TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Forst-, Geo- und Hydrowissenschaften

Uncertainty Assessment of Hydrogeological Models Based on Information Theory

Dissertation zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)

vorgelegt von
José Guillermo De Aguinaga, MSc.

Gutachter:

Herr Prof. Dr. Rudolf Liedl
Technische Universität Dresden

Herr Prof. Dr. Gunnar Nützmann
Humboldt-Universität zu Berlin

Dresden, 3. Dezember 2010

Erklärung des Promovenden

Die Übereinstimmung dieses Exemplars mit dem Original der Dissertation zum Thema:

„ Uncertainty Assessment of Hydrogeological Models Based on Information Theory“

wird hiermit bestätigt.

……..…
Ort, Datum

……..…
Unterschrift (Vorname Name)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

α	Gamma and Inverse Gaussian distribution parameter
α	empirical parameter of the soil water retention function [L^{-1}]
α_L	longitudinal dispersivity
α_2	empirical parameter of the soil water retention function for soil material 2 [L^{-1}]
α_3	empirical parameter of the soil water retention function for soil material 3 [L^{-1}]
a_i^2	square residuals; in Hill's exercise are squared random numbers
a_w	empirical crop parameter [-]
AIC	Akaike's information Criterion
AIC_c	a second-order AIC, necessary for small samples
AIC_c*	symmetric AIC_c criterion
β	Gamma and Inverse Gaussian distribution parameter
$\bar{\hat{\beta}}$	averaged estimated parameter
$\hat{\beta}_i$	estimated parameter value of each model *i*
β_r	regression coefficients
BIC	Bayesian Information Criterion
C	constant term in the K-L equation
$\hat{c}$	QAIC estimated variance inflation factor. When no overdispersion occur *c* = 1 and QAIC is equal to AIC
Co	Courant criterion, it has to be smaller or equal to one for numerical stability
$Cond_{Riverbed}$	Conductance of the riverbed (m^2/s)
C_P	partition coefficient also known as surface cover fraction [-]
CV	Cross Validation
D	dispersion coefficient
Δ_i	Akaike differences; difference between each AIC value and the minimum AIC value of the set
Δt	transport time step

Δx cell length in flow direction

df degrees of freedom

e Euler's number; mathematical constant e = 2.71828… Base of the natural logarithm

ε residual, deviation from observed value

e vector of residuals of the form [y-y'] where y is an observation vector and y' the simulated value vector. All vectors have a dimension equal to the observed values

E soil water potential evaporation

ET evapotranspiration [L]

ET_0 reference potential evaporation measured at the lysimeter station

$e^T \omega e$ weighted residual sum of squares

exp exponential function e^x

F observed Fisher information matrix

$\langle F \rangle$ expected Fisher information matrix

f(x) "truth", the processes that produce the measured data

GI God existence indicator

g(n,k|p ;binomial) binomial probability function of data *n, k* given parameter *p*

g(x) a model, which is expected to provide an approximation to the measured data

$g(x \mid \theta)$ probability function of data *x* given parameters *θ*

h hydraulic pressure, also known as water pressure or head [L]

h hydraulic pressure, also known as water pressure or head [L]

hCritA absolute value of the minimum allowed pressure head at the soil surface [L]

HK_1 horizontal hydraulic conductivity of layer 1 (m/s)

HK_2 horizontal hydraulic conductivity of layer 2 (m/s)

ini. h initial heads

$I(\theta)$ matrix of K x K dimensions

$I(f,g)$ K-L number; information lost when *g* is used to approximate *f*

I-direction Y axis

J Jacobian (sensitivity) matrix

$J(\theta)$ matrix of K x K dimensions

J-direction X axis

k number of successes

K hydraulic conductivity [LT^{-1}]

K number of estimable parameters. If σ^2 is known K = number of hydrologic parameters; otherwise, K = number of hydrologic parameters + 1 for the unknown statistical parameter $\hat{\sigma}^2_{ML}$

K_C crop coefficient [-]

K_h horizontal hydraulic conductivity [LT^{-1}]

KIC Kashyap Information Criterion

$K_{Lay\ 1}$ hydraulic conductivity in layer 1 (m/s)

$K_{Lay\ 1\ channel}$ hydraulic conductivity zone in layer 1 (m/s); resembles a paleo-channel

$K_{Lay\ 1\ without\ channel}$ hydraulic conductivity zone of layer 1 excluding the channel (m/s)

$K_{Lay\ 2}$ hydraulic conductivity in layer 2 (m/s)

$K_{Lay\ 2\ left}$ left half hydraulic conductivity zone in layer 2 (m/s)

$K_{Lay\ 2\ right}$ right half hydraulic conductivity zone in layer 2 (m/s)

$K_{relative}$ relative hydraulic conductivity [-]

K_s saturated hydraulic conductivity [LT^{-1}]

$K_{saturated}$ saturated hydraulic conductivity [LT^{-1}]

K_V vertical hydraulic conductivity [LT^{-1}]

$K_{V\ North}$ northern half zonation of K_V (m/s)

$K_{V\ South}$ southern half zonation of K_V (m/s)

K-L Kullback-Leibler (information, distance, discrepancy, divergence, or number)

K_RB hydraulic conductivity of the riverbed (m/s)

l tortuosity [-]

LAI measured leaf area index [-]

ln natural logarithm

LSE Least Squares Estimation

$L(g_i\ |y)$ Likelihood of a model given data

L(p|n,k ;binomial) binomial likelihood function of parameter *p*, given data *n* and *k*

L(θ |x) likelihood function of parameters *θ* given data *x*

$L(\hat{\theta}\,|\,y)$ maximum likelihood function of the parameter estimators given data

m empirical parameter of the soil water retention function [-]

MFI2K Modflow Interface 2000

MLE Maximum Likelihood Estimation

MT3D Modular 3D Transport Model for Advection, Dispersion and Chemical Reactions Simulation in Groundwater

n empirical parameter of the soil water retention function [-]

n number of observations

ND number of observations

Ne Neuman criterion, it has to be smaller or equal to 0 for numerical stability

n_{eff} effective porosity of the medium

NP number of estimated parameters

NPR number of prior information

p a parameter named *p*

p parameter probability of success

Pe Peclet number

PEST inverse code for automated hydrologic calibration

p_i parameter *p* at current position *i*

p_i true probability distribution of outcome i

π Pi; mathematical constant π=3.14159…

p_{i+1} parameter *p* varied by a small amount $i+1$

p_{MLE} MLE Estimate of parameter *p*

PMWIN Processing Modflow Program

$p_{reference}$ reference parameter

PTFs pedotransfer functions

P0 pressure head below which roots start to extract water [L]

P0pt pressure head below which roots start to extract water at maximum rate [L]

$P(A)$ prior or marginal probability of A, regardless of information about B

$P(A')$ prior probability of an alternative to A

$P(A|B)$ posterior probability or conditional probability of A given B

$P(B)$ prior or marginal probability of B

$P(B|A)$ likelihood or conditional probability of B given A

$P(B|A')$ complement of the conditional probability of B given an alternative to A

$p(\hat{\theta})$ prior probability of the parameter θ evaluated at the estimated parameter $\hat{\theta}$

QAIC modified AIC for overdispersion

r number of variables *x* in a linear regression model

R number of models

RCH_1 recharge in zone 1 (cm/a)

RCH _2 recharge in zone 2 (cm/a)

RSS Residual Sum of Squares

RU seepage water [L]

s^2 error variance

σ standard deviation

σ^2 variance

$\sigma_{h,\ K,\ or\ C}$ respective standard deviation of observed heads, hydraulic conductivity and concentration data sets

$\hat{\sigma}^2_{ML}$ variance maximum likelihood estimate

$S'(b')$ maximum likelihood function

T potential transpiration

^{T} transpose matrix operator

tr matrix trace function

θ parameters of a linear regression $(\beta_0, \beta_1, ..., \beta_r, \sigma)$

θ water content [L^3L^{-3}]

$\hat{\theta}$ estimator of parameter θ

θ_e effective water content [-]

θ_r residual water content [-]

θ_s saturated water content [-]

TIC Takeuchi information criterion

μ mean value

UCODE_2005 inverse code for automated hydrologic calibration

USGS United States Geological Survey

v fluid velocity within the pores (seepage velocity), which is the Darcy velocity divided by n_{eff}

VK_CB hydraulic conductivity of the confining bed (m/s)

ω weight

w weight

w_i Akaike weights; can be interpreted as the probability that model i is the best model supported by the data

x spatial coordinate [L] (positive upward)

χ^2 goodness-of-fit chi-square statistic

y finite data assumed by Akaike to approximate the K-L number

$\hat{\bar{y}}$ averaged estimated result

$\hat{y}_i$ estimated result of each model i

$\hat{y}(p_i)$ calculated value obtained at current parameter position i

$\hat{y}(p_{i+1})$ calculated value obtained at parameter p varied by a small amount $i+1$

$\hat{y}(p_{reference})$ reference output value

$\hat{z}$ predicted system state variables in space-time

$\approx$ approximately equal to

$\sum$ Cramér-Rao lower bound of the covariance matrix of the p_{MLE}

$|\ |$ determinant of a matrix

$|\omega|$ determinant of a weight matrix with dimension [n x n]

^ symbol over a variable or constant meaning estimator; not the theoretical true value but an estimated or calculated value of it

∇h hydraulic pressure gradient

∞ infinity

! permutation

ABSTRACT

There is a great deal of uncertainty in hydrogeological modeling. Overparametrized models increase uncertainty since the information of the observations is distributed through all of the parameters. The present study proposes a new option to reduce this uncertainty. A way to achieve this goal is to select a model which provides good performance with as few calibrated parameters as possible (parsimonious model) and to calibrate it using many sources of information.

Akaike's Information Criterion (AIC), proposed by Hirotugu Akaike in 1973, is a statistic-probabilistic criterion based on the Information Theory, which allows us to select a parsimonious model. AIC formulates the problem of parsimonious model selection as an optimization problem across a set of proposed conceptual models. The AIC assessment is relatively new in groundwater modeling and it presents a challenge to apply it with different sources of observations.

In this dissertation, important findings in the application of AIC in hydrogeological modeling using different sources of observations are discussed. AIC is tested on groundwater models using three sets of synthetic data: hydraulic pressure, horizontal hydraulic conductivity, and tracer concentration. In the present study, the impact of the following factors is analyzed: number of observations, types of observations, and order of calibrated parameters. These

analyses reveal not only that the number of observations determine how complex a model can be but also that its diversity allows for further complexity in the parsimonious model. However, a truly parsimonious model was only achieved when the order of calibrated parameters was properly considered. This means that parameters which provide bigger improvements in model fit should be first considered.

The approach to obtain a parsimonious model applying AIC with different types of information was successfully applied to an unbiased lysimeter model using two different types of real data: evapotranspiration and seepage water. With this additional independent model assessment it was possible to underpin the general validity of this AIC approach.

ZUSAMMENFASSUNG

Hydrogeologische Modellierung ist von erheblicher Unsicherheit geprägt. Überparametrisierte Modelle erhöhen die Unsicherheit, da gemessene Informationen auf alle Parameter verteilt sind. Die vorliegende Arbeit schlägt einen neuen Ansatz vor, um diese Unsicherheit zu reduzieren. Eine Möglichkeit, um dieses Ziel zu erreichen, besteht darin, ein Modell auszuwählen, das ein gutes Ergebnis mit möglichst wenigen Parametern liefert („parsimonious model“), und es zu kalibrieren, indem viele Informationsquellen genutzt werden.

Das 1973 von Hirotugu Akaike vorgeschlagene Informationskriterium, bekannt als Akaike-Informationskriterium (engl. Akaike’s Information Criterion; AIC), ist ein statistisches Wahrscheinlichkeitskriterium basierend auf der Informationstheorie, welches die Auswahl eines Modells mit möglichst wenigen Parametern erlaubt. AIC formuliert das Problem der Entscheidung für ein gering parametrisiertes Modell als ein modellübergreifendes Optimierungsproblem. Die Anwendung von AIC in der Grundwassermodellierung ist relativ neu und stellt eine Herausforderung in der Anwendung verschiedener Messquellen dar.

In der vorliegenden Dissertation werden maßgebliche Forschungsergebnisse in der Anwendung des AIC in hydrogeologischer Modellierung unter Anwendung unterschiedlicher Messquellen diskutiert. AIC wird an Grundwassermodellen getestet, bei denen drei

synthetische Datensätze angewendet werden: Wasserstand, horizontale hydraulische Leitfähigkeit und Tracer-Konzentration. Die vorliegende Arbeit analysiert den Einfluss folgender Faktoren: Anzahl der Messungen, Arten der Messungen und Reihenfolge der kalibrierten Parameter. Diese Analysen machen nicht nur deutlich, dass die Anzahl der gemessenen Parameter die Komplexität eines Modells bestimmt, sondern auch, dass seine Diversität weitere Komplexität für gering parametrisierte Modelle erlaubt. Allerdings konnte ein solches Modell nur erreicht werden, wenn eine bestimmte Reihenfolge der kalibrierten Parameter berücksichtigt wurde. Folglich sollten zuerst jene Parameter in Betracht gezogen werden, die deutliche Verbesserungen in der Modellanpassung liefern.

Der Ansatz, ein gering parametrisiertes Modell durch die Anwendung des AIC mit unterschiedlichen Informationsarten zu erhalten, wurde erfolgreich auf einen Lysimeterstandort übertragen. Dabei wurden zwei unterschiedliche reale Messwertarten genutzt: Evapotranspiration und Sickerwasser. Mit Hilfe dieser weiteren, unabhängigen Modellbewertung konnte die Gültigkeit dieses AIC-Ansatzes gezeigt werden.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Rudolf Liedl who has supported me since the beginning with his constructive discussions and ideas. I would also like to thank Dr. Robert Schwarze and Dipl.-Hydrol. Verena Maleska from the TUD Institute of Hydrology and Meteorology for providing the lysimeter data used in Chapter 5.

Furthermore, I would like to thank my family for their support, particularly, my wife Alexandra and my mother Elizabeth for reading through the manuscript and making remarks which helped me to improve the readability of my dissertation.

The dissertation was financed with a CONACYT/DAAD scholarship granted by the National Council on Science and Technology of Mexico (CONACYT) in cooperation with the German Academic Exchange Service (DAAD) and with support of the Institute for Groundwater Management at the Technische Universität Dresden (TUD).

It is better to be roughly right than precisely wrong.

(John Maynard Keynes; English economist, 1883-1946)

1

INTRODUCTION

Hydrogeological modeling is a complicated science. Mathematicians, chemists, and engineers probably find modeling a very frustrating task due to the great deal of uncertainty which they are confronted with. There are uncertainties in the value of the parameters, scale chosen, data quality, boundary conditions, etc. At a given point these can be quite overwhelming and decrease confidence in the results. Modeling is not an exact science and plenty of compromises must be made. For example, the measured field values can not be directly used to parametrize the model and the same is true for the values from the laboratory. They have to be all "averaged" to account for wrong boundary conditions, a chosen scale, and unparametrized variables. Several attempts have been tried to reduce these uncertainties and one of those is the present dissertation. A way to reduce uncertainty could be to choose a model with as few calibrated parameters as possible and still having good performance (parsimonious model) and to use for calibration many sources of information (for example: heads, flows, concentration, etc.).

Akaike's Information Criterion (AIC) is a model selection criterion based on Information Theory which allows us to choose a parsimonious model. A recent tendency among hydrogeologists is to postulate several alternative hydrogeological models for a site and use a model selection criteria in order to: 1) rank the models, 2) eliminate some of them and/or, 3) weight and average predictions and statistics generated by the models (Ye et al. 2008).

Since the application of model selection criteria is relatively new in groundwater modeling, there is still some insecurity about the correct implementation of the model selection criteria. A challenge is always present when applying AIC and other model selection criteria in hydrogeological models, especially when using different sources of observations.

