field requires observed H values from observation wells and optionally known parameter (i.e. K) values (Tonkin, et al., 2007; Tonkin and Doherty, 2009). Although a range of strategies such as uniformly random sampling or regionally-stratified sampling could be done (Amini et al., 2008), we employed an evenly-distributed sampling design to for simplicity and reproducibility and to reduce the number of experiments (Figure 4). Water head at each observation well is extracted from a forward simulation with aforementioned synthetic K fields, recharge and boundary conditions. Based on these synthetic observation wells, we randomly assign synthetic K values to x% of them.

![Figure 4. Locations of observation points (black) and pilot points (red). Boundaries of cells constituting domain grid are shown with black lines.](image)

As a more advanced alternative to zonation (Moore and Doherty 2006), pilot point method has been extensively used in groundwater model calibration. The earliest
application of pilot points dates back to 1990s, for instance, Certes & de Marsily (1991); LaVenue & Pickens (1992); Ramarao et al. (1995), just to name a few. Some of them included locations of pilot points in their study interest in terms of sensitivity analysis. This study employs the approach of pilot point to parameterize the model as well. It has been suggested that pilot points can be more intensive in areas with more observation data (Doherty et al., 2010). Considering the regularity of our domain geometry, similar with observation locations, pilot points are also evenly spread over the domain.

Recalling the iterative inverse process expressed in Section 2.2.3, parameter vector $b$ is now comprised by values assigned to pilot points. Together with observations, locations of pilot points are depicted in Figure 4.

In accordance with Section 2.2.5, Tikhonov Regularization is selected in our model to provide parameters (i.e. pilot points) with prior information, the implementation of which is homogeneity preference meaning that in regardless of other necessity, the neighboring pilot points should ultimately have the same or very similar values (Doherty, 2010).

### 2.3.3 Termination criteria of inverse modeling iterations

The inverse modeling processes iteratively until the value of objective function reaching certain calibration objective as being discussed at the end of Section 2.2.3. As PEST is highly case-sensitive, there is no explicit instruction of “correct” values that can be assigned to termination control factors. However, the PEST developer, Doherty (2013) proposed some “conservative settings” of the termination criteria, which are dedicated to preventing a slowing down of optimization from leading to premature termination. The PEST manual (Doherty, 2010) also suggests a series of values that are suitable for most exercises to the criteria. In addition that we are presently operating a synthetic case with simplification and assumptions, termination criteria listed in Table 1 are weighed and validated.
Table 1 Values of calibration termination control factors

<table>
<thead>
<tr>
<th>Control</th>
<th>Full name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOPTMAX</td>
<td>Max Number of Iterations</td>
<td>30</td>
</tr>
<tr>
<td>NPHINORED</td>
<td>Max Number of Iterations with No Improvement</td>
<td>3</td>
</tr>
<tr>
<td>NPHISTP</td>
<td>Max Number of Relative Convergence Iteration</td>
<td>3</td>
</tr>
<tr>
<td>PHIRESTP</td>
<td>Relative Convergence Limit</td>
<td>0.005</td>
</tr>
<tr>
<td>NRELPAR</td>
<td>Max Number of Parameter Change Iterations</td>
<td>3</td>
</tr>
<tr>
<td>RELPARSTP</td>
<td>Relative Parameter Change Criterion</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Note: The value of NOPTMAX should never be the reason that parameter estimation stops executing, so that optimization of parameter set is guaranteed. As a matter of fact, all the calibrations experiments in present study are terminated within 30 iterations.

2.4 Dimensionless control variables for data densities and calibration errors

The density of observation data needs to be examined relative to the spatial heterogeneity of the field, which is characterized by $\lambda$. Smaller $\lambda$ indicates more rapidly varying $K$ in space which requires more data measurements to constrain. To reduce the degrees of freedom, we propose a dimensionless number, the effective data density, $\mu_H$, which quantifies the ratio between correlation length and effective observation data density:

$$
\mu_H = \frac{\lambda}{d_H} = \frac{\lambda}{\sqrt{A/n}}
$$

Equation 2.16

where $\lambda$ is the correlation length of the log($K$) field that is simulated by spherical model variogram; $A$ is the domain area; $n$ is the number of wells where $H$ is measured. $d_H$ measures the average distance between data points. $\mu_H$ can be understood as “the square root of the number of data points in a square box with an area of $\lambda^2$”. Greater $\mu_H$ indicates slower variation of $K$ relative to the distance between measurement points.
and thus more available information about the field. Although the hypothetical observation wells in present case are evenly located and the intervals of one to next others are all the same along with both coordinate axis, such a definition essentially integrates the water head information content with domain hydrogeology state (i.e. heterogeneity level), which makes the quantification of data acquisition to be more representative when being applied to different domains.