1.1 OBJECTIVE

The scientific premise in this dissertation is the application of Akaike's Information Criterion to hydrogeological models while using different sources of observations, such as head, hydraulic conductivity, and concentration values for calibration. If this criterion can be applied, then, it is necessary to determine the impact of the number and type of observations, as well as the order of the calibrated parameters. This is to my information still not achieved in the field of hydrogeological modeling, although, some attempts were already made by Hill and Tiedeman (2007a). Akaike's Information Criterion is rarely applied in hydrogeological modeling. The same is true for calibrating with different information types. The combination of these two practices is not well developed yet and I believe that this combination will substantially reduce uncertainty in hydrogeological modeling.

1.2 STRUCTURE

Chapter 1 presents the justification of the study, the research objective and gives a short overview of the whole dissertation.

In Chapter 2, the scientific foundations, assumptions, and formulas involved in the Information Theory are explained.

Chapter 3 describes the Model Selection theory in depth. At the beginning, an overview of model selection methods is provided, followed by a detailed explanation of the AIC method, assumptions, formulas, and its most important modifications. There is also a discussion of whether the model selection methods AIC, AIC_c, BIC, and KIC were correctly derivated assuming the truth was not included within the candidate models. Furthermore, the results of the comparison of model selection methods in hydrogeology are mentioned. Finally, the implementation problems in hydrogeology of model selection methods are presented and discussed.

In Chapter 4, the combination of applying AIC to groundwater models and calibrating them with different types of observations as a practical method to reduce conceptual model

uncertainty is tested using synthetic data. The way to norm the different data types is revealed. The model that generates the synthetic data is presented, along with the candidate models proposed to simulate the data. This is followed by the set up and evaluation of three experiments with synthetic data. These experiments test the application of AIC in groundwater models using first, only hydraulic pressure observations for calibration; second, calibrating additionally with the conductivity observations; and third, making use of all three data sets, including the concentration values. A discussion follows about the impact of number of observations, type of observations, and the influence of parameter order on the model selection ranking. The discussion provides important insights into the optimal way to set up the model selection evaluation according to the amount and types of observations available.

Chapter 5 applies the method proposed in the previous chapter to a real case using two different types of observations from a lysimeter. To this end, the order of the parameters was optimally arranged to obtain the best possible AIC value.

Chapter 6 provides the closure to the research by presenting the conclusions and discussing the implications of using the combination of applying AIC to hydrogeological models and calibrating the models using different types of observations as an innovative method to reduce model uncertainty. The areas where further research is needed are also identified.

Finally, important information about basic mathematical subjects which are essential to understand AIC are annexed at the end, along with relevant model selection topics related to the present dissertation.

2

INFORMATION THEORY

Information Theory is a science that began in the Second World War. Codified information was sent to allies and intercepted by enemies. Efforts on both sides were made to decode this information. Under these conditions, Kullback and Leibler (1951) found out the mathematical expression which defines information loss between a model and the “truth” (parametrized truth or truth-model). This expression is known as the Kullback-Leibler (K-L) information, also known as K-L divergence, relative entropy, or I-divergence (Seghouane and Amari 2007). The mathematical efforts to codify, send, and decode information constitute the Information Theory. The main applications are found on the field of telecommunication, which includes transmitting information (audio, images, data files, etc.) through different channels, but the discovery of Kullback and Leibler can be applied to other fields such as modeling.

2.1 HOW TO MEASURE INFORMATION LOSS

K-L information (Information loss) can indeed be measured and it is defined for continuous functions as the integral:

$$I(f,g) = \int f(x) \ln \frac{f(x)}{g(x \mid \theta)} dx \qquad 1$$

where $I(f,g)$ denotes the information lost when g is used to approximate f (Burnham and Anderson 2002:51); $g(x|\theta)$ is a model which is expected to provide an approximation of the truth; while $f(x)$ is the representation of the truth; and θ stands for model parameters.

We can precisely calculate this loss of information with the previous equation as shown in the following illustrative example modified from Burnham and Anderson (2002:54). Let the true model $f(x)$ be the probability density function of a gamma distribution with two parameters (α = 4, β = 4), defined as

$$f(x) = \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \qquad 2$$

and consider three different probability density functions $g(x|\theta)$, Lognormal, Inverse Gaussian, and F distribution, denoted g_1, g_2, and g_3 respectively, each with two parameters, as approximation models, shown below:

$$g_1(x \mid \theta, \sigma) = \frac{e^{\frac{-(\ln x-\theta)^2}{2\sigma^2}}}{x\sigma\sqrt{2\pi}} \qquad 3$$

$$g_2(x \mid \alpha, \beta) = \sqrt{\frac{\beta}{2\pi x^3}}\, e^{\frac{-\beta(x-\alpha)^2}{2\alpha^2 x}} \qquad 4$$

$$g_3(x \mid \alpha, \beta) = \frac{\Gamma\left[\frac{\alpha+\beta}{2}\right]}{\Gamma\left(\frac{\alpha}{2}\right)\Gamma\left(\frac{\beta}{2}\right)}\left(\frac{\alpha}{\beta}\right)^{\frac{\alpha}{2}} \frac{x^{\frac{\alpha-2}{2}}}{\left[1+\left(\frac{\alpha}{\beta}\right)x\right]^{\frac{\alpha+\beta}{2}}} \qquad 5$$

These functions are substituted in Equation 1 and evaluated in the range [0, 50]. The one with the minimum value has the smallest information loss, also interpreted as the shortest K-L "distance," which means the closest approximation to $f(x)$. Burnham and Anderson (2002:52) note that according to Kullback (1959) the K-L information could be seen as the direct "distance" from g (approximation) to f (truth), but not as the distance from f to g since it is an oriented distance in which both directions are different[1]. The three models and their corresponding "distances" to $f(x)$ are listed in the following table:

[1] Seghouane and Amari (2007) symmetrized the K-L divergence using the geometric and harmonic mean and apply it as a base for a model selection criteria (AIC_c*, more on this subject in Section 3.4.3).

Table 1: Information loss between $f(x)$ Gamma distribution (α = 4, β = 4) and three approximating models g_i

	Approximating model	$I(f, g_i)$
g_1	Lognormal distribution (θ = 2, $\sigma = \sqrt{2}$)	0.672
g_2	Inverse Gaussian (α = 16, β = 64)	0.060
g_3	F distribution (α = 4, β = 10)	5.745

The closest approximation to $f(x)$ is the probability density function g_2, followed by g_1 and far behind by g_3. It can also be visually verified (for this simple example) by plotting the three functions against the gamma distribution as it is presented in Figure 1.

In this example, we know exactly the truth and its parameters; what in practice seldom happens. This makes out of the K-L divergence a theoretical formula which in the real praxis is not possible to be calculated.

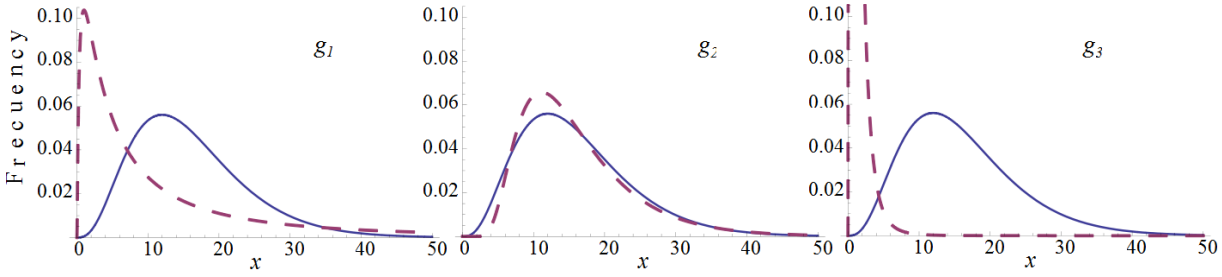

Figure 1: Plots of truth $f(x)$ gamma distribution (4, 4) (solid line), against three $g(x)$ approximating models (dashed lines): g_1 = lognormal ($2, \sqrt{2}$), g_2 = inverse Gaussian (16, 64), and g_3 = F distribution (4, 10)

If one is familiarized with the entropy concept, then one could notice that the K-L information and the entropy are related. Burnham and Anderson (2002:54, 86) note that Boltzmann (1877) derive the theorem that entropy is proportional to minus logarithm of the probability. Hence, Kullback and Leibler derive a measure that in fact happens to be the negative of Boltzmann's entropy:

$$\textit{Boltzmann's entropy} = -\ln\frac{f(x)}{g(x)} \qquad 6$$

The K-L information [$I(f,g)$] is the expectation with respect to f of minus Boltzmann's entropy:

$$I(f,g) = \mathrm{E}_f\left(\ln\frac{f(x)}{g(x)}\right) = \int f(x)\ln\frac{f(x)}{g(x)}dx \qquad 7$$

Hence, maximizing the entropy equals minimizing the K-L distance. Burnham and Anderson (2002:86) explain that entropy means “disorder,” so by maximizing the entropy we get the maximum of disorder, which means minimum of information. It is, however, the “noise” which is disordered at its maximum and forced to a minimum of information. A good model has the information on the data, leaving only “noise” as a residual. Burnham and Anderson summarize that this leaves a model which is maximally justified by the data and with a minimum of noise when the K-L distance is minimized.

At least theoretically, the information loss between the model and the truth is known. It seems as if the K-L information would not be useful in practice since the truth is generally unknown and therefore not quantifiable. The K-L information is, though, relevant since Akaike (1973) found an estimation to the K-L information when the truth is unknown.

2.2 HOW TO ESTIMATE INFORMATION LOSS WHEN THE TRUTH IS UNKNOWN

The K-L information expressed in extended form is the following:

$$I(f,g) = \int f(x)\ln[f(x)]dx - \int f(x)\ln[g(x\mid\theta)]dx \qquad 8$$

Clearly, every term at the right is an expectation with respect to $f(x)$. The expectation of the logarithm of the truth with respect to the truth is not of interest since it is always constant. Therefore, we can take it to the other side as a constant C as follows:

$$I(f,g) - C = -\int f(x)\ln[g(x\mid\theta)]dx \qquad 9$$

The left hand side term is no more the exact value of the K-L information but a numerical value of the K-L information minus a constant which depends on the unknown truth. Calculating the right hand side will not give us the exact information loss, it will be just a value. However, this numerical value is useful since the truth remains constant for a specific situation and can be compared to those from other approximating models. Akaike (1973) found a way to estimate the right hand term when $f(x)$ is unknown based on the maximum likelihood function. An estimate of relative information loss can be obtained. Since the x is

integrated out and there is no data on the above expression, Akaike assumed then some finite y data to estimate the parameters θ as follows:

$$I(f,g) - C = -\int f(y)\left(\int f(x)\ln[g(x \mid \hat{\theta}(y))]dx\right)dy \qquad 10$$

At this point, we introduce the mathematical notation of estimators "^." The parameters θ are not known, they must be estimated $\hat{\theta}$. A term with a "hat" is not the theoretical true value but an estimated or calculated value of it. Akaike found out that the expectation of x and y from $\ln[g(x \mid \hat{\theta}(y)]$ with respect to the truth could be estimated by substituting the unknown truth with the maximum likelihood estimate which is asymptotically chi squared distributed when considering large samples and "good" approximating models (close to the truth) [2]. This gives (Burnham and Anderson 2002:353, 364):

$$I(f,g) - C \approx \ln[L(\hat{\theta} \mid y)] - K \qquad 11$$

where $\ln[L(\hat{\theta} \mid y)]$ is the natural logarithm of the maximum likelihood function of the parameter estimators $\hat{\theta}$ given some data. It is not a problem to solve the likelihood function since it can be found in mathematical handbooks for different distributions. Akaike determined that the bias introduced by substituting the maximum likelihood estimate is approximately equal to K, which is the number of unknown estimable parameters. These parameters can be hydrological as well as statistical.

We have seen that information loss between the truth and an approximating model can be measured and that Akaike's approach can estimate the information loss when the truth is unknown by relying on sample data of that truth. The relation found by Akaike between the K-L information and the maximum likelihood estimate has allowed major practical and theoretical advantages in model selection and the analysis of complex data sets (Burnham and Anderson 2002:261, deLeeuw 1992). In the next chapter, we will explore the applications of Akaike's approximation in Model Selection.

---

[2] Detail information on the estimation of $\ln[g(x \mid \hat{\theta}(y)]$ in Akaike (1973), Sawa (1978), Linhart and Zucchini (1986:41-47, 76-78, 243-245), deLeeuw (1992), Konishi and Kitagawa (1996), and Burnham and Anderson (2002:Chapter 7).

3

MODEL SELECTION METHODS

In this chapter, an overview of the variety of model selection methods found in the literature is presented. However, we will focus on the Akaike Information Criterion and AIC_c, the corrected version for small data samples. These two methods will be explained in detail, while other related selection models such as QAIC, $QAIC_c$, AIC_c*, BIC, and KIC will be just shortly introduced. Then, a literature discussion will be presented about doubts of proper mathematical derivation of AIC and BIC. This will be followed by a literature comparison of the performance of model selection methods in hydrogeology. And to conclude, the implementation problems faced by our hydrogeology colleagues will be presented, focusing on the implementation problem resulting from considering different kind of observations.

3.1 MODEL SELECTION METHODS IN THE LITERATURE

If data is available, one could always find several models that fit equally well the data. For example, given two points, we could fit beside a line also any curve. Without making any further assumptions, there is no reason to prefer a model over other one. In other words, we are forced to make assumptions to fit the data with a model. Several model selection methods have been proposed in order to enable us to choose wisely among the different possibilities. Just to give an overview, a list of methods classified as theoretical or empirical, modified from Sewell (2008) are listed below:

Theoretical methods:

- AIC_c (Sugiura 1978, Hurvich and Tsai 1989);
- AIC_c* (Seghouane and Amari 2007);
- AICi (Bengtsson and Cavanaugh 2006);
- AICW (Wilks 1995);
- Akaike Information Criterion (AIC, Akaike 1973);
- Bayesian Information Criterion (BIC, also Schwarz Criterion, Schwarz Information Criterion (SIC), or Schwarz-Bayesian Information Criterion; Schwarz 1978);
- CAT (Parzen 1974, 1977);
- Consistent AIC with Fisher Information (CAICF, Bozdogan 1987);
- Deviance Information Criterion (DIC, Spiegelhalter et al. 2002);
- FIC (Wei 1992);
- Final prediction error (FPE, Akaike 1969);
- FPEα (Bhansali and Downham 1977);
- FPEC (de Luna 1998);
- FPER (Larsen and Hansen 1994);
- Gamma Test (Han et al. 2009);
- Generalized cross-validation (GCV, Craven and Wahba 1979);
- Generalized prediction error (GPE; Moody 1991, 1992);
- GM (Geweke and Meese 1981);
- Hannan and Quinn Criterion (HQ, Hannan and Quinn 1979);
- ICOMP(IFIM) Inverse Fisher Information Matrix (Bozdogan 2000);
- Informational Complexity Criterion (ICOMP, Bozdogan and Haughton 1998);
- KIC (Kashyap 1982);
- KICc (Cavanaugh 2004);
- Mallows' Cp (Cp, Mallows 1973);
- Minimum description length (MDL, Rissanen 1978);
- Minimum message length (MML, Wallace and Boulton 1968);
- Predicted squared error (PSE, Barron 1984);
- PRESS (Allen 1974);
- QAIC, $QAIC_c$ (Lebreton et al. 1992);
- Structural risk minimization (SRM, Vapnik and Chervonenkis 1974);
- Takeuchi's information criterion (TIC, also Generalized AIC (GAIC) or AIC_T; Takeuchi 1976); and
- VC-dimension (Vapnik and Chervonenkis 1968, 1971; Vapnik 1979).

Empirical methods:

- Adjusted R^2 (Wherry 1931);
- Cross-validation (Stone 1974, Geisser 1975);
  - Bootstrap (Efron 1979)
  - Jacknife
  - K-fold cross-validation
  - Leave-one-out cross-validation
- Linear regression;
- Shibata's model selector (sms, Shibata 1981);
- Signal-to-noise ratio; and
- Test set validation.

If one goes into the details of every single method, one would get easily lost in such a diversity of model selection methods. The literature is highly technical and dispersed throughout research articles and books. Therefore, it is not my intent to explain every single method, but to focus on AIC and briefly mention some other methods, especially, those which have been used to evaluate and rank models in hydrogeology.

3.2 PRINCIPLE OF PARSIMONY

In numerical modeling context, the principle of parsimony means that models are favoured for fitting the observed values and penalized for building additional model complexity. The principle of parsimony is one of the main principles of many selection methods and stands for "as less as possible." It is also known as Occam's razor: *It is vain to do with more what can be done with fewer,* William of Occam (Grünwald 2000:133). An increase in estimated parameters means an increase in complexity. However, the addition of a parameter may result on substantial improvement of model fit which would compensate for the penalty. If the improvement does not compensate for the penalty, model selection criteria do not rank the model higher than the competing models. In this manner, model selection criteria choose the parsimonious model.

3.3 AKAIKE'S INFORMATION CRITERION (AIC, AKAIKE 1973): MODEL SELECTION BASED ON THE ESTIMATION OF THE INFORMATION LOSS

Akaike defined a model selection criterion called Akaike's Information Criterion (AIC) based on his estimation of the information loss between an approximating model and an unknown

parametrized truth. With this criterion, Akaike could estimate a "relative" information loss. AIC can distinguish if an approximating model loses less information than another one, and thus, identify the one which is closer to the unknown truth which generated the data. He achieved this by multiplying by -2, due to historical reasons, the unbiased estimation of the K-L information, $\ln[L(\hat{\theta} \mid y)] - K$, and defining AIC as follows:

$$AIC = -2\ln[L(\hat{\theta} \mid y)] + 2K \qquad 12$$

One could think that these "historical reasons" have no justification and could be left out. However, deLeeuw (1992) adds that although the justification in Akaike's paper for using -2 may seem weak, the result is a natural distance measure between probability densities. He also mentions that this result has strong connections with the Shannon-Wiener information criterion, the Fisher information, and entropy measures used in thermodynamics. Burnham and Anderson (2002:64) emphasize that -2 times the natural logarithm of the ratio of two maximized likelihood values is asymptotically chi-squared distributed under certain conditions and assumptions. Thus, the term -2 is not unreasonable.

AIC, as given in Equation 12, cannot be directly entered into the calculator. First, one should substitute the corresponding likelihood function, which depends on the distribution of the data. In hydrology, most annual hydrologic events are described by the normal distribution (Reddy 2005:58, 63). When analysing model results, we are dealing mostly with normal distributed residuals. Residuals are a kind of experimental error obtained by subtracting the model results from the observed data. Therefore, the general assumptions for errors apply: *One expects them to be (roughly) normal and (approximately) independently distributed with a mean of 0 and some constant variance* (NIST/SEMATECH 2003:Section 5.2.4). The normal distribution, also called Gaussian or Bell Curve, is probably the most useful distribution. The Central Limit Theorem states that with increasing number of observations, data which is roughly normal, not necessarily independent, to some extent random, and with finite variance, tends to a normal distribution[3] (Brutsaert 2005:524). This means that as *n* increases, our data will tend to be normally distributed even if we relax the assumptions of independence, randomness, and normal distribution of the observations. Independence means that the value of an observation does not depend on the values of previous observations. The likelihood function of the normal distribution function is

[3] Appropriate transformations to make non normal data approximately normal may be needed.

$$L(\hat{\theta} \mid y; normal) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_{ML}^2}} \right)^n e^{-\frac{1}{2}n} \quad 13$$

where $\hat{\sigma}_{ML}^2$ is defined as

$$\hat{\sigma}_{ML}^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \quad 14$$

The notation stands for $\hat{\sigma}_{ML}^2$: maximum likelihood variance estimate, ε: residual (observed minus calculated value), and *n*: number of observations. More information about the maximum likelihood estimation of the normal distribution can be found in Appendix A.2 (p. 91).

Inserting the just mentioned likelihood function into the general AIC formula (Equation 12) will provide us with an AIC equation for normally distributed residuals which can be easily computed with a calculator:

$$AIC = n\ln(\hat{\sigma}_{ML}^2) + n\ln(2\pi) + n + 2K \quad 15$$

where K is the number of estimable parameters. If σ^2 of the data is known, K equals the number of hydrological parameters. However, if σ^2 is unknown, as is often the case, K equals the number of hydrologic parameters plus one (Ye et al. 2008). This is because the unknown statistical parameter σ^2 has to be approximated by Maximum Likelihood to obtain the estimate $\hat{\sigma}_{ML}^2$.

Since the terms $n\ln(2\pi)$ and n are constant, they do not influence model selection and can be dropped, leaving us with a practical Gaussian AIC formula:

$$AIC = n\ln(\hat{\sigma}_{ML}^2) + 2K \quad 16$$

AIC can be regarded as a balance equation between closeness to the truth and model complexity. Assuming that there is a set of candidate models with varying complexity, the optimal model would be the one with the smallest AIC, as noted by the black triangle with the smallest value in Figure 2. The term $n\ln(\hat{\sigma}_{ML}^2)$ is a measure of model fit, indicated by the blue diamonds in the same figure. Usually, the model fit improves substantially from non-calibrated models to models calibrated with increasing number of parameters. However, the improvement is not linear. Frequently, allowing more parameters to be calibrated results in little improvement. On the other hand, the term $2K$, indicated in the figure as red squares,

penalizes model complexity in linear form. Adding this last term to the model fit term gives us the value of AIC. This proves to be a very simple criterion to identify parsimonious models.

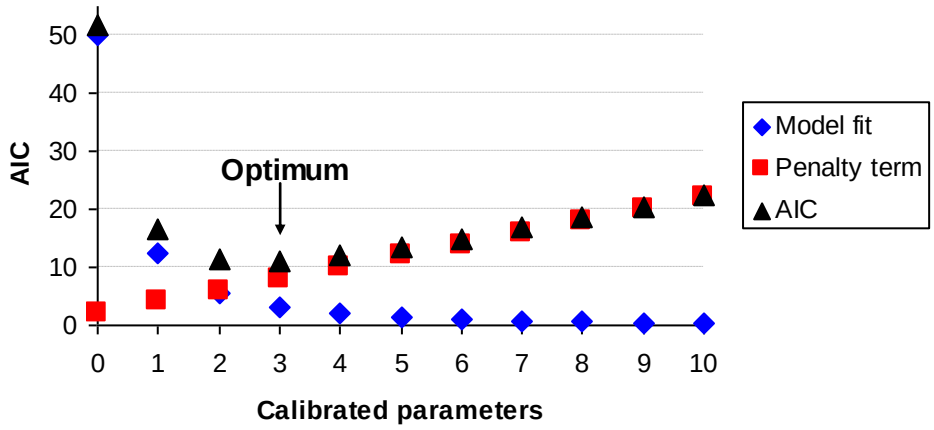

Figure 2: Schematic example of AIC principle

Burnham and Anderson (2004) suggest taking the Akaike differences to rank the models:

$$\Delta_i = AIC_i - AIC_{\min} \qquad 17$$

where Δ_i is the difference between each AIC value and the minimum AIC value of the set. The best model will have a $\Delta = 0$ while the others will show positive numbers. As a rough guideline, one can consider that models characterized by $\Delta_i \leq 2$ have support to a large extent; $4 \leq \Delta_i \leq 7$ have considerably less support; and $\Delta_i > 10$ have basically no support (Burnham and Anderson 2004).

The Akaike differences allow us to easily calculate the likelihood of a model given the data. According to Burnham and Anderson (2004), the likelihood of the model given the data is

$$L(g_i|y) = \exp\left(-\frac{\Delta_i}{2}\right) \qquad 18$$

It should be recalled that Akaike multiplied the unbiased estimation of the K-L information by -2 to obtain AIC; otherwise, the likelihood would be simply:

$$L(g_i|y) = \exp(\Delta_i) \qquad 19$$

The likelihood of the model is important since the Akaike weights w_i are defined by normalizing the likelihoods such that they all sum 1 (Burnham and Anderson 2004):

$$w_i = \frac{\exp\left(-\frac{\Delta_i}{2}\right)}{\sum_{i=1}^{R}\exp\left(-\frac{\Delta_i}{2}\right)}, \qquad \text{where } i = 1, 2, \ldots, R \text{ (number of models).} \qquad 20$$

The w_i results can be interpreted as the probability that model i is the best model supported by the data.

3.4 AIC MODIFICATIONS (AIC_C, QAIC, $QAIC_C$, AND AIC_C*)

AIC found widespread acceptance and received immediate attention from the scientific community. A part of this community searched applications for AIC while another one tried to find a “better model selection criterion.” From these last efforts, several modifications of AIC were developed. It seems that there is no best model selection criterion for all cases. Instead, a certain model works best for a determined case.

3.4.1 Small sample correction: AIC_c (Sugiura 1978, Hurvich and Tsai 1989)

A correction to the bias, which Akaike found out to be approximately equal to 2K, was made by Sugiura (1978), and applied by Hurvich and Tsai (1989) for the case of having few observations:

$$AIC_c = -2\ln[L(\hat{\theta}\mid y)] + 2K\left(\frac{n}{n-K-1}\right) \qquad 21$$

in Gaussian form:

$$AIC_c = n\ln(\hat{\sigma}^2_{ML}) + n\ln(2\pi) + n + 2K\left(\frac{n}{n-K-1}\right) \qquad 22$$

dropping out the constant terms we get

$$AIC_c = n\ln(\hat{\sigma}^2_{ML}) + 2K\left(\frac{n}{n-K-1}\right) \qquad 23$$

It could also be expressed in terms of AIC as presented by Burnham and Anderson (1999):

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1} \qquad 24$$

As the ratio of number of observations to number of estimated parameters (n / K) decreases, the bias of the estimated K-L information in AIC increases (Ye et al. 2008). Therefore, Burnham and Anderson (2002:66) suggest using AIC_c when n / K < 40 for the candidate model with the most of parameters. For the opposite case, when the number of observations increases with respect to the estimable parameters, AIC_c tends to AIC.

The AIC_c is probably the most useful modification of AIC, since modelers often have few observations to work with. The following example illustrates the AIC_c. Suppose we have the following truth:

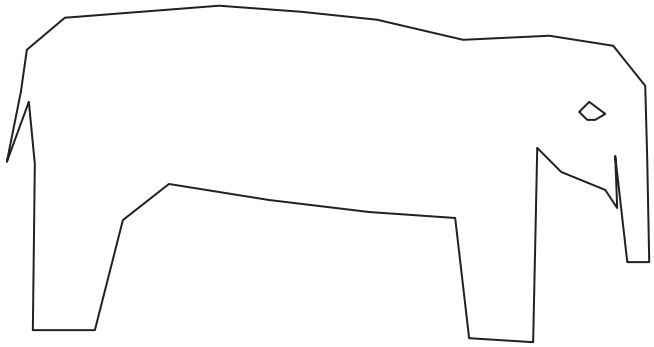

Figure 3: Elephant drawing (Wel 1975)

The elephant was defined by Wel (1975) with 36 points and simulated with four different models using the least square Fourier sine series (Equations 25 and 26) with 5, 10, 20 and 30 terms. The results are shown in the following figure:

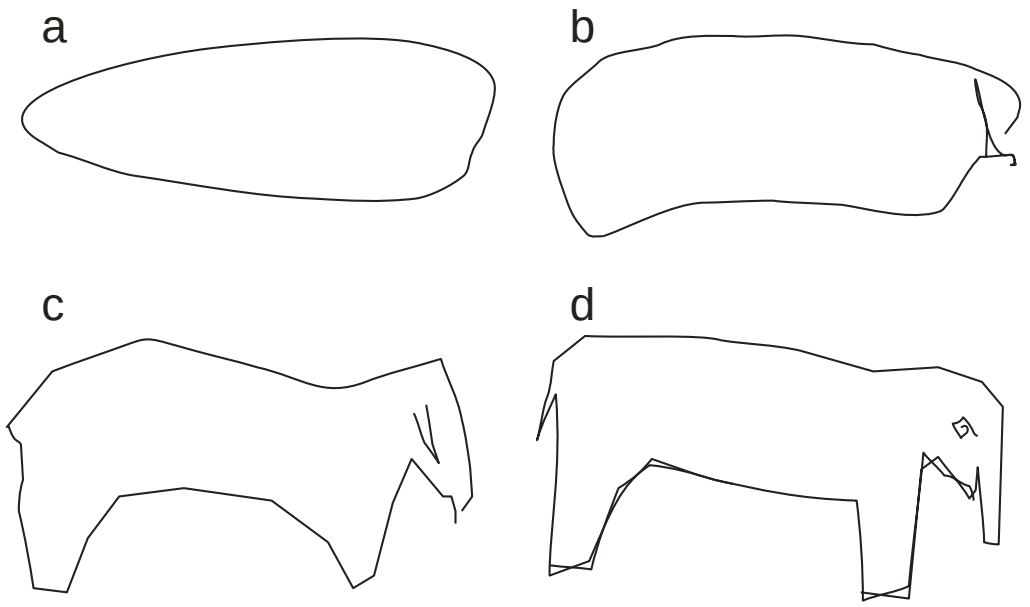

Figure 4: Elephant fit using the least square Fourier sine series with 5 (a), 10 (b), 20 (c), and 30 (d) terms (Wel 1975)

From the above figure we can tell that with more terms, we obtain a better fit. And that model "d" seems to be a very good approximation of the elephant. But, was the effort of adding so many parameters worthy? How can this be quantified? To answer this, we implement AIC_c.