Data worth of K to model calibration is of primary concern in this study. Similarly, we can quantify the relative density of wells with known conductivity values, which are presumably obtained from pumping tests. We implement this by randomly assigning synthetic K values to $x\%$ of observation wells. $x\%$ is systematically varied at each level of $\mu_H$ to examine the effect of conductivity measurements, their interactions with H measurements, and the effect of improved empirical variogram as a result of more known K values. It gives rise to the average distance between known K values, $d_K$:

$$d_K = \sqrt{x\%} \times d_H$$ \hspace{1cm} \text{Equation 2.17}

where $x$ is observation well proportion where K is measured besides H. Then, similar to Equation 2.16, the dimensionless factor to quantify conductivity data $\mu_K$ is derived as:

$$\mu_K = \frac{\lambda}{\sqrt{A/(n \times x\%)}}$$ \hspace{1cm} \text{Equation 2.18}

After each iteration in parameter estimation executions, updated parameter values at pilot points are interpolated to all other cells in the domain. We selected different interpolation methods when the executing the model calibration with and without parameter measurement. When there are certain amount of parameter sampling data serving the estimation of parameter field, we used Kriging as interpolation method as Kriging approach has been agreed to be an optimum unbiased interpolation method with minimum mean residual and has been widely applied to groundwater head and table evaluation study (Desbarats, et. al., 2002; Rivest, et. al., 2008; Theodossiou and
Latinopoulos, 2006). With the particular K data measurements, Equation (2.14) was recalled to plot empirical variogram pertaining to the given parameter data points and again, the spherical model variogram (Equation 2.15) was fitted by Kriging for it can provide a specific contribution value for any distance between the interpolated cell and the pilot points. On the other hand, when there is no conductivity data in the calibration execution, we used a simpler approach, the Inverse Distance Weighted (IDW) interpolation to project the estimated values at pilot points to all the grid cells.

When processing parameter sampling to supply model calibration, in another word, when \( \mu_K \) is non-zero, selections among wells being conducted with pumping test are randomly considered. However, as described in Introduction section, parameter estimation result can vary depending on exact locations where field investigations are conducted. In all experiments, water head observation wells are held constant in locations. However, the locations of known K are randomly assigned among the presumed wells. To obtain robust results, for each of \( x\% \) other than zero, three sets of random locations are chosen for known K, leading to three separate calibrations, the results of which are then averaged to obtain final calibration errors corresponding to each K field generation.

The model-to-measurement misfit of hydraulic water head is firstly calculated at each of the wells after model run. The scattered water head residuals are thereafter cast to Root Mean Square Error (RMSE). The equation of such a derivation can be shown as:

\[
\text{RMSR} = \sqrt{\frac{\sum_{i=1}^{n}(H_{i}^{o} - H_{i}^{c})^2}{n}} \quad \text{Equation 2.19}
\]

where \( H_{i}^{o} \) and \( H_{i}^{c} \) are the synthetic and calibrated groundwater head at the location \( i \), respectively; \( n \) is the total number of observation wells. To eliminate the disturbance of water head range all over the domain to parameter estimation, we developed a normalized water head error \( e_H \) as following:
\[ e_H = \frac{RMSE}{H_{max} - H_{min}} \]  

Equation 2.20

where \( H_{max} \) and \( H_{min} \) are maximum and minimum water head across the domain.

Similarly, for hydraulic conductivity, since \( K \) is lognormally distributed, the error, \( e_K \), is calculated as a normalized Root Mean Squared Logarithmic Error (RMSLE) as the following:

\[ RMSLE = \sqrt{\frac{\sum_{y=1}^{N} \sum_{x=1}^{M} [\log(K_{(x,y)}^o) - \log(K_{(x,y)}^c)]^2}{M \times N}} \]  

Equation 2.21

\[ e_K = \frac{RMSLE}{\log(P_{90}) - \log(P_{10})} \]  

Equation 2.22

where \( K_{(x,y)}^o \) and \( K_{(x,y)}^c \) are the synthetic and calibrated hydraulic conductivity values at the cell \((x,y)\), respectively; \( M \) and \( N \) are total numbers of columns and rows in the domain, respectively; \( \log(P_{90}) \) and \( \log(P_{10}) \) are 90\(^{th}\) and 10\(^{th}\) percentile of synthetic conductivity values of each field.

Manipulations above on normalized calibration errors and effective date densities extinguished the units thus to reduce the number of experiments and simplify the experimental design. If they are proven to effectively and adequately characterize the system, we can avoid simultaneously varying data density and spatial heterogeneity, but instead we adjusted only the latter. To enhance the dimensional validation of the data densities to normalized calibration errors, we compared the errors of calibrated \( K \) from a series of experiments, which contain different combinations of \( \lambda \), \( n \) and \( x\% \) that produce similar dimensionless numbers. 3 different values \( \mu_H \) values and 3 different \( \mu_K \) were tested while keeping the other variable constant. The relative numbers of observation and parameter can affect inverse process dramatically (Christensen and Doherty, 2008; Lavenue and de Marsily, 2001; Floris et al., 2001), therefore to remove an extra matter that can disturb the verification, number of pilot points is always set to be 35\%-45\% of observation points, as, again, even distribution of both pilot points and observation wells impedes exact same calculations in different
experiments.

As our synthetic domain is highly simplified, when PEST ceases its execution, the model estimated water head values at the hypothetical wells are all well matched to observed values. Figure 5 presents an example of all the calibration experiments and such plots of all the other calibrations are very similar. The goodness of model calibration establishes a valid foundation of the following research.