The 36 points which Wel used to define his elephant and to calibrate the four models are located at the line inflections of the elephant model:

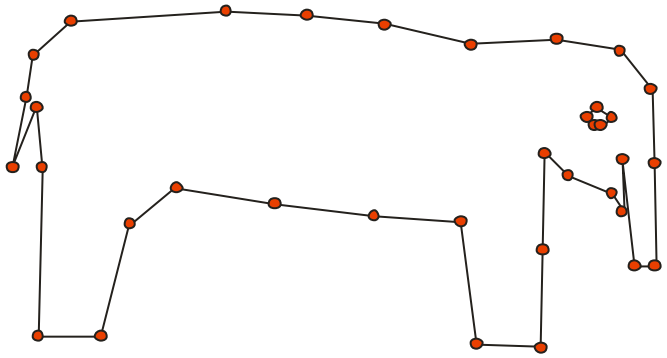

Figure 5: Elephant data

The x and y coordinates of the points were parametrized by Wel in terms of a variable *t*, which assumes integer values from 1 to 36. The following least square Fourier sine series were used to fit the data:

$$x(t) = a_0 + \sum_i a_i \sin \frac{it\pi}{36} \qquad (25)$$

$$y(t) = b_0 + \sum_i b_i \sin \frac{it\pi}{36} \qquad 26$$

By overlaying the data (Figure 5) to the model results shown in Figure 4, we obtain the following “original data – approximating models” superpositions:

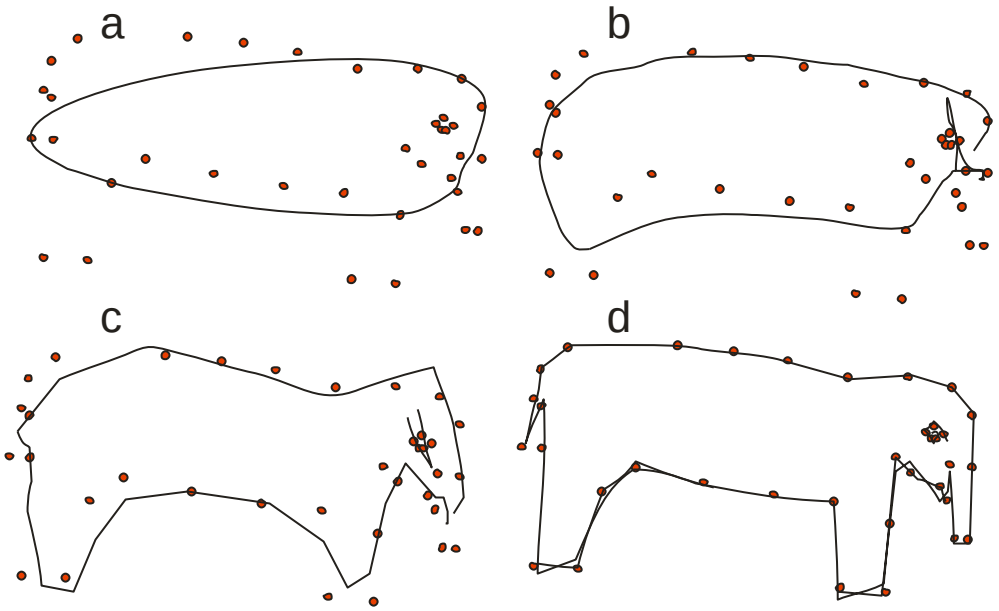

Figure 6: Elephant point data and fit of the least square Fourier sine series with 5 (a), 10 (b), 20 (c), and 30 (d) terms

The shortest distance of the points to the fitted curves in the above figure is taken as the residual. With the residuals, we can calculate the model fit (first term of Equation 23). K can be calculated as the number of parameters (terms of least square Fourier sine series) + 1 statistical parameter ($\hat{\sigma}^2_{ML}$). Furthermore, the Akaike differences ($AIC_c\,\Delta_i$) are the differences of the respective model to the best model of the set. These differences are calculated analogously to those of AIC Δ_i with Equation 17 (p. 14). The Akaike weights ($AIC_c\ w_i$) are calculated with Equation 20 (p. 15). The results are summarized in the following table and visualized in the following graph:

Table 2: AIC_c evaluation of least square Fourier sine series with 5 (a), 10 (b), 20 (c), and 30 (d) terms

Model	K	Model Fit	AIC_c Penalty	AIC_c	$AIC_c\ \Delta_i$	$AIC_c\ w_i$
a	6	91.8	14.9	106.7	2.4	0.24
b	11	71.4	33.0	104.4	0.0	0.76
c	21	18.3	108.0	126.3	22.0	0.00
d	31	-129.7	558.0	428.3	323.9	0.00

n = 36

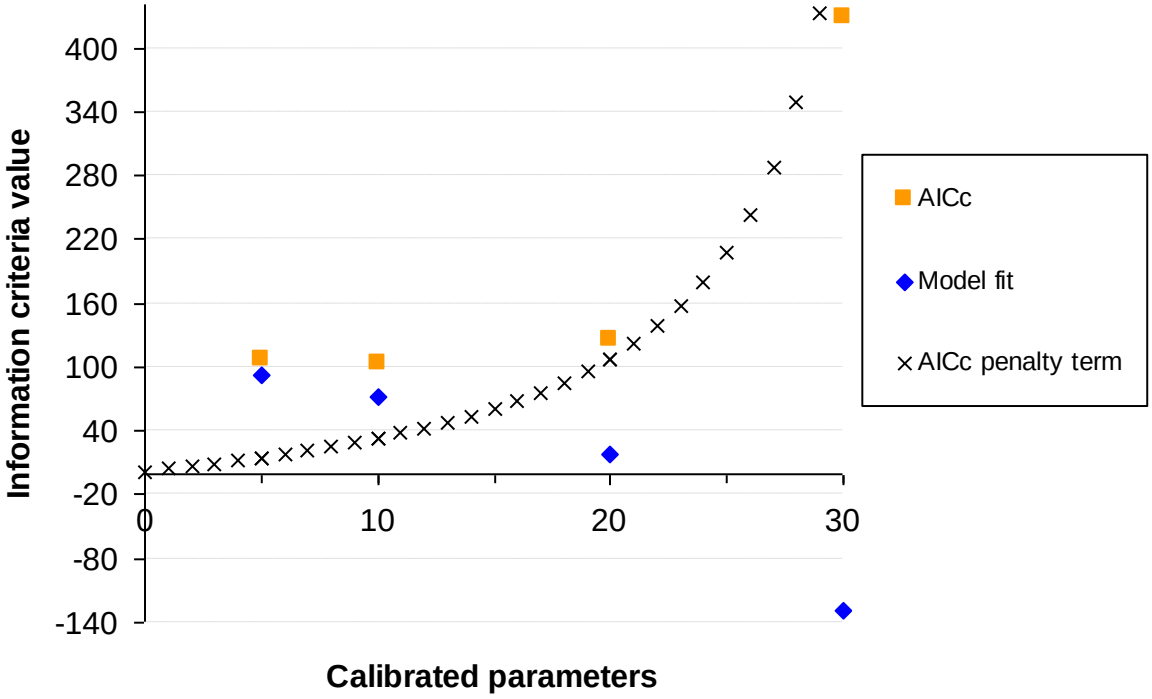

Figure 7: AIC_c evaluation of the four elephant candidate models

The ratio n / K of the model with more parameters (Model "d") is 36 / 31 = 1.2. This value is much lower than 40 and therefore, the set of models is to be analized by AIC_c. The bias of the estimated K-L information is too high to apply AIC.

In the previous graph, one can see that the model fit improves from model to model. This was already evident in Figure 6. Not evident was that the penalty term increases as in Figure 7. The penalty term is the second term of Equation 23 (p. 15). The form of the penalty term curve depends on the number of observations. The less observations, the more exponential the curve behaves; the more observations, the flatter it is (AIC_c tends to AIC). For this data set with 36 observations we obtain almost an exponential penalty term curve with a limit at 34 calibrated parameters. There, K = 35 and $n - K - 1 = 0$ (division by zero is not defined). The mathematical limit of the penalty term is the number of observations available minus three (since with minus two we have already the division by zero). In practice one should calibrate with far less parameters than available observations.

Out of the four models, the best approximation is Model "b," with the lowest AIC_c value (Figure 7). This is the best compromise between data availability and model complexity. It has approximately 75 % probability of being the best model, as indicated by the Akaike weights (AIC_c w_i) in Table 2. Model "d," has almost as much parameters as data is available.

Such a model is clearly overfitted. This means that given another set of data, for example, by choosing different points along the elephant of Figure 5, Model “d” will most probably provide a very different result to that in Figure 6 d. In contrast, Model “b” results will be more consistent and still look similar to the result of Figure 6 b. Model “a” is slightly underfitted with approx. 25 % probability of being the best model. Model “c” is overfitted and has no support from the Akaike weights.

With this exercise we do not just illustrate AIC_c but also ratify the common English expression that *given enough parameters one can fit an elephant*, or rather: overfit an elephant.

3.4.2 Modified criteria for overdispersion in count data: QAIC and $QAIC_c$ (Lebreton et al. 1992)

When count data sampling variance differs from the predicted Poisson or multinomial probability distributions, count data is said to be overdispersed (Burnham and Anderson 2001). This modification of AIC is probably not relevant in hydrogeological modeling since we do not work with count data. However, it is certainly of relevance for ecological population models where AIC has been mostly applied. An appropriate modification of AIC and AIC_c to take overdispersion into account was found by Lebreton et al. (1992):

$$QAIC = -\frac{2\ln L(\hat{\theta}|y)}{\hat{c}} + 2K \qquad 27$$

$$QAIC_c = -\frac{2\ln L(\hat{\theta}|y)}{\hat{c}} + 2K\left(\frac{n}{n-K-1}\right) \qquad 28$$

where

$$\hat{c} = \frac{\chi^2}{df} \qquad 29$$

and where χ^2 is the goodness-of-fit chi-square statistic with df degrees of freedom, which are the additional number of parameters needed to specify the most complex model fitted to the available data (Richards 2008). Burnham and Anderson (2001) note that only one value of $\hat{c}$ should be used along a set of candidate models. They add that a value of $\hat{c}$ lower than 1 should not be used and when c is estimated, it counts as a parameter and should be included in K. $\hat{c}$ is an estimated variance inflation factor which should be estimated from the

highest dimensioned model in the set of candidates (Burnham and Anderson 1999). When no overdispersion occurs c = 1 and QAIC is equivalent to AIC.

3.4.3 Symmetric criterion: AIC_c* (Seghouane and Amari 2007)

As noted in Section 2.1, the K-L information on which the AIC is based is an oriented "distance" where both directions are different. Seghouane and Bekara (2004) found that the K-L distance may reflect errors due to overfitting while the alternative direction may reflect errors due to underfitting. They argue that symmetrizing the K-L information would provide an AIC modification that provides a better balance of model disparity than any of the both directions.

By applying the mean average to the K-L Information, Seghouane and Amari (2007) obtain a symmetric AIC_c criterion called AIC_c*:

$$AIC_c^* \approx -2\ln[L(\hat{\theta} \mid y)] + \frac{(K+1)(2n-K-2)}{n-K-2} + \frac{K}{n-K} \qquad 30$$

Seghouane and Amari (2007) symmetrized AIC_c also using the geometric and the harmonic mean, but they are not sure about the benefits of using one or the other mean. They conclude that using a symmetrized AIC_c criterion could be preferable to the normal AIC_c, but more research is needed on the properties of the different criteria resulting by applying different average methods in order to make some further suggestions about them.

However, their formula is not based on Equation 21 but on a version of this AIC_c equation in which K is added 1. They claim to have it also from Hurvich and Tsai (1989):

$$AIC_{c,\,Seghouaneand\,Amari} = -2\ln[L(\hat{\theta} \mid y)] + 2(K+1)\left(\frac{n}{n-(K+1)-1}\right) \qquad 31$$

3.5 BAYESIAN INFORMATION CRITERION (BIC, SCHWARZ 1978) AND KASHYAP INFORMATION CRITERION (KIC, KASHYAP 1982)

BIC is after AIC one of the most popular model selection methods. It was developed by Schwarz (1978) under a Bayesian context as an asymptotic approximation of a transformation of the posterior probability of a candidate model (Ye et al. 2008). Background

information on Bayesian inference is presented in Appendix A.3 (p. 93). Even though, the derivation of BIC sounds complicated, the result is as simple as AIC:

$$BIC = -2\ln[L(\hat{\theta} \mid y)] + K \ln n \qquad 32$$

In the same manner as with AIC, the terms with $-2\ln[L(\hat{\theta}|y)]$ are measures of goodness of fit; the smaller the term the better the fit between predicted and observed values. Terms containing K are measures of model complexity. The goodness of fit term is the same for AIC as for BIC; the different model selection criteria differ only in the approximation of the bias.

KIC was also developed under a Bayesian context by Kashyap (1982) as an asymptotic approximation to the model likelihood (Ye et al. 2008):

$$KIC = -2\ln[L(\hat{\theta} \mid y)] - 2\ln\, p(\hat{\theta}) + K\ln(n/2\pi) - K\ln n + \ln|F| \qquad 33$$

which can be simplified to

$$KIC = -2\ln[L(\hat{\theta} \mid y)] - 2\ln\, p(\hat{\theta}) - K\ln(2\pi) - \ln|\Sigma| \qquad 34$$

where $p(\hat{\theta})$ is the prior probability of the parameter θ evaluated at the estimated parameter value $\hat{\theta}$. The term $-2\ln\, p(\hat{\theta})$ drops out when no previous information of the parameters is available; the two vertical lines | | denote the determinant of a matrix; F is the observed Fisher information matrix which is usually interchanged by $\langle F \rangle$, the expected Fisher information matrix having the following elements (Ye et al. 2008):

$$\langle F_{ij} \rangle = -E\left[\frac{\partial^2 \ln[L(\theta|y)]}{\partial\theta_i \partial\theta_j}\right]_{\theta=\hat{\theta}} \qquad 35$$

In Equation 34, $|\sum|$ denotes the Cramér-Rao lower bound of the covariance matrix of the Maximum Likelihood parameter estimates which is equal to the inverse of the determinant of the expected Fisher information matrix (Ye et al. 2008):

$$|\Sigma| = \frac{1}{|\langle F_{ij} \rangle|} \qquad 36$$

AIC_c and KIC provide a better approximation for the bias than AIC and BIC when only few observations are present. With increasing number of observations, AIC_c tends to AIC and KIC to BIC (Ye et al. 2008).

As noted by Ye et al. (2008), dropping out constant terms, the Gaussian forms of BIC and KIC are as follows:

$$BIC = n\ln(\hat{\sigma}^2_{ML}) + K\ln n \tag{37}$$

$$KIC = n\ln(\hat{\sigma}^2_{ML}) - 2\ln\, p(\hat{\theta}) - K\ln(2\pi) - \ln|\Sigma| \tag{38}$$

When the prior information about the hydrological parameters is not available, the term $-2\ln\, p(\hat{\theta})$ drops out and

$$\ln|\Sigma| = K\ln(\hat{\sigma}^2_{ML}) - \ln\left|J^T wJ\right| \tag{39}$$

where J stands for the Jacobian (sensitivity) matrix having elements:

$$J_{ij} = -\left.\frac{\partial \hat{z}_i}{\partial \theta_j}\right|_{\theta=\hat{\theta}} \tag{40}$$

and $\hat{z}$ are the predicted system state variables in space-time; ^{T} is the Transpose matrix operator and w stands for a weight.

KIC can be then expressed as:

$$KIC = (n-K)\ln(\hat{\sigma}^2_{ML}) - K\ln(2\pi) + \ln\left|J^T wJ\right| \tag{41}$$

Now that the topic model selection has received more attention in hydrogeology, Neuman (2003) proposes to apply KIC in hydrogeological modeling. This proposal was implemented by his colleagues (Ye et al. 2008, 2005, 2004, and Meyer et al. 2007). Ye et al. (2008) note that in KIC there are unique terms which not only make possible to evaluate models based on the number of observations and parameters but also based on their quality. On the other hand, KIC has also a major drawback in comparison to AIC_c. The terms that make KIC so special, like the Fisher information matrix, are not easy to be calculated. This contrasts to AIC_c, which is very easy to calculate and to apply. AIC_c seems to incorporate the parsimony

principle itself. This contributes in a great part to its popularity, especially outside the field of hydrogeology where AIC and AIC_c have been mainly applied.

3.6 DISCUSSION ABOUT THE TRUE MODEL NOT BEING PRESENT IN THE SET OF CANDIDATE MODELS FOR CORRECT DERIVATION OF AIC AND BIC

AIC and therefore, also AIC_c were derived assuming the true model was NOT in the set of candidate models (Poeter and Anderson 2005). However, Ye et al. (2008) examined all rigorous deviations of AIC they could find and concluded that the “true model” which generated the observed data needs to be present for a correct mathematical derivation. On the other hand, BIC and therefore also KIC (which tends to BIC as number of observations become large relative to estimated parameters) were derived assuming that the true model IS present in the set of candidate models. However, Cavanaugh and Neath (1999) rederived BIC without assuming that the true model is included in the set of candidate models (Ye et al. 2008, Burnham and Anderson 2002:293). Takeuchi (1976) derived also an information criterion (TIC) without assuming the truth to be in the set of candidate models (more about this criterion in Appendix B.1). According to Shibata (1989) and Burnham and Anderson (2002:65), AIC is the best estimator of TIC. Ye et al. (2008) do not object the TIC derivation.

The debate has been a main issue for model selection in hydrogeology. Model selection criteria such as AIC, AIC_c, BIC, and KIC may perform similarly in applications (Ye et al. 2008, Poeter and Anderson 2005). Therefore, presenting a solid theoretical argument such as “a derivation without assuming the truth to be present in the set of models” would be a good argument to justify the use of a specific selection method. The assumption of including the true model in the set of candidate models for derivation purposes is neither easy to prove nor to deny, since the mathematics to derivate the equations are not simple. Poeter and Anderson (2005) prefer to use AIC_c (which tends to AIC as observations increase) instead of BIC and KIC. They cite primarily Burnham and Anderson (2002) and argue that even if the model selection methods may perform similarly, AIC_c was derived assuming the truth was not considered in the set of models while for the derivation of the other methods this was not the case. The group of Neuman (Ye et al. 2008) prefers to use KIC (which tends to BIC as observations increase) instead of AIC and AIC_c. They argue that even if model selection methods perform similarly they found out that contrary to Poeter and Anderson (2005) and Burnham and Anderson (2002), AIC must, in contrast to BIC, indeed assume that the truth is in the set of models.

In summary, there is an agreement that BIC and TIC do not actually require the "true model" to be present in the set of candidate models for derivation. It also seems reasonable that KIC does not require either to include the "true model" since it is an asymptotic limit of BIC. It also seems convincing that AIC does not need to include the "true model" since it is a reliable approximation of TIC when a good performing model is included in the set of candidate models. Consequently, neither AIC nor BIC assume the truth to be included in the set of candidate models.

3.7 COMPARISON OF MODEL SELECTION METHODS IN HYDROGEOLOGY

As mentioned in the previous section, Ye et al. (2008) and Poeter and Anderson (2005) agree that model selection criteria such as AIC, AIC_c, BIC, and KIC may perform similarly in applications. Ye et al. (2008) carried out a comparison between the predictive performance of the model selection methods AIC, AIC_c, BIC, and KIC. They used real hydrological data and cross-validation to test the prediction performance[4]. From the four methods compared by Ye et al. (2008), KIC showed a better performance. However, they noticed that the difference between model averaging predictions[5] for the different methods was much smaller than the difference between any individual best model results, independent from model selection method. As a result, they conclude that the decision to apply model averaging is more important than the choice of the model selection method. Also, Burnham and Anderson (2004) note that model averaging produces better results than either AIC or BIC under all circumstances.

Foglia et al. (2007) also compared AIC_c, BIC, and GCV[6] with Cross Validation (CV) using hydrogeological data and concluded that model selection methods are as good as the computationally expensive cross validation test.

---

[4] Details on their applied cross-validation method and model averaging procedures in Ye et al. (2004). For a general explanation of cross-validation methods see Browne (2000).

[5] Information on "multimodal inference", also known as "model averaging" is given in Appendix B.2.

[6] In Foglia et al. (2007) defined with a logarithmic transformation to compare it with the other selection methods as follows: $n \ln GCV = n \ln \sigma^2_{ML} - 2n[\ln(n-K) - \ln n]$. The original version without the logarithmic transformation is (Regonda et al. 2005):

$$GCV = \frac{\sum_{i=1}^{n} \frac{\varepsilon_i^2}{n}}{\left(1 - \frac{K}{n}\right)^2}$$

3.8 IMPLEMENTATION PROBLEMS OF MODEL SELECTION METHODS IN HYDROGEOLOGY

Since the late eighties, model selection methods have been sporadically applied in hydrogeology. However, implementation problems have occurred. According to Ye et al. (2008), Carrera and Neuman (1986) applied KIC to a hydrogeology problem but the term $-K\ln n$, a penalty term, was missing in their equation. This omission continued with Hernández et al. (2006) and with Poeter and Anderson (2005) which also seem to miss the term $-K\ln\hat{\sigma}^2_{ML}$ due to a wrong specification of the Fisher information matrix (Ye et al. 2008).

Carrera and Neuman (1986) proposed including the term $-2\ln p(\hat{\theta})$ which appears in KIC also into the leading term $-2\ln[L(\hat{\theta}\,|\,y)]$ of AIC and BIC. However, Ye et al. (2008) explain that the term $-2\ln p(\hat{\theta})$ drops out of KIC as the number of observations increase and therefore, should not be included into BIC, AIC, or AIC_c as allowed by Hill (1998), Hill and Tiedeman (2007a), and Poeter and Hill (2007). Ye et al. (2008) conclude that the term $-2\ln p(\hat{\theta})$ is a unique feature of KIC.

In a serious attempt to apply AIC to groundwater models calibrated with different kinds of observations (hence, with different kinds of units), Hill (1998) normalized the model residuals with the help of weights (ω) making them unitless. However, the weight matrix is defined such that the common error variance of the residuals equals one. This solves the unit problem, but prevents the use of the AIC formula as defined by Akaike (Equation 16, p. 13). The leading term of the Gaussian form of AIC, $n\ln(\hat{\sigma}^2_{ML})$, drops out ($n\ln 1=0$). The formula used by Hill (Equation 42) is not the AIC any more as it is claimed, since the maximum likelihood term was dropped through an inconvenient mathematical decision which stripped the AIC of its maximum-likelihood goodness-of-fit term.

$$AIC_{Hill\ and\ Tiedeman} = n\ln(2\pi) - \ln|\omega| + e^T\omega e + 2K \qquad 42$$

This formula is implemented in MODFLOW-2000 (Harbaugh et al. 2000, Hill et al. 2000) and UCODE_2005 (Poeter et al. 2005)[7]. Hill's (1998) approach is described in Appendix B.3 (p. 97). The AIC implementation problems of Hill and Tiedeman[8] are discussed below in detail

[7] MODFLOW-2000 and UCODE_2005 are open source, public domain computer programs developed by the USGS. The first one solves the 3-D groundwater flow equation using a control-volume finite-difference numerical method (Hill and Tiedeman 2007a:19). UCODE_2005 performs general inverse modeling and can be used to calibrate models, perform sensitivity and uncertainty analysis.

[8] Hill (1998) publication was superseded by the book of Hill and Tiedeman (2007a) which includes not just all the information in Hill (1998) but it is also reviewed and embedded in a wider modeling context.

with the help of two exercises from Hill and Tiedeman's (2007a) book. Exercise 6.1 is about analyzing a groundwater model previously created using synthetic data in exercise 5.2c. Both exercises serve as a practical example of Hill and Tiedeman's method to obtain the value of AIC for different types of observations. Hill and Tiedeman provide detailed instructions on how to create this groundwater model (exercise 5.2c) in their book's web site[9]. This information includes the complete input files for the model, as well as answers to the specific exercise's questions. The project file is EX5.2C.NAM and the output files can be easily calculated. The input data can be also verified through the program interface, MFI2K (Harbaugh 2002). The software is freely available at the USGS website[10]. For the following exercise, MODFLOW-2000 latest version: 1.18.01 (20th June, 2008) was downloaded in June 2009. The main output files are EX5.2C.GLO and EX5.2C.LST. Here we can verify the observations, the estimated parameters, the residuals, and the AIC value. The calculated value of AIC is -5.6713, found near the bottom of file EX5.2C.GLO. The way in which Hill and Tiedeman come to this result is explained below.

The data which Hill and Tiedeman provide consist of 13 values: 10 head values (m); 1 flow value (m^3/s) equal to the groundwater discharge to a river reach (the reach extends along the entire length of the river); 1 hydraulic conductivity value of the riverbed (m/s); and 1 value of vertical hydraulic conductivity of the confining bed (m/s).

Enough information to solve the first term of Hill and Tiedeman's AIC formula (Equation 42):

$$13 \ln 2\pi = 23.89 \qquad 43$$

To solve the next term, we need to determine which weights are used for each observation data type. To make the data unitless Hill and Tiedeman use weights. The weights are the inverse of an assumed measurement error variance. For this exercise, Hill and Tiedeman (2007a:38) assumed a measurement error variance of 1.0025 m^2 (1 m^2 well elevation measurement error + 0.0025 m^2 water-level measurement error) for head values. The weight is defined as the inverse of the variance: 0.9975 m^{-2}.

For the flow data, an error measurement of 0.10 coefficient of variation was assumed (Hill and Tiedeman 2007c). By multiplying the coefficient of variation by the flow measurement we obtain the standard deviation, $\sigma = 0.10$ x (-4.4 m^3/d). The variance is the square of this last one, $\sigma^2 = 0.1936$ m^6/d^2. Taking the inverse gives us the weight 5.165 d^2/m^6.

[9] http://wwwbrr.cr.usgs.gov/projects/GW_ModUncert/hill_tiedeman_book/ (accessed February 2008).

[10] http://water.usgs.gov/software / (accessed June 2009).

A coefficient of variation of 0.30 was assumed to calculate the weights of both hydraulic conductivity parameters: of the riverbed (K_RB in m/s) and the confining bed (VK_CB in m/s). The weights are calculated in the same manner as previously carried out with the flow data:

$$K\_RB_{weight} = \frac{1}{(0.3 \times (1.2 \times 10^{-3} m/s))^2} = 7.716 \times 10^6 m^{-2} s^{-2} \quad 44$$

$$VK\_CB_{weight} = \frac{1}{(0.3 \times (1.0 \times 10^{-7} m/s))^2} = 1.111 \times 10^{15} m^{-2} s^{-2} \quad 45$$

Hence, the 2nd term of Equation 42 is calculated as follows:

$$\ln|\omega| = \ln(0.9975^{10} \times 5.165^1 \times (7.716 \times 10^{-6})^1 \times (1.111 \times 10^{15})^1) = 52.12 \quad 46$$

In Equation 46, the units are omitted since the natural log is dimensionless (Hill and Tiedeman 2007c).

The third term of Equation 42 is equivalent to the weighted residual sum of squares (RSS) which is equal to 10.561, as shown in the next table at the bottom-right corner:

Table 3: RSS obtained from Hill and Tiedeman's (2007a) exercises

Name	Observed	Calculated	Residual	(Weight)$^{1/2}$	(Weighted residual)^2
head01	101.804 m	100.210 m	1.594 m	0.999 m^{-1}	2.536 (-)
head02	128.117 m	126.942 m	1.175 m	0.999 m^{-1}	1.378 (-)
head03	156.678 m	157.185 m	-0.507 m	0.999 m^{-1}	0.256 (-)
head04	124.893 m	126.942 m	-2.049 m	0.999 m^{-1}	4.190 (-)
head05	140.961 m	141.076 m	-0.115 m	0.999 m^{-1}	0.013 (-)
head06	126.537 m	127.090 m	-0.553 m	0.999 m^{-1}	0.305 (-)
head07	101.112 m	100.934 m	0.178 m	0.999 m^{-1}	0.032 (-)
head08	158.135 m	157.141 m	0.994 m	0.999 m^{-1}	0.987 (-)
head09	176.374 m	176.650 m	-0.276 m	0.999 m^{-1}	0.076 (-)
head10	142.020 m	141.137 m	0.883 m	0.999 m^{-1}	0.779 (-)
flow01	-4.400 m^3/d	-4.417 m^3/d	1.71E-02 m^3/d	2.27 d/m^3	0.002 (-)
K_RB	1.20E-03 m/s	1.17E-03 m/s	3.08E-05 m/s	2.78E+03 s/m	0.007 (-)
VK_CB	1.00E-07 m/s	9.90E-08 m/s	1.02E-09 m/s	3.33E+07 s/m	0.001 (-)
∑					**10.561** (-)

It should be noticed that in Table 3, the residuals are not multiplied by the weights but by the square roots of the weights. This operation is necessary since the variances do not have the same units as the residuals. Multiplying the residuals by the inverse of the assumed standard deviation provides unitless weighted residuals.

Finally, the last term in Equation 42 is simply two times the number of calibrated parameters, $2K$. The parameters being estimated are six, listed on the following table:

Table 4: Estimable parameters for exercise 5.2C from Hill and Tiedeman (2007a)

Parameter	Initial value	Reasonable lower limit	Reasonable upper limit
HK_1	3.00E-04	3.00E-05	3.00E-03
HK_2	4.00E-05	4.00E-06	4.00E-04
VK_CB	1.00E-07	1.00E-08	1.00E-06
K_RB	1.20E-03	1.20E-04	1.20E-02
RCH_1	63.072	32	126
RCH_2	31.536	16	63

MODFLOW-2000 calculates $2K$ as 2×6. In this case, there are no statistical parameters to be calculated, since it is assumed unit standard deviation ($\sigma = 1$). As explained in Section 3.3, K equals then just the number of hydrological parameters.

Substituting all terms in Equation 42 we obtain what is calculated in UCODE_2005 and MODFLOW-2000:

$$AIC_{Hill\ and\ Tiedeman} = 23.89 - 52.12 + 10.56 + 6\times 2 = -5.67 \qquad 47$$

In exercise 6.1b, Hill and Tiedeman propose a simple exercise to demonstrate that the expected value of both the calculated error variance and the standard error is 1. With Excel one can generate a list of random numbers with the function "norminv(rand(),0,σ)" and then, calculate the error variance with the equation below:

$$s^2 = \frac{\sum_{i=1}^{n} \frac{a_i^2}{\sigma^2}}{n} \qquad 48$$

where s^2 is the error variance, a_i^2 represents square residuals, which in this exercise are squared random numbers generated with a given variance. $1/\sigma^2$ represents the weights,

which are 1 over the variance of data measurement error, for this exercise the variance of the random numbers distribution.

The result of the previous calculation (Exercise 6.1b) is indeed one. However, in groundwater modeling, **the variance of the residuals distribution depends on the model and reflects a fit to the conceptual model and not a fit to measurement error**. This concept is ignored by Hill and Tiedeman (2007a). In consequence, one cannot assume unit variance. This in turn means that the leading term of the Akaike equation should not be dropped. Accordingly, the penalty term should take into consideration an extra statistical parameter, the variance. It should also be emphasized that contrary to Akaike's Equation 16 (p. 13), in Hill and Tiedeman's underlying AIC formula (dropping constant terms, Equation 96, p. 98) there is no influence of the number of observations since n is not present. Last but not least, as mentioned in Section 3.3, there is no need to retain constant values which remain the same for all models, as it is the case in Equation 42 (p. 26), which is calculated by MODFLOW-2000 and UCODE_2005.

This chapter provided an overview of model selection methods. It presented in detail the theory, formulas and principles of the AIC and AIC_c model selection methods; including a practical example of how the method works using an elephant as a model. Furthermore, it also presented AIC modifications and the model selection methods BIC and KIC. The current debate in the literature about the assumption for the correct mathematical derivation of AIC and BIC that the truth is not in the set of candidate models was also discussed. Also studies about comparison of these four model selection methods were presented. The chapter concluded by presenting implementation problems, in particular those faced by Hill and Tiedeman (2007a) when using different types of observations for calibration.

The next chapter presents a suggestion for a proper implementation of AIC and AIC_c considering different types of observations using synthetic data.

4

MODEL ASSESSMENT WITH DIFFERENT TYPES OF OBSERVATIONS USING SYNTHETIC DATA

Another approach to implement AIC to models calibrated with different kind of observations is proposed in this dissertation. However, theory by itself is not enough in applied sciences. We need to demonstrate that the theory works with data. Hence, this chapter will present an approach to apply Akaike's Information Criterion to models calibrated with synthetic data using different types of hydrogeological information. To accomplish this, a groundwater playground for model assessment was created. The playground consists of synthetic data and several sets of models of varying complexity which were calibrated to these data.