![Figure 5. Goodness of match between model computed and observed water head values at 256 hypothetical observation wells](image)

2.5 Calibration errors to different recharge and boundary condition

Inaccurate boundary condition may also play a role in groundwater model error and uncertainty (Refsgaard, 1997) as the solution of groundwater flow equation (recalling Equation 2.1) is partially relies on domain boundaries, and is probably the most critical step in the conceptual model establishment (Franke, et al., 1987). In the work of Beven and Binley, (1992) acceptable error of predicted boundary condition is regarded as one of the criteria to admit a certain set of parameter values.

We examined the impacts of recharge and boundary condition on calibration errors. 6 recharges ranging from 100mm/year to 1000mm/year are supplied to domain and PEST was run for two different magnitudes of $\mu_H$. To prevent the increase in recharge from
raising the water table above ground surface elevation, the specific water head boundary was decreased to 115m while all the other three are kept to be no-flow still. Other model settings are identical to previously described. Different recharge inevitably bring different water head range in the domain (see Table 2), thus the absolute calibration errors can show separate pattern from normalized errors.

![Figure 6. Boundary condition transaction from initial (left) to the comparison (right). Red cells represent those of specific head and gray color is for no-flow boundaries. Blue means the cell is activated in numerical simulation and the black exhibits the procedure of transaction.](image)

We tested how the calibrations errors change with the difference of the Dirichlet boundary conditions by defining another specific-head boundary to replaced the no-flow boundary on the left domain edge. Two H data densities are selected to process model calibrations: zero parameter estimation and 50% of the 256 wells being conducted pumping test to obtain K, both of which are consistently associated with original $\mu_H$ values. In such a stage of study, we kept recharge uniformly to be 500 mm/yr thus the measurements at the two specific-head boundaries are reversed back to be 130m (Figure 6).
Table 2. Ranges of hydraulic water head with each of the recharges under two H data densities regarding $\mu_H$

<table>
<thead>
<tr>
<th>$\mu_H$</th>
<th>Recharge mm/year</th>
<th>$H_{max}$</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.45</td>
<td>100</td>
<td>117.94</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>120.85</td>
<td>5.85</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>123.72</td>
<td>8.72</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>129.35</td>
<td>14.35</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>134.84</td>
<td>19.84</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>124.85</td>
<td>27.85</td>
</tr>
<tr>
<td>7.24</td>
<td>100</td>
<td>118.54</td>
<td>3.54</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>122.03</td>
<td>7.03</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>125.46</td>
<td>10.46</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>132.18</td>
<td>17.18</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>138.73</td>
<td>23.73</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>148.17</td>
<td>33.17</td>
</tr>
</tbody>
</table>

Note: Minimum water head, $H_{min}$, is constantly 115m as a result of specific head boundary, and is not shown in the table.

2.6 Experimental design and multi-variable polynomial curve fitting

Different $\mu_H$ is achieved by the 6 values of $\lambda$ in the K field generation, which keeping observation wells consistent. We tested a total of 30 pairs of $(\mu_H, \mu_K)$, with 6 $\mu_H$ and 5 $x\%$ levels, which will be shown further results. For each pair where $\mu_K$ is non-zero, we ran three random assignments of known K values. This experimental design results in a total of 78 calibration experiments. As discussed previously, all experiments have the same domain geometries and water head observation wells.

To produce a reusable formula for errors, we fitted error as a polynomial function of $\mu_H$ and $\mu_K$ values:

$$e(x, y) = P \cdot \bar{\mu}$$

Equation 2.23

where $P$ is a vector of the coefficient in polynomial fitting, $\bar{\mu} = [1, \mu_H, \mu_H^2, ..., \mu_K, \mu_K^2, ..., \mu_H \mu_K]$ is the vector of predictors, $e(x, y)$ is the calibration error. Our experiments are constrained within $\mu$ values tested. $P$ is adjusted
in Matlab® curvefitting Toolbox to minimize the squared residual. Firstly the fitting is
processed with original polynomial Equation (2.23) of 2\textsuperscript{nd} degree concerning the 30
data points in terms of both \( e_H \) and \( e_K \). The term \( \mu_H \mu_K \) in Equation (2.23) indicates
the interaction of the two separate field investigations. We performed stepwise T-test to
each of the curve fitting coefficients to evaluate their statistical significance.
Furthermore, the goodness of fitting is evaluated using Adjusted R-Square value (R\textsuperscript{2})
and the Root Mean of Squared Error of fitting (\( e_f \)). As the 2\textsuperscript{nd} orders terms have been
demonstrated to be unnecessary, we did not further examine higher orders embodied in
Equation 2.23.

### 2.7 Relative data worth and economic analysis

First, our experimental design presumes K to be measured at a certain percentage of
observation wells as we assumed when an investigator could obtain a conductivity
estimate, e.g., through pumping test, a water head reading can always be obtained.
Therefore, water head observations encompass conductivity estimates. Based on the
chain rule, we can evaluate the partial derivative of error with respect to the number of
observations:

\[
\frac{\partial e_K(\mu_H, \mu_K)}{\partial n_H} = \frac{\partial e_K}{\partial \mu_H} \cdot \frac{\partial \mu_H}{\partial n_H} + \frac{\partial e_K}{\partial \mu_K} \cdot \frac{\partial \mu_K}{\partial n_H} \tag{Equation 2.24a}
\]

\[
\frac{\partial e_K(\mu_H, \mu_K)}{\partial n_K} = \frac{\partial e_K}{\partial \mu_H} \cdot \frac{\partial \mu_H}{\partial n_K} + \frac{\partial e_K}{\partial \mu_K} \cdot \frac{\partial \mu_K}{\partial n_K} \tag{Equation 2.24b}
\]

where \( n_H \) is the number of head observations and \( n_K \) is the number of K
measurements. \( \frac{\partial e}{\partial n} \) provides a measure of marginal K error reduction due to adding one
head measurement. To help us make decisions about whether to obtain H or K data
points, we can compute two ratios that measures the relative worth of data:

\[
R_\mu = \frac{\frac{\partial e}{\partial \mu_H}}{\frac{\partial e}{\partial \mu_K}} \tag{Equation 2.25}
\]

\[
R_n = \frac{\frac{\partial e_K}{\partial n_H}}{\frac{\partial e_K}{\partial n_K}} \tag{Equation 2.26}
\]
$R_\mu$ is the relative data worth ratio with respect to unit increase in $\mu_H$ or $\mu_K$. Since $\mu$ is a nonlinear function of $n$, far more data points are needed to increase a unit of $\mu$ when $\mu$ is high. This can make $R_\mu$ more difficult to interpret and use. $R_n$, which has a stronger practical meaning, indicates whether it is more beneficial to add an H or a K data point. However, $R_n$ has three control variables: $n_H$, $n_K$ and $\lambda$. We will show both variables for the sake of comparison.

We examined relativeness of conductivity error changing with H and K data numbers. As derivation of water head error directly involves number of observation wells (i.e. $n$), hence that dimensional verification suggests that there might be other factors impacting $e_H$ besides the effective data densities, error changing rate with varying data numbers is only exposed with conductivity data densities. Such exposure is implemented both numerically and practically with monetary considerations, which will be specified in the following section as the expressions are established upon the adjusted polynomial fitting equation.

We consider two scenarios when calculating $R_n$: (A) a new K data point is always accompanied by a new H data point, which is relevant when we plan to install wells. In this scenario, $\frac{\partial \mu_H}{\partial n_K} = \frac{\partial \mu_H}{\partial n_H}$; (B) we can add a K measurement without adding H data. This scenario is relevant when we can conduct a pumping test from an existing well, or we can extract K estimates from interpreting existing literature. Under this scenario, $\frac{\partial \mu_H}{\partial n_K}$ is 0. We denote $R_n$ by $R'_n$ when it is calculated under scenario (B). Finally, if we include the cost ratio into our consideration, the relative data worth per dollar is $R_c = R_n \times \frac{c_K}{c_H}$, where $\frac{c_K}{c_H}$ is the ratio of the cost of a new K measurement and a new H monitoring well. With some preliminary market survey, we arbitrarily fixed $\frac{c_K}{c_H}$ at 5.
Chapter 3 Result and Discussion

3.1 Verification of the dimensionless numbers

The normalized conductivity error $e_K$ for calibrations with similar values of $\mu_H$ and $\mu_K$ are tightly clustered (Figure 7). In the three $\mu_H$ clusters, water head is the only data measurement in the model thus their $\mu_K$ are zero ($x\% = 0$). In each of the three $\mu_K$ clusters, all the three scatter points have the similar amount of conductivity data as well as $x\%$, thus the water head data densities they have are also very close. Although $\lambda$, $x\%$ and $n$ are varied substantially, the errors are fairly constant with similar $\mu_H$ and $\mu_K$. This behavior verifies $\mu_H$ and $\mu_K$ as effective dimensionless numbers to characterize $e_K$ of the system, and allows us to alter only $\lambda$ in later experiments. In addition, Figure 7 also suggests $e_K$ can be described as a smooth function of $\mu_H$ and $\mu_K$. In these preliminary experiments, we did not note any non-monotonicity or fluctuations. Such smoothness and monotonicity serve as the basis of fitting a polynomial function to the relationships between normalized conductivity error and effective data densities.

However, we did not obtain such a pattern for water head error. The $e_H$ values show obvious scattering even with similar effective water head data densities. Such a fact makes it meaningless to further test $e_H$ relative to $\mu_H$. In summary, $\mu_H$ and $\mu_K$ are effective dimensionless numbers to characterize the system for calibrated conductivity errors but not so for calibrated water head errors. As a result, the conclusions we later draw for $e_K$ as a function of $\mu_H$ and $\mu_K$ are applicable to different $\lambda$, $x\%$ combinations, while those for $e_H$ are only valid for $\mu_H$ and $\mu_K$ values that we have specified.
Figure 7. Dimensional verification of effective data density of both H and K to normalized conductivity errors.