4.1 NORMING DIFFERENT KINDS OF OBSERVATIONS

When evaluating model performance with a model selection method, we are normally interested in the estimated parameters and the residuals. The number of calibrated parameters is related to model complexity, while the residuals are the values which evaluate the model fit to the data. Residuals result from substracting observed "real" values which we want to reproduce, to "model-produced" or "model-given" values which the model calculates or needs as parameters. Small residuals mean that "real" and "model" values are similar. Residuals are the basis of model calibration. The calibration procedure consists basically in changing the values of the parameters until the residuals are minimized. Usually, different

information is available for model calibration and it would be valuable to make use of it. However, different types of observations have usually different types of units. This poses a mathematical dilemma when attempting to analyse a data set of mixed elements. Since we cannot add apples and oranges, we have to transform them such that all of them have either the same units, or even better, no units. Furthermore, the values lay on different order of magnitudes and have to be evaluated together as a single data set. To transform them so that they are unitless and have similar order of magnitudes is not an easy task. Several methods were tested for this purpose:

- observations of a given type were divided by their respective mean, maximum, or minimum value;
- simulated values were divided by observed values;
- observed and simulated values were scaled first such that the observed values equal 1 and then normed by the scaled simulated values;
- the observed and simulated values are normed by their respective observed standard deviation.

The best method proved to be the last mentioned. The previous ones may solve the unit problem, but not the difference of magnitude issue. The procedure to norm the data is explained schematically in the figure below.

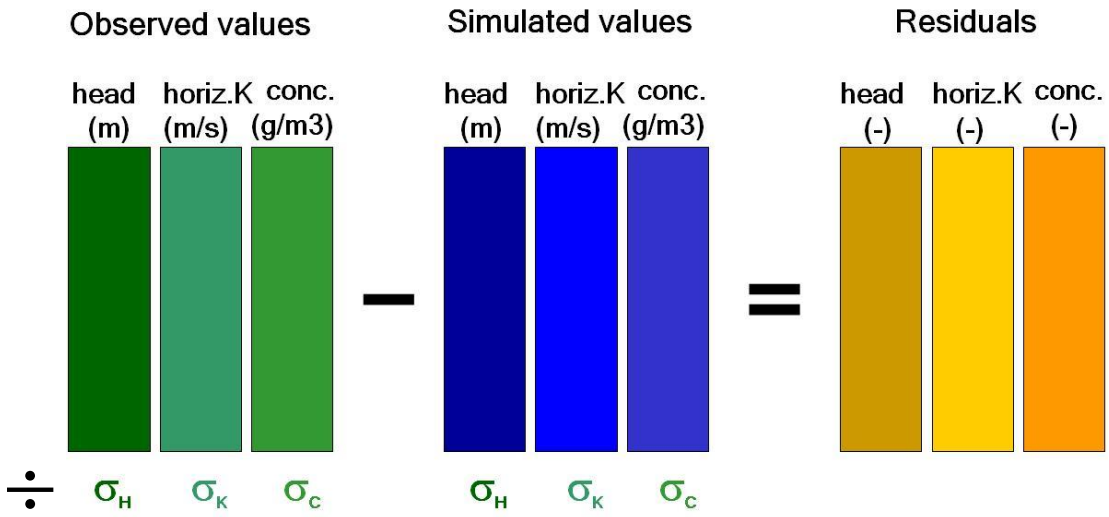

Figure 8: Schematic example of norming procedure of observed and simulated values

Three different observed data sets are represented in different green tones: head (m), horizontal hydraulic conductivity (m/s), and concentration (g/m^3). The hydraulic conductivity and concentration values have to be log transformed (base 10) in order to have, as the head residuals, a normal distribution instead of a logarithmic one. The arithmetic mean of each set

of observations is taken to center the data sets at zero and obtain their respective standard deviation value ($\sigma_{h,\ K,\ or\ C}$, which has the same units as the data). The observed values are divided then by their standard deviations (normalized). A practical example using synthetic data of this norming procedure can be found in Appendix B.5 (p. 101). The blue sets of simulated head, log K, and log concentration values are also normed with the same observed (green) standard deviations and not with the simulated standard deviations. In this manner, observed and simulated values are consistently normed and provide unitless residuals of the same order of magnitude. They can be finally evaluated as a single data set to calculate AIC (Equation 16, p. 13) or AIC_c (Equation 23, p. 15). In order to estimate the maximum likelihood variance (Equation 14, p. 13) the set is centered at zero. However, the norming mass can not be always estimated. If there is not enough data to calculate a representative standard deviation for a data type, then it has to be assumed. In this case, an appropriate estimation will depend on the experience of the modeler.

4.2 SYNTHETIC DATA GENERATING MODEL

In order to obtain space and time distributed synthetic data, it is first necessary to create the spatial-temporal environment for which a virtual model has to be built and parametrized. To make a realistic environment, heterogeneity has to be added. The advantage of such a model is that it does not need to be calibrated since it represents an ideal situation. Its objective is to be a realistic model to provide data.

For this purpose, I created a two layered, steady-state groundwater model using the PMWIN program (Processing Modflow) version 5.3.0 (Chiang and Kinzelbach 1991-2001), a free pre- and postprocessing tool for modeling algorithms such as MODFLOW, MODPATH, and MT3D. This version of PMWIN comes with MODFLOW 96, which was applied in this dissertation to simulate the groundwater flow system.

4.2.1 Conceptual model and boundary conditions

The model is designed such that water enters the system at a constant head boundary and as groundwater recharge. The upper layer is unconfined while the lower is confined. After flowing through the layers, groundwater leaves the system at a well and a river boundary. The boundary conditions, illustrated in Figure 9, were defined as in the following table:

Table 5: Boundary conditions of data generating model

Boundary	Condition
Right side of domain	No flow
Left side of domain	No flow
Top of domain	Constant recharge
Bottom of domain	No flow
Back of layer 1	Constant head
Back of layer 2	Constant head
Front of layer 1	River condition
Front of layer 2	No flow

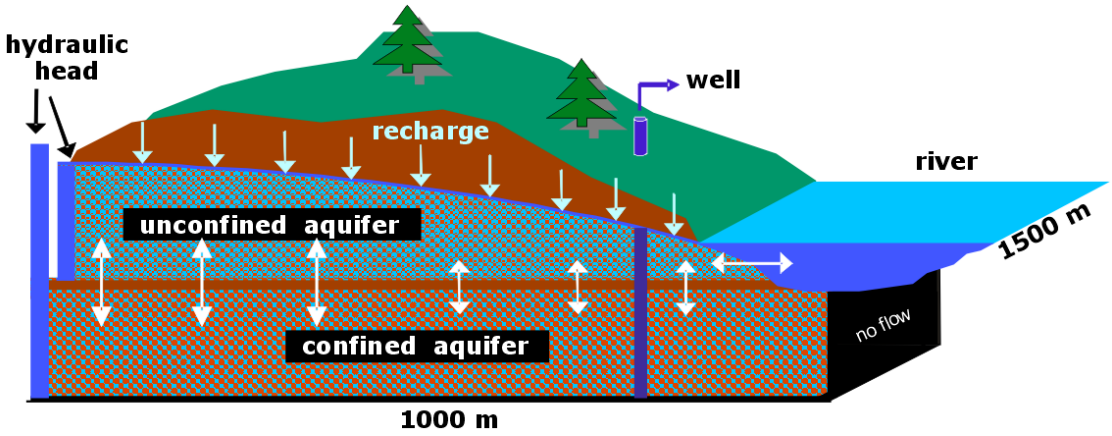

Figure 9: Conceptual model and boundary conditions of data generating model (modified from Liedl 2007)

4.2.2 Geometry

The geometric features can be summarized in the following table:

Table 6: Geometric parameters to generate the model

Attributes	Value
Number of layers	2
Number of rows	100
Number of columns	150
Number of cells	30 000
Cell size	10 x 10 m
Domain length	1000 m
Domain width	1500 m
Top of layer 1	40 m
Top of layer 2	20 m
Bottom of layer 2	0 m

4.2.3 Initial heads

The values of the constant head boundary of layer 1 are evenly spaced: 37.68 m to the west, 37.63 m at the center, and 37.58 m to the east (not shown in figure above). For layer 2, the

value remains constant at 39.7 m. Due to this hydraulic pressure difference, water from the lower layer tries to flow into the upper one. Initial heads were set iteratively for the cells not marked as constant heads. First, 37 m was assigned to those cells in layers 1 and 2. This value is the mean between layer 1 constant head boundary and average river height. Then, after being fully parametrized, the model was executed and the resulting heads were used as initial head values.

4.2.4 Recharge and hydraulic conductivity fields

To make the model realistic, heterogeneity was built in. For the recharge, plausible average values were taken from the literature. A graph of water percolation depending on soil type and precipitation for a typical middle European basin (temperate climate) in Hölting (1992:59) served as reference. Furthermore, hydraulic conductivity of layer 1 and groundwater recharge were correlated. In the same area where the highest recharge values are found, we also find the highest conductivity values. The same is true for lower and middle values. This can be easily appreciated in the following figure comparing the highest recharge zones (dark blue areas) of the recharge distribution (upper right image) with the highest hydraulic conductivity zones (white areas) of the hydraulic conductivity distribution of layer 1 (upper left image).

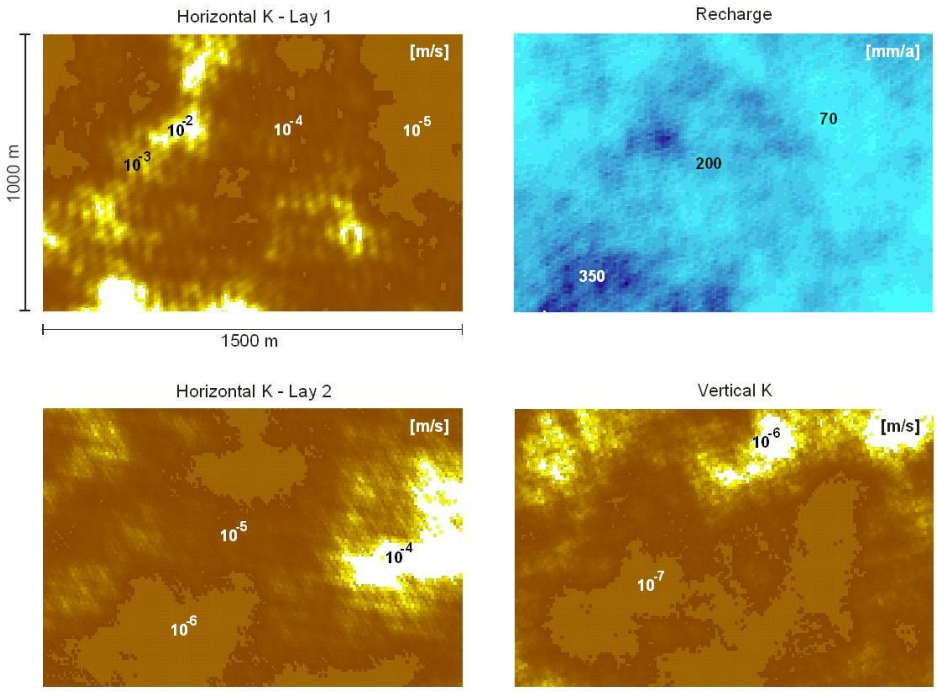

Figure 10: Hydraulic conductivity and groundwater recharge heterogeneity distributions

The values of hydraulic conductivity shown in Figure 10 lay on the order of magnitude which is found in unconsolidated rocks, see the following table:

Table 7: Typical hydraulic conductivity values (Hölting 1992:104)

Material	Hydraulic conductivity, K (m/s)
Pure gravel	10^{-1} - 10^{-2}
Coarse grained sand	10^{-3}
Medium grained sand	10^{-3} - 10^{-4}
Fine grained sand	10^{-4} - 10^{-5}
Silty sand	10^{-5} - 10^{-7}
Clayey silt	10^{-6} - 10^{-9}
Clay	< 10^{-9}

Conductivities according to	DIN 18130, Tl.1 (m/s)
Very high conductivity	> 10^{-2}
High conductivity	10^{-2} - 10^{-4}
Permeable	10^{-4} - 10^{-6}
Low conductivity	10^{-6} - 10^{-8}
Very low conductivity	< 10^{-8}

4.2.5 Generation of heterogeneity

The hydraulic conductivity and groundwater recharge heterogeneity were built using the Field Generator (Frenzel 1995) which comes with PMWIN.

For the user, the Field Generator consists of a single dialog window where the values of seven parameters should be given. Figure 11 shows the values used to create the heterogeneous recharge field.

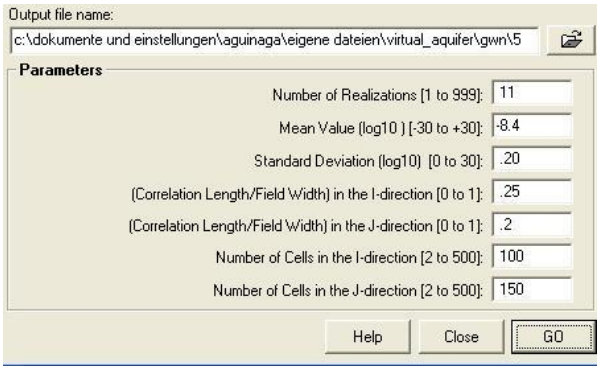

Figure 11: The Field Generator dialog window with parameter values used to generate the heterogeneous groundwater recharge field

The number of created fields (realizations) can be set high for the purpose of stochastic modeling. For my case, I just need one realization. Eleven realizations where made for the generation of recharge distribution, but just one of them was selected. Moreover, since the mean should be given in log10, the value -8.4 represents a mean of $10^{-8.4}$ (m/s) which is approx. 125 mm/a. The standard deviation value was chosen to allow variations of one order of magnitude. Other parameters are ratios, like the “Correlation Length/Field Width” in I- or J-

direction. The I-direction represents the Y axis, for this model 1000 m length (100 cells of 10 m width). The J-direction is the X axis, in this case 1500 m. Therefore, a ratio of 0.25 in the I-direction means 250 m / 1000 m. A ratio of 0.20 in the J-direction is 300 m / 1500 m. The ratio of "Correlation Length / Field Width" in I-direction to J-direction of the recharge field is 1.25 = 25/20. The same ratio, but for the hydraulic conductivity field of layer one is 1.3158 = 25/19. The previous two ratios are similar, but intentionally not the same. This similarity gives a touch of realistic correlation between both sets.

To create the three heterogeneous fields of hydraulic conductivity with the Field Generator, the following parameters were used:

Table 8: Field Generator parameter values to generate hydraulic conductivity heterogeneity fields

Field Generator Parameters	K_h Layer 1	K_h Layer 2	K_V
Number of Realizations	3	24	3
Mean Value (log10)	-3.09691	-5.04576	-6.22185
Standard Deviation (log10)	0.2	0.2	0.4
Correlation Length / Field Width in the I-direction	0.5	0.5	0.25
Correlation Length / Field Width in the J-direction	0.38	0.38	0.20
Number of Cells in the I-direction	100	100	100
Number of Cells in the J-direction	150	150	150

In the three cases, the last realization was always taken while the others were discarded. According to Chiang (2005:167), the hydraulic conductivity is commonly assumed to be log-normally distributed, Y = log(X). This results in a log-normal distribution of X when Y is normally distributed with a mean value μ and standard deviation σ.

4.2.6 Effective porosity

Two zones of effective porosity were assigned to layer 1 (Figure 12). The orange zone with a value of 0.28 corresponds roughly to the higher hydraulic conductivity values in layer 1 while the brown zone (0.22) corresponds to the lower ones. The second layer remained homogeneous and received an effective porosity value of 0.25.