3.2 Impact of recharge and boundary condition on model calibration errors

Normalized calibration errors are independent to recharge with the two selected data densities (Figure 8). Various recharge inputs lead to different water head ranges yielding different absolute error values. However, after the error is normalized with respect to the ranges of H and K, the error curves become flat and irrespective of recharge. In summary, these experiments show that the influence of recharge is linear, and can be removed by normalizing the error.

When there is no K data in the domain, the difference in the boundary condition does not matter for the $e_K$ (the upper two lines in Figure 9a). However, when there is K data, BC2 apparently becomes higher than BC1 (the lower two lines in Figure 9a). Generally the curves produce the same pattern with the two separate boundary conditions and data densities. The polynomial curve fitting equation, which will be specified later, constitutes a constant term and multiple coefficients. The similarity of the curves in their shape embraces the possible interpretation that based on different boundary conditions, the resultant curve fitting equation may vary primarily in the constant term, which will be eliminated when calculating the partial derivatives of the independent variables of the equation. Hence that the experimental data that will be used to conduct the curve fitting are all obtained under uniform boundary condition, thus such a matter
does not undermine the following study.

Figure 8. Normalized water head error (a) and conductivity error (b) changing with recharge values under two different water head data amount.
Figure 9. Normalized water head error (a) and conductivity error (b) changing with recharge values under various amount of data collection pertaining to both water head and conductivity. BC1 means domain with one specific water head boundary and BC2 stands for the domain with two specific water head boundaries.

3.3 Errors as a function of effective data densities

When we hold $\mu_H$ steady and increased $\mu_K$, $e_K$ gradually decreases as one expects (Figure 10). As described in Methods section, each point on the plot when $x\%$ is
non-zero represents the mean of three calibrations. The decline in $e_K$ is almost linear. The slopes of the equi-$\mu_H$ lines decrease slightly for higher $\mu_H$ values, and the distances between the lines become smaller at higher $\mu_K$, indicating a moderate interaction between $\mu_K$ and $\mu_H$. This narrowing means that when relative density of known conductivity is lower, the marginal gain of adding $\mu_H$ decreases. $e_H$ shows a generally similar trend but there is a more noticeable quadratic trend (Figure 10 a & b). When $\mu_K$ increase, different equi-$\mu_H$ lines can cluster together, potentially resulting from a more prominent role of randomness when there is no known $K$ in the field.

The linear decline in $K$ is somewhat surprising. Toward higher $\mu_K$ we should be able to build more accurate variogram model for the regularization procedure during calibration. Therefore, one might expect an accelerating $e_K$ decline as a function of $\mu_K$. However, at least in the range of $\mu_H$ and $\mu_K$, such an effect is hardly observable. At the same time, as $e_K$ is computed from comparing the $K$ of the entire domain, the monotonously and smoothly varying $e_K$ suggest the regularization approach is effective in reducing overfitting errors when provided with additional data.

Viewing the data in a different way, when we keep a constant $\alpha\%$, the error apparently decreases smoothly as we increase the effective data densities (Figure 10 c & d). In this Figure, as $\mu_H$ increases, $\mu_K$ also increases proportionally. When $\alpha\%$ is increased from 0 m to 10%, the reduction of both RMSE and RMSLE are more significant than it is added from 10% to 20%. This pattern is perhaps due to the moderate interaction between $\mu_H$ and $\mu_K$ mentioned previously. This happens can also partially because of the increasing amount of tested $K$ values improves the estimates of variogram which is used during pilot-point based calibration. The interpolation method of Kriging expedited by the spherical model variogram provides the inverse process with more prior information of spatial variance of conductivity. Such words can be enhanced with the further distance between calibration errors without and with parameter measurements at greater $\mu_H$ compared to that at smaller $\mu_H$ as given $K$ values are
assigned to certain percentages of H observations, in which way model abstracts more information regarding aquifer spatial variability even with the same percentage.

![Normalized K error (a) and H error (b) as well as absolute K error (c) and absolute H error (d) relative to effective data densities](image)

**Figure 10.** Normalized K error (a) and H error (b) as well as absolute K error (c) and absolute H error (d) relative to effective data densities

### 3.4 Multi-variable polynomial curve fitting

Stepwise regression tests show that high order terms (i.e. the 2\textsuperscript{nd} order) are statistically insignificant for $e_K$ (Table 3), so that the system is mostly linear in the range of $\mu_K$ and $\mu_H$ tested. The small value of the coefficient for the $\mu_K \mu_H$ compared to other terms confirms that the interaction between the two variables is mild. In the interest of parsimony, we created a 3D surface using three terms: $\mu_K$, $\mu_H$ and $\mu_K \mu_H$. Therefore, the final fitted equation that will serve further study can be written as:
\[ e(x, y) = p_0 + p_1 \mu_H + p_2 \mu_K + p_5 \mu_H \mu_K \]  

where \( p_0 \) to \( p_5 \) are fitting parameters encapsulated in \( \mathbf{p} \) of Equation (2.23). Meanwhile, the quadratic term, \( \mu_K^2 \), is statistically significant (p value =0.001) for \( e_H \). However, the \( R^2 \) without the quadratic term is high enough, and adding the term does not notably increase. In the interest of parsimony, we chose not to include \( \mu_K^2 \) in the fitted formula for \( e_H \). For the first order terms, the difference between \( p_1 \) and \( p_2 \) is \(~15\%\), suggesting that in regions where the first-order terms dominate and where \( \mu_H \) and \( \mu_K \) are similar, the data worth of new H and K data points are also similar.