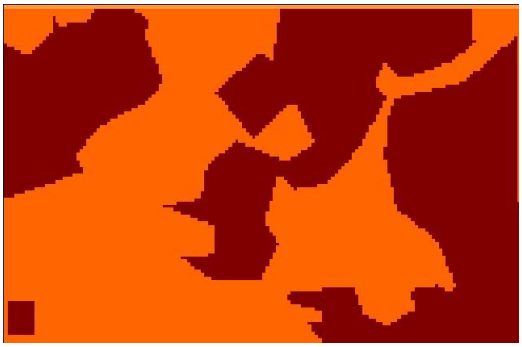

Figure 12: Effective porosity in layer 1: Orange = 0.28, Brown = 0.22

4.2.7 Parametrization of MODFLOW packages: River, Well and Solver

The river is located at the South boundary of our area and flows with a small gradient from left to right. The river water level does not vary through time and is at the left border 36.6 m, in the middle 36.5 m, and 36.4 m at the right border. The river has a constant width of 10 m and a riverbed bottom elevation of 30 m. The riverbed conductance was set to 0.003 m^2/s which is equivalent to a 10 cm thick riverbed layer with 3 x 10^{-6} m/s hydraulic conductivity.

A water supply well was defined at the following x, y coordinates: 1045 m, 295 m. At this location there was no influence of the domain boundaries. The well is screened on the interval 20 to 40 m depth (2^{nd} layer) and discharges 1000 m^3/d.

The solver chosen to solve the resulting system of simultaneous finite difference linear equations is the PCG2 (Preconditioned Conjugate Gradient Package 2), with the default preconditioning method: Modified Incomplete Cholesky.

4.2.8 Transport model

The transport of a non-reactive compound (just advection and dispersion, no chemical reaction transport) was simulated with MT3D[11]. The initial (background) concentration was set to zero. In MT3D, there are four solution schemes to simulate advection. According to Chiang (2005:94), the Upstream Finite Difference Method is only suitable for solving transport problems not dominated by advection due to the problems of numerical dispersion and artificial oscillation. The Method of Characteristics (MOC) is virtually free of numerical dispersion but it can be slow and requires a large amount of computer memory. The Modified Method of Characteristics (MMOC) is normally faster than the MOC but introduces some numerical dispersion. And the hybrid method of characteristics (HMOC) combines both,

[11] Modular three-Dimensional Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems, Version DOD_1.5, July 1996; comes with PMWIN.

when sharp concentrations are present advection is solved by MOC, otherwise by MMOC. For this model, the MMOC solution scheme was selected because this method is a compromise between performance and numerical dispersion and it has much less parameters to be determined than the HMOC. The parameters of the MMOC were left as default, except the Courant number. This number stands in MMOC for the fraction of a cell that a particle is allowed to move in any direction in one transport step and it was set to 0.5. One restrains the particle shift per time step to avoid numerical instability, which could happen if a very high velocity is calculated for a particle and makes it "jump" a cell. At the cells with the highest hydraulic gradient values, velocities up to 3 m/d are contemplated, thus, no particle jumps are expected.

MT3D provides three particle tracking algorithms: 1st-order Euler, 4th-order Runge-Kutta, or a hybrid of these two. According to Chiang (2005:95), the 1st-order Euler has a big numerical error unless the transport step is small. The 4th-order Runge-Kutta needs more computational effort because it provides the average velocity of the particle at four different times, at the beginning of the time step, at the end, and at two midpoints in between. The hybrid approach was selected as the particle tracking algorithm. This approach uses the 4th-order Runge-Kutta algorithm just on cells adjacent to or containing sink and source cells; otherwise, it uses the 1st-order Euler approach.

4.2.9 Dispersion

There are four parameters on the dispersion package of MT3D: Horizontal longitudinal dispersivity, horizontal transverse dispersivity, vertical dispersivity, and molecular diffusion. The values for the first three parameters are given in the table below. The fourth parameter is not taken into account since dispersion by molecular diffusion is very small and just important when there are very slow groundwater velocities. These are not present in this model.

Table 9: Dispersivity values of data generating model

Dispersivity	Layer 1 (m)	Layer 2 (m)
Longitudinal	12.00	12.00
Transverse	1.32	0.84
Vertical	0.66	0.60

Dispersion is scale dependent, the larger the domain (in flow direction), the bigger the dispersivity value. For 1000 m distance a value of longitudinal dispersivity greater than 10 m can be expected (Walther et al. 2008:195). Furthermore, the transverse and vertical dispersivities are usually one to two orders of magnitude smaller than the longitudinal dispersivity. However, these values are not seen by the model since they are much lower than the cell size. Nevertheless, we obtain plausible model results since the longitudinal

dispersion and the heterogeneity of the porous media are intended to be the main contributors to dispersion in the model.

4.2.10 Concentration source

As concentration source, a “recharge” value of 1000 g “tracer” per cubic meter water per day was constantly supplied for a period of 28 days. The element introduced in the system stands for a virtual non reactive substance, analogous to the tracers used by hydrogeologists in the field. A saturated cell of 10 x 10 x 20 m (x, y, z) with an effective porosity of 0.22 provides for 440 m^3 water. By multiplying this last value times 1000 g/(m^3d) tracer, we obtain approx. 5 g “tracer” recharge input at every source cell. The source is located in four cells of the upper layer (column,row): (40,19), (41,19), (41,18), and (42,18), see small polygon between borehole 1 and 61 in the top left corner of Figure 13 (p. 42).

4.2.11 Time discretization

In order to simulate transport of the substance, 364 days were discretized into two stress periods: a first period of 28 days with constant tracer input and a second one of 336 days where no input takes place. A summary of the time discretization values is shown in the following table:

Table 10: Time discretization overview of data generating model

Attribute	Value
Simulation flow type	Steady state
Number of stress periods	2
Length Period 1	28 days
Length Period 2	336 days
Total simulation time	364 days
Transport step size	1 day

The value of the transport step size was carefully chosen. Numerical problems appear, if the transport step is too big. On the other hand, if the transport step is too small, the computer demand increases. The optimal value depends on the velocity of the fluid, the dispersion and cell length in flow direction (Δx). The relationships between these parameters are given by the three terms shown in the following section (Co, Ne and Pe).

4.2.12 Numerical stability

The stability in groundwater numerical models is usually represented by unitless numbers. To maintain numerical stability during simulation, the Courant criterion (Co) has to be smaller or equal to one and the Neuman criterion (Ne) smaller or equal to 0.5 (Kolditz 1997:69):

$$Co = \frac{\Delta t \cdot v}{\Delta x} \leq 1 \qquad 49$$

and

$$Ne = \frac{D \cdot \Delta t}{\Delta x^2} \leq 0.5 \qquad 50$$

where Δt is the transport time step; v stands for the fluid velocity within the pores (seepage velocity), which is the Darcy velocity divided by the effective porosity of the medium (n_{eff}), $v = -K \nabla h / n_{eff}$, where K is the hydraulic conductivity and ∇h is the hydraulic pressure gradient; D is the dispersion coefficient, $D = \alpha_L \cdot v$, where α_L is the longitudinal dispersivity.

These conditions were evaluated twice in layer 1: once in the high conductivity zone which resembles a paleo-channel (Co = 0.3, Ne = 0.4), and outside of it (Co = 0.009, Ne = 0.01). The conditions were fulfilled for both cases, which suggests stability of the numerical transport computation even at the high velocity zones of the channel. The highest average velocity attained was 3.5 m/d in some narrow channel segments. Average velocity in the channel was around 3 m/d. The average velocity of selected channel segments was calculated using PMPATH v.6.1.0 (Chiang and Kinzelbach 1994-2001). This particle tracking program reads the Modflow result files and calculates flowlines, pathlines and velocities at single cells.

The Peclet number (Pe), which expresses the relationship between advective and dispersive transport (Kolditz 1997:69), was also calculated:

$$Pe = \frac{v \cdot \Delta x}{D} = \frac{\Delta x}{\alpha_L} \qquad 51$$

The value obtained for layers one and two was Pe = 0.8. This suggests that the transport process is more affected by dispersion than by advection, the differential equation is more parabolic and stability is expected, see following table:

Table 11: Interpretation of the Peclet Number (Kolditz 1997:69)

Physical	proportion ratio between advective to dispersive transport	Pe < 1 more dispersive Pe > 1 more advective
Mathematical	Differential equation characterization for transport properties	Pe < 1 more parabolic Pe > 1 more hyperbolic
Numerical	Stability criteria	Pe $\leq$ 2 linear elements Pe $\leq$ 4 quadratic elements

4.2.13 Boreholes and observations

In order to collect data, 100 points were selected. The boreholes were randomly chosen but taking into account the following guidelines: as in real life, there are more boreholes in upper layers as in lower layers, since deeper boreholes involve higher costs; 57% of the wells are in layer one and 43% in layer two. Also more boreholes are found in higher hydraulic conductivity zones as in lower ones since the first ones are of more interest for water extraction or injection. Therefore, more wells are located to the left side in the upper layer and to the right side in the lower one, as it can be appreciated in the following figure:

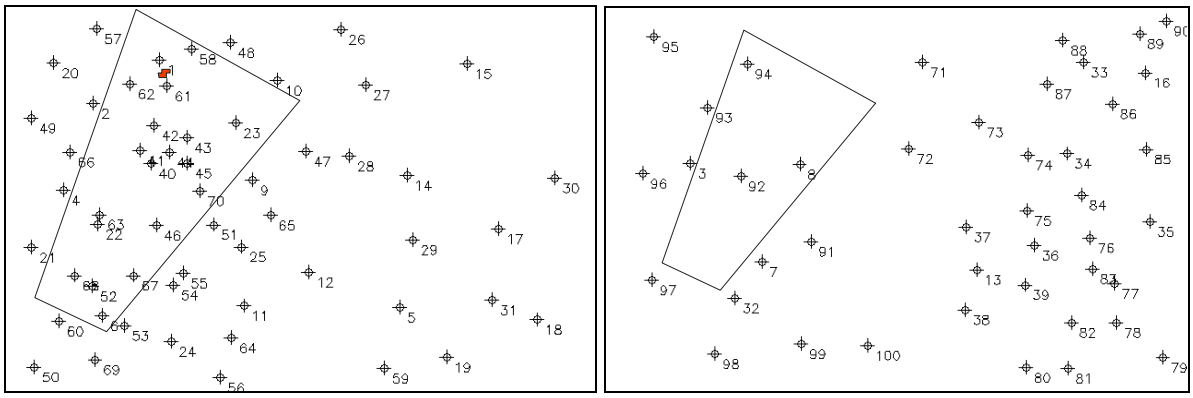

Figure 13: Boreholes distribution for reading hydraulic pressure and conductivity values in layer 1 (left) and layer 2 (right). Polygon areas encompass boreholes which were additionally used to read concentration values. The concentration input zone is shown in the small red polygon between boreholes 1 and 61 in the top left corner of layer 1

However, not all boreholes are of interest to sample concentration. Most of them do not register concentration values at any time. Therefore, just the boreholes inside the polygons shown in Figure 13 were selected to read concentration values. The polygons are oriented more or less along the high conductivity zone which resembles a paleo-channel (see Figure 10, p. 35). The polygon in layer two is 1/3 smaller than in layer one. At the output control of MT3D, it was specified that every 28 days the concentration is read. The limit of detection was set to $1x10^{-7}$ g/m^3 per cell, any value smaller than this was not seen and noted as zero.

4.2.14 Model results

After completing the parametrization and successfully running the model, we obtain spatially distributed results of hydraulic pressure, velocity vectors, and concentration (see Figure 14 and 15). Groundwater flows from high hydraulic pressure (dark blue in the figure below) to low hydraulic pressure (light blue). This produces a flow roughly from North to South.

However, groundwater does not flow through every cell in the same quantity. Most of the flow occurs in specific regions, as it can be seen in the velocity vector distributions (Figure 14, middle section). These regions correspond to high conductivity areas of the model (see Figure 10, p. 35). The concentration distributions shown at the bottom of Figure 14 are those of day 364. This is the last day of the simulation and maximum extent of the plume in the model. The dark orange regions are of high concentration and the contours decrease by one order of magnitude until $1x10^{-7}$ g/m^3, which is the detection limit (light orange in the following figure). The concentration plume is present in both layers and follows the “channel” of high conductivity values. The synthetic data generated at the “boreholes” is annexed in Appendix B.4 (p. 99).

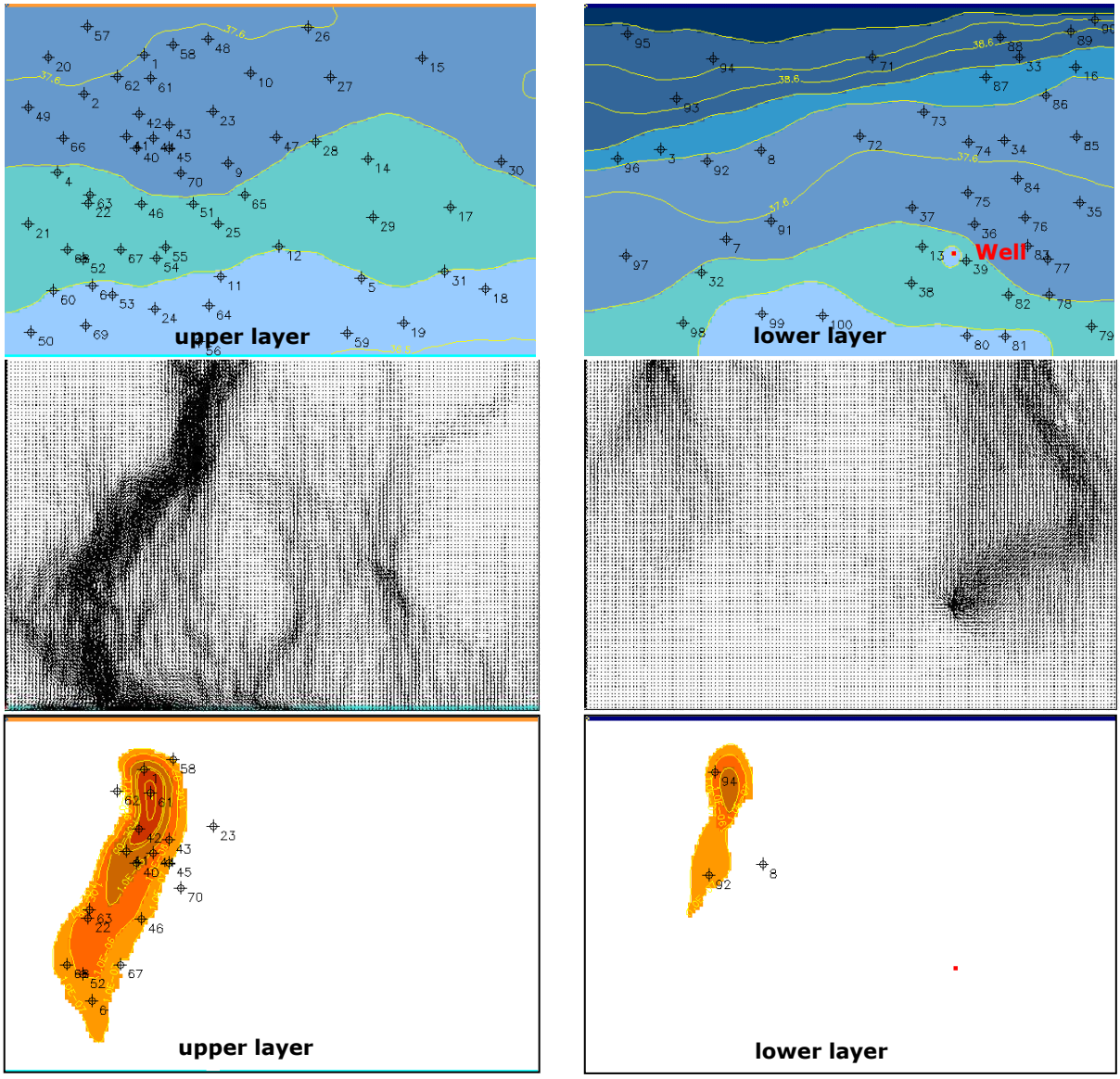

Figure 14: Model distributions of hydraulic pressure, velocity vectors, and tracer concentration at day 364; upper layer (left) and lower layer (right)

Figure 15 shows the velocity vectors in North - South profile at the middle of the model area. The velocity vectors, calculated with PMPATH, represent magnitude and direction of the flow velocity. Flow in layer 1 is more horizontal and faster as in layer 2. Groundwater flows from the constant head boundaries (in the figure below in dark blue) and it discharges in the river (marked in light blue).