<table>
<thead>
<tr>
<th>Table 3 Multi-polynomial curve fitting result and P-values of T-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration Errors</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( e_K )</td>
</tr>
<tr>
<td>P-Value</td>
</tr>
<tr>
<td>Value</td>
</tr>
<tr>
<td>( e_H )</td>
</tr>
<tr>
<td>P-Value</td>
</tr>
<tr>
<td>Value</td>
</tr>
</tbody>
</table>

Note: All the P-values are calculated on 95% confidence level in T-test. The P-value of \( p_4 \) coefficient for \( e_H \) is small but still set to be 0. The two values inside the parentheses are the values including the quadratic term, \( \mu_K^2 \) when fitting \( e_H \).

Because \( \mu_K \) cannot be greater than \( \mu_H \), as we assumed that all the pumping tests are implemented at the existing wells where water head have been observed, the valid region for fitting is limited in the left lower triangle of the \( \mu_H \sim \mu_K \) plane. The fitting surface well describes the errors as a function of \( \mu_H \) and \( \mu_K \) (Figure 11) as most points are scattered closely near the surface.
Figure 11. 3D visualization of adjusted polynomial curve fitting of effective data densities to normalized K (a) and H (b) errors with original scatters of \((\mu_K, \mu_H)\), which is represented by black spots, used in the fitting process.
Figure 12. Contour representation of adjusted polynomial curve fitting results of normalized water head error (a) and normalized conductivity error (b) changing with data densities. We also provided contour plot for the surface for more numerically accurate representation (Figure 12). The patterns of H and K data density collaboratively reduce $e_H$ and $e_K$ are similar. Additional filed investigation values most when the original available data is scarce. The magnitude that calibration errors decrease with increasing
data availability slightly recedes. The ranges of the two calibration error types varies due to their generations, $e_K$ encloses discrepancy between parameter estimated and synthetic K value cell by cell, while $e_H$ is comprised by model-to-measurement misfit at abstracted observation wells only.

Substituting the fitted polynomial equation into Equation 2.26, we can derive the ratio between marginal error reduction rates. For example, $R_\mu$ becomes

$$R_\mu = \frac{P_1 + p_3\mu_K}{P_2 + p_3\mu_H}$$

Equation 3.2

which is clearly a function of both $\mu_K$ and $\mu_H$. Our experimentally specified values of $\mu_K$ scatter mostly within the range of 0 to 4.5 (Figure 10). We further constrain $\mu_K$ within the experimentally tested range of 0.05 to 4.5. The lower bound is set to be slightly greater than 0 for the following mathematical procedure can reject 0 variables.

Given a unit increase in the effective data density, H data appears to be more effective in reducing $e_K$ (Figure 13). $R_\mu$ is >1 for most part of the plane, and it rises quickly toward the high $\mu_H$, low $\mu_K$ region near the upper left corner of the figure. This seemingly counter-intuitive pattern results from the calculation of $R_\mu$: $\mu$ is not a linear function of $n$. To increase one unit in $\mu$, a lot of data points are needed when $\mu$ is high. $\mu_H$ dominates most in model calibration improvement when the available data of the domain provides modeler with much more water head level relative to domain property (i.e. hydraulic conductivity). As the water head density gets decreasingly more than conductivity density, the domination of $\mu_H$ to calibration error vanishes and further replaced by $\mu_K$. Such a transition can be reached faster with the concentration of water head error, and in the $R_\mu$ plane of conductivity error, the worth of $\mu_K$ does not become equivalent to $\mu_H$ until the numerical values of the two are almost the same as each other.
Figure 13. $R_\mu$ of K error (a) and H error (b) with effective data densities $\mu_H$ and $\mu_K$. Bold lines highlight the less-than-one values among the contour.

3.5 Economic worth of data acquisition against model calibration error reduction

Given the polynomial form of the error and Equation 2.26, we can further obtain the
equation for \( R_n \), for scenario (A): 

\[
R_n = \frac{\partial e_K}{\partial n_H} \frac{\partial e_K}{\partial n_K} = (p_1 + p_5\mu_H) \frac{\lambda}{2\sqrt{\pi}} \frac{1}{\sqrt{n_H}} : [(p_2 + p_5\mu_H) \frac{\lambda}{2\sqrt{\pi}} \frac{1}{\sqrt{n_H}} + (p_1 + p_5\mu_K) \frac{\lambda}{2\sqrt{\pi}} \frac{1}{\sqrt{n_K}}] 
\]

Equation 3.3a

\[
R_n = \frac{\partial e_K}{\partial n_H} \frac{\partial e_K}{\partial n_K} = (p_1 + p_5\mu_H) \frac{\lambda}{2\sqrt{\pi}} \frac{1}{\sqrt{n_H}} : (p_2 + p_5\mu_K) \frac{\lambda}{2\sqrt{\pi}} \frac{1}{\sqrt{n_K}} = \frac{(p_1 + p_5\mu_H)\sqrt{n_K}}{(p_2 + p_5\mu_K)\sqrt{n_H}} 
\]

Equation 3.3b

\( R_n \) is easier to interpret as it simply shows the ratio of information content brought in by the next data point of H vs K, measured by the reduction in calibrated K error, but to examine it we must consider the nonlinear influence of \( \lambda \). When \( \lambda \) is small (Figure 14a & b), \( R_n \) ranges between 0.1 and 0.45 which means a K data point will always bring in more than twice the information value than an H data point. This difference may be counter-intuitive given Equation 3.1 since the magnitude of coefficients \( p_1 \) (-0.053) is only slightly smaller than \( p_2 \) (-0.061), but an important factor is that \( n_K \) is always much smaller than \( n_H \) in our tested ranges, which is normally the case with available groundwater data. \( R_n \) contours radiate out almost linearly in the shape of a fan, and is dense near the left edge of the figure. The linear pattern of the contours suggest \( R_n \) is almost a function of \( x\% \) for this high-heterogeneity case. When \( n_K < 10 \), \( R_n < 0.1 \), suggesting when K measurement is sparse compared to H, there is 10 times more information value in new K data points than H. However, when \( \lambda \) is large, which means when we are in a relatively data-rich environment, the influence of the term involving \( p_5 \) becomes more important, and \( R_n \) becomes markedly larger and more nonlinear (Figure 14c &d). Toward the left-edge, contours are dense and mostly vertical, indicating that, in that region, \( R_n \) mainly depends on the density of K data and H has little impact. \( R_n \) then increases toward the upper-right corner. In a very heterogeneous field, K points are more favored than in a homogeneous field. As the
fraction of K data points increases, new H data points become more and more useful relative to K. Because we cannot visualize the map for all $\lambda$ values, we provide a code (Appendix II) to do the calculations for user-specified $\lambda$ values. Finally, using a cost ratio of $c_K : c_H = 5$, we note that the majority of the $n_K \sim n_H$ plane favors new H measurements based on information content per dollar (Figure 15). Only when the fraction of K wells is very small that it favors new K wells. This plot will obviously depend on the actual cost ratio. As a summary, Figure 14 suggests that the amount of presently-available data plays the most important role in determine the marginal value of new H or K data. To put in a concise statement, we should invest in the data that is sparse.

Figure 14. $R_n$ when $\mu_H$ change with K data measurements (a, c) and $R_n'$ when $\mu_H$ is independent (b, d) based on two different domains regarding $\lambda$. 
Comparing $R_n$ of scenario (A) and $R'_n$ of scenario (B), the values of the latter are higher. That is, when K data measurement is exclusive form being associated with H data, the advantage of conducting pumping test at existing wells recesses. In a more homogeneous domain when there have already been more than 50 K data measurements (i.e. $\mu_K$ is greater than 1.6) the numerical benefits of the two data types can be almost equivalent. Different scenarios do not change the pattern of contours, which is mostly dependent on the domain heterogeneity. In a relatively homogeneous domain, the contours are characterized with higher non-linearity. This is because $\lambda$ is involved in the interaction term with coefficient $p_5$, which brings non-linearity to the ratio pertaining to data numbers (Equation 3.3). Such a fact can draw the interpretation that in the domain of higher homogeneity, the collective effectiveness of the H and K data can be more significant.

Based on the results above, we can provide some first-order guidance on how our investment should be made. Suppose we have some resources at hand for collecting a number of data points, be it drilling wells or locating data from different sources. We first need to evaluate the likely heterogeneity of the field, compile existing data, and collect rudimentary information about the cost ratios. When the K data is very sparse or non-existent, it is almost always more beneficial to be invested in some high-quality K measurements, as $R_n$ is also very small under this scenario. These values should also help us better understand the heterogeneity of the field, if not established a priori using geologic information. We can take a multi-stage, low-regret approach that add data points along the line of $R_C = 1$, here termed the decision line. When we are far away from the decision line, we should allocate investment so that we approach it as fast as possible, which means that we add H points to move vertically up toward the line when we are under it, or add K data points to approach it horizontally when we are above it. If we are close to the decision line, more data points can take us across the decision line when it favors the other type of investment. Such a strategy will optimize the allocation
of resources with low regrets.

Figure 15. \( R_c \) based on assumption that field investigation of \( K \) is often associated with an additional \( H \) data (Scenario A) within two different domains regarding \( \lambda \). Bold lines highlight less-than-one values among the contour.
Chapter 4 Conclusions and Limitations

We found that the calibrated K error, $e_K$, can be well described as a function of $\mu_H$ and $\mu_K$, which are densities of H and K data points non-dimensionalized by a normalization with respect to the spatial heterogeneity. This finding allows us to reduce the number of control factors for $e_K$. The function contains only first order terms of $\mu_H$ and $\mu_K$ and a mild interaction term. The interaction term only becomes more important when the spatial correlation length of the field is larger. The absolute value of the coefficients for the first-order terms differs only around 16% between $\mu_H$ and $\mu_K$. However, the H to K data worth ratio for the next data point, $R_n$, is strongly dependent on density of presently available data points. If we assume that a new K well entails a new H observation, in most of the tested range of data densities, a new H measurement is less than 40% of the value of a new K measurement ($R_n < 0.4$). The relative value of K increases substantially as K density decreases, and when K is sparse, this ratio can be less than 10% ($R_n < 0.1$). In a domain with higher heterogeneity, $R_n$ is mostly determined by the fraction of wells with K measurements. When the aquifer is simpler, the interaction term becomes more important and $R_n$ gets higher. Assuming a K to H data cost ratio of 5:1, we found that most the parameter-plane favors new H data point, except when the number of K data points is less than 10% of that of H. Based on these results, we suggest there is a decision-support line on the $n_K$~$n_H$ plane. Finally, a simple interpretation of our results is that groundwater investigators should first examine the existing data they have. When K data is very scarce, high priorities should be placed on obtaining K data. Our investment should move toward the decision-support line if the initial condition is distant from it, and evolve around it if we are close. We attach the code to compute $R_n$ as a function of $\lambda$, $n_H$ and $n_K$.

As the ultimate result we have determined the decision line, which is the certain contour separating values greater and less than 1. Such a line can switch with the difference in
costs measuring the two data types. Although being case-sensitive, such a finding can qualitatively enlighten the decision making that one should seek more conductivity information or acquire more water head observations. Our 1st-order evaluation suggests putting the endeavor to the rare one compared to the other. However, such rareness needs to be more dramatic for conductivity data to make it more cost-effective. Our conclusion is not only sensitive to existing data availability, but also to domain heterogeneity. Considering the high cost of conductivity investigation and its superior value in improving inverse model performance, acquiring more water head data can be safer in a more homogeneous domain.

We established a synthetic domain with a series of hypothetically generated data, and the domain is highly simplified on its hydrogeological features. Further study can involve some realistic case to further enhance the validation of present study. We did not fully expose the variables that can characterize water head error, which directs another object for further study. More advanced methods should be used to determine the optimum location to place new data points. Our work reveals that the boundary conditions do change the constant term in the function of $e_K$, but it does not alter conclusions about $R_n$. Finally, our result is done using the PEST algorithm in the environment of Groundwater Modeling System, and the results are thus conditioned on some decisions of the program, e.g., the choice of the regularization parameter.
Appendix A: Random log(K) Fields and Evaluations with Variograms

Here we visually display the random fields of log(K) generated with FIELDGEN. The data are used as “true” Hydraulic Conductivity (K) values of model domains. Empirical variograms (squared scatters) fitted with spherical model variograms (solid blue lines) are plotted exhibited beside, which quantify various heterogeneity levels of domain under study.
Appendix B: MATLAB Function to Plot $R_n$ and $R'_n$

The following code enables the plotting of $R_n$ and $R'_n$ planes of various $\lambda$ values, which is the only input (i.e. L) to execute the function. To maintain the effectiveness of the polynomial fitting coefficients, the range of valid $\lambda$ is constrained within 500 to 3000.

```matlab
function plot_Rn_and_Rnprime(L)
    if L>=725 & L<=3000
        a =0.4637;
        b = -0.05326;
        c = -0.06103; % fitting coefficient for $e_K$
        d = 0.007531;
        A=8000^2; %A is domain are
        n=24:12:256; nk=1:10:200;
        for i=1:length(n)
            for j=1:length(nk)
                fe_n(i)=(b+d*L/A^0.5*nk(j)^0.5)*L/2/(A*n(i))^0.5;
                fe_nk(j)=(c+d*L/A^0.5*(n(i))^0.5)*L/2/(A*nk(j))^0.5+
                    (b+d*L/A^0.5*nk(j)^0.5)*L/2/(A*n(i))^0.5; %mu_H is dependent
                fe_nk2(j)=(c+d*L/A^0.5*n(i)^0.5)*L/2/(A*nk(j))^0.5; %mu_H is independent to n_K
                if nk(j)<n(i)&fe_n(i)<0&fe_nk(j)<0&fe_nk2(j)<0
                    r_numb(i,j)=fe_n(i)/fe_nk(j); %dependent
                    r_numb2(i,j)=fe_n(i)/fe_nk2(j); %independent
                else
                    r_numb(i,j)=NaN;
                    r_numb2(i,j)=NaN;
                end
            end
        end
        %% plot when mu_H is dependent to n_K
        figure('Position',[0 0 1500 600])
        subplot(1,2,1)
        contour(nk,n,r_numb,20,'ShowText','off')
        colorbar
        xlabel('n_K')
```

56
ylabel('n_H')
title(['a) R_n; \lambda=',num2str(L)]);

subplot(1,2,2)
contour(nk,n,r_numb2,20,'ShowText','off')
colorbar
title(['b) R\prime_n; \lambda=',num2str(L)]);
xlabel('n_K')
ylabel('n_H')
else
    fprintf('L need to be within 725 to 3000');
end
References


Gates, J., ... & Kisiel, C. C. (1974). Worth of additional data to a digital computer model


Wada, Y., & Ludovicus P. H. van Beek, C. M. (n.d.).

Wagner, B. J. (1999). Evaluating data worth for ground-water management under