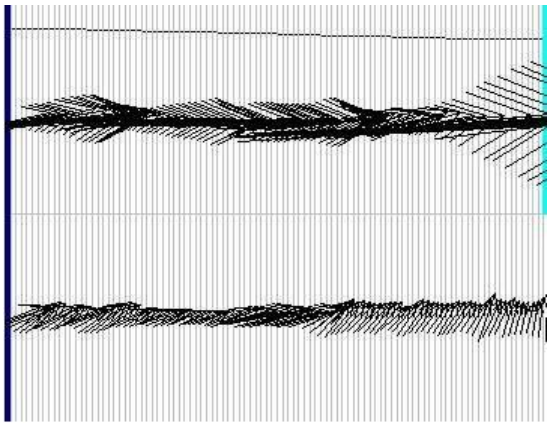

Figure 15: Velocity vectors North - South profile at the middle of the model

4.3 CANDIDATE MODELS

Seven models, each of them with increasing number of parameters to be calibrated were created to simulate the synthetic data. The parameters are basically different zones of hydraulic conductivity (K), either in the riverbed ($Cond_{Riverbed}$), or as horizontal (K_h), or vertical hydraulic conductivity (K_V) zones in the model layers.

Model 1 is the simplest of all seven models because it consists of one homogeneous layer with just one parameter to be calibrated, the horizontal hydraulic conductivity ($K_{Lay\ 1}$). Model 2 consists of two homogeneous layers with one horizontal hydraulic conductivity value ($K_{Lay\ 2}$ is equal to $K_{Lay\ 1}$) and a value of vertical hydraulic conductivity (K_V). Model 3 consists of two independent homogeneous layers ($K_{Lay\ 1}$ and $K_{Lay\ 2}$, which are separated by K_V). In Model 4, the K_h of the upper layer is divided in a channel zone ($K_{Lay\ 1\ channel}$) where the highest conductivity values are observed and in a second zone which comprises the rest of the layer ($K_{Lay\ 1\ without\ channel}$). Model 5 adds to it a zonation of the lower layer in right ($K_{Lay\ 2\ right}$) and left zone ($K_{Lay\ 2\ left}$). Model 6 introduces a North - South zonation of K_V ($K_{V\ North}$ and $K_{V\ South}$). Finally, Model 7 adds the hydraulic conductivity of the riverbed ($Cond_{Riverbed}$) as calibration parameter. A summary of the models calibration parameters is shown below:

Table 12: Calibration parameters of candidate models (newly introduced parameters are marked bold)

Model name and number of calibration parameters	Calibration parameters
1	**$K_{Lay\ 1}$**
2	$K_{Lay\ 1}$ and **K_V** ($K_{Lay\ 2}$ is tied to $K_{Lay\ 1}$)
3	$K_{Lay\ 1}$, **$K_{Lay\ 2}$** and K_V
4	**$K_{Lay\ 1\ channel}$, $K_{Lay\ 1\ without\ channel}$**, $K_{Lay\ 2}$ and K_V
5	$K_{Lay\ 1\ channel}$, $K_{Lay\ 1\ without\ channel}$, **$K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$** and K_V
6	$K_{Lay\ 1\ channel}$, $K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$, **$K_{V\ North}$** and **$K_{V\ South}$**
7	$K_{Lay\ 1\ channel}$, $K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$, $K_{V\ North}$, $K_{V\ South}$ and **$Cond_{Riverbed}$**

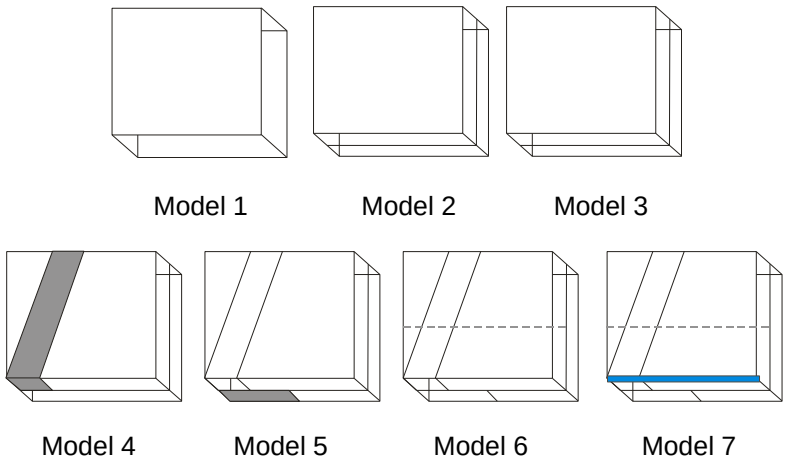

Figure 16: Hydraulic conductivity zonation of the seven candidate models

All the candidate models share the following properties:

- the constant head values of the back of the layer (North border) were estimated from the head observations near the boundary. For Models 2 to 7, this resulted in evenly decreasing values at the upper layer: 37.9 m at the west then, 37.8 m, 37.7 m and 37.6 m to the east. For the lower layer (west - eastward): 39.6 m, 39.4 m, 39.2 m and 39.0 m. Since Model 1 is a one layer model, all of the observations were considered to take place in this layer. The observations near the boundary suggest evenly decreasing head values, starting at the west with 38.75 m, then 38.60 m, 38.45 m and 38.30 m to the east;
- effective porosity 0.25, which is a value that can occur in medium, coarse, gravelly sand and fine gravel (Hölting 1992:79);

- longitudinal dispersivity of 10 m and 1 m horizontal and vertical transverse dispersivity;
- the calibration data: 100 hydraulic pressure values, 100 hydraulic conductivity values and 100 concentration values (see Appendix B.4, p. 99).

The following properties shared by the candidate models are also the same as in the data generating model:

- geometry (see Section 4.2.2, p. 34), except for Model 1 which is a single layer model where just the upper layer is considered;
- upper layer unconfined and lower layer confined;
- boundary conditions of Models 1 to 6 differ from those of the data generating model (Table 5, p. 34) only in that instead of a river boundary, the front of layer 1 (Southern boundary) is constant head with the same pressure values as in the river stage (see Section 4.2.7, p. 38);
- Model 7 boundary conditions;
- steady state flow;
- recharge (see Figure 10, p. 35);
- time discretization (see Table 10, p. 40);
- the well in the lower layer (see Section 4.2.7, p. 38). Except in Model 1 where it is in the upper layer, since the lower layer is not present;
- initial background concentration: zero;
- 1000 $g/(m^3_{water}d)$ "tracer" input supplied as recharge during the first stress period in the same 4 cells;
- MODFLOW solver, MT3D solution scheme and particle tracking algorithm.

4.4 MODEL SELECTION ANALYSIS WITH DIFFERENT TYPES OF OBSERVATIONS

4.4.1 Model selection analysis with one type of observations

The candidate models were calibrated to 100 hydraulic pressure (h) observations using PEST (Doherty 2006), an automated hydrologic calibration inverse code. Besides PEST there is another inverse code for automated hydrologic calibration: UCODE_2005 (Poeter et al. 2005). Both are available in MODFLOW-2000. They even calculate $AIC_{Hill\ and\ Tiedeman}$ (Equation 42, p. 26) and $BIC_{Hill\ and\ Tiedeman}$ based on Equation 93 (p. 98). However, due to the problems mentioned in Section 3.8 (p. 26), the criteria are not correctly applied, and thus, not

considered here. The parameter values obtained through PEST calibration are shown in the following table:

Table 13: Parameter values obtained from automated calibration to 100 h observations

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
$K_{Lay\ 1\ without\ channel}$ (m/s)	0.222	0.02384	0.03193	0.02082	0.02091	0.01901	0.03242
$K_{Lay\ 1\ channel}$ (m/s)				0.05338	0.05413	0.04449	0.04830
$K_{Lay\ 2\ right}$ (m/s)		= $K_{Lay\ 1}$	0.02306	0.02099	0.02106	0.03250	0.01785
$K_{Lay\ 2\ left}$ (m/s)					0.02012	0.03141	0.02069
$K_{V\ North}$ (m/s)		1.9E-07	1.8E-07	1.7E-07	1.6E-07	1.4E-07	9.1E-08
$K_{V\ South}$ (m/s)						3.9E-07	2.3E-07
$Cond_{Riverbed}$ (m^2/s)							0.12246

The horizontal conductivity in the channel ($K_{Lay\ 1\ channel}$) and the values of vertical hydraulic conductivities ($K_{V\ North}$ and $K_{V\ South}$) are in the same order of magnitude as the ones used in the data generating model (see Figure 10, p. 35). The value of the other parameters ($K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$ and $Cond_{Riverbed}$) lay much higher than in the data generating model. The best fit of each model is obtained by using horizontal hydraulic conductivity values equivalent to those of pure gravel. A satisfactory explanation for this behaviour could not be found. One would have expected that some conductivity values lay indeed much higher, yet others much lower than those of the reference model.

The calibrated models attained the best possible fit at their respective complexity level. With every additional parameter model fit improves. A summary of the results from the model selection evaluation for the seven candidate models is given in the following table and graph:

Table 14: Model selection analysis of candidate models calibrated to 100 h observations

Model			AIC_c				AIC				BIC			
Nr.	Fit	K	Penalty	AIC_c	Δ_i	w_i	Penalty	AIC	Δ_i	w_i	Penalty	BIC	Δ_i	w_i
1	-68.4	2	4.1	-64.2	214.9	0.00	4	-64.4	216.4	0.00	9.2	-59.1	212.0	0.00
2	-285.0	3	6.3	-278.7	0.5	0.27	6	-279.0	1.8	0.18	13.8	-271.1	0.0	0.87
3	-285.3	4	8.4	-276.9	2.3	0.11	8	-277.3	3.4	0.08	18.4	-266.9	4.3	0.10
4	-286.0	5	10.6	-275.4	3.7	0.05	10	-276.0	4.7	0.04	23.0	-263.0	8.1	0.02
5	-286.9	6	12.9	-274.0	5.2	0.03	12	-274.9	5.9	0.02	27.6	-259.3	11.9	0.00
6	-293.4	7	15.2	-278.2	0.9	0.21	14	-279.4	1.3	0.23	32.2	-261.2	9.9	0.01
7	-296.7	8	17.6	-279.2	0.0	0.34	16	-280.7	0.0	0.44	36.8	-259.9	11.2	0.00

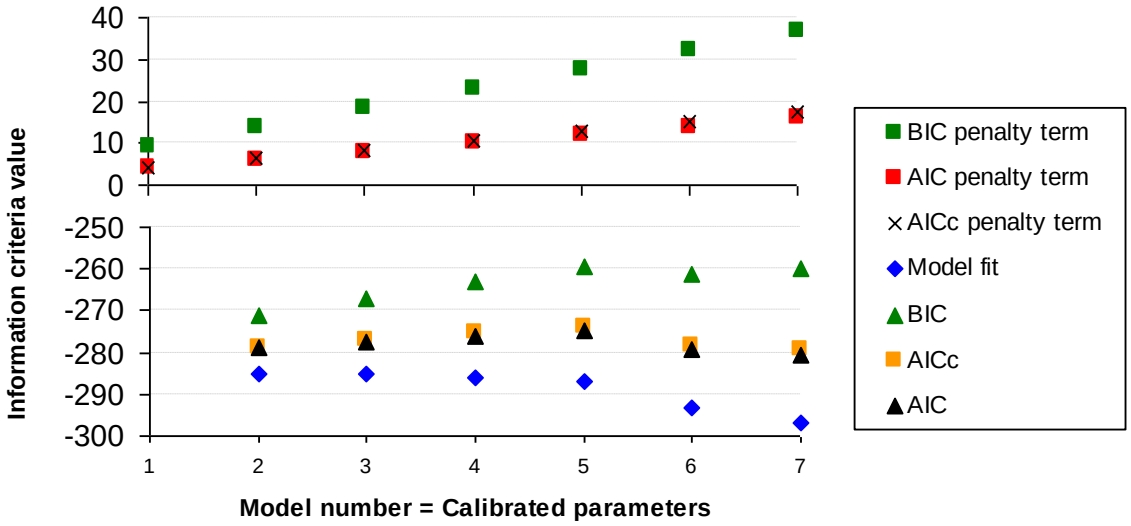

Figure 17: AIC_c, AIC, and BIC analysis of candidate models calibrated to 100 h observations

As mentioned in Section 3.3, K represents the number of hydrological plus statistical calibration parameters, Δ_i the difference to the best model and w_i the probability of having the best model. The difference between the fit of the first model (fit = -68.4) and the rest of them is so large that Model 1 results do not appear on the range of Figure 17. The ratio n / K is 12.5 (= 100 / 8). This suggests the use of AIC_c for ranking purposes instead of AIC. AIC_c selects Model 7 with 34 % certainty as the best model, followed by Model 2 (27 %) and Model 6 (21 %). The rankings of AIC and BIC are disregarded. However, they are presented here to corroborate the observation of Ye et al. (2008) and Poeter and Anderson (2005) that they perform similarly in applications (Section 3.6). The best models selected by AIC are also the same as those of AIC_c, however, in another ranking order: Model 7 (44 %) ranks also first, followed this time by Model 6 (23 %) and Model 2 (18 %). For BIC, the best model is number 2 with a high certainty (87 %).

In summary, with 100 h observations available for calibration, Model 7 (the most complicated) or Model 2 (one of the simplest ones) are chosen by AIC_c as the best models with a certainty of around 30 % each. The other models are either not simple enough or their model fits are not good enough to justify the increase in complexity.

4.4.2 Model selection analysis with two types of observations

The same candidate models as in Section 4.3 (p. 44) are now calibrated to two types of observations: as done before, to hydraulic pressure (h) and additionally to horizontal

hydraulic conductivity (K_h). This last parameter is not a model result like h or concentrations. To emphasize this difference K_h observations are usually called “prior information.” However, in the same manner as typical observations are compared to model results, the simulated K_h values are here also compared to the “observed” K_h values. Therefore, in the present study they are all referred as “observations” since there is here no practical reason to distinguish K_h observed values from other observed values.

The models were calibrated to 100 h and 100 K_h observed values using a mixture of PEST (Doherty 2006) and “trial and error” calibration methods. Such a combination is the most appropriate calibration strategy (Kresic 1997:336). K_V and $Cond_{Riverbed}$ parameters are first calibrated with PEST to fit the h observations; their fit to the K_h observations is evaluated manually. The PEST-calibrated parameters are used then as departure values for “trial and error” calibration runs. The K_h input values which fit the best the K_h observations are known. These values are the mean of the K_h observations. For example, the best input value for Model 1 is simply the mean of all K_h observations. However, the K_h average might not be necessarily the value which provides the best fit to the h observations. Hence, the value of K_h is manually changed and the model fit to h and K_h is evaluated norming the values as in Section 4.1 (p. 31). It should be recalled that in the just referred section we obtain unitless residuals which can be evaluated as a single data set by norming the observed and calculated values of h and log K_h. The norming values are their respective observations standard deviation values (σ_h and σ_{logK}). From the resulting unitless residuals we estimate the maximum likelihood variance (Equation 14, p. 13), which allows us to implement AIC. Notice that when calculating the variance or standard deviations, the data sets are previously centered at zero. The final parameter values obtained through calibration are shown in the following table:

Table 15: Parameter values obtained from calibration to 100 h and 100 K_h observations

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Observed mean
$K_{Lay\ 1\ without\ channel}$ (m/s)	0.001	0.002	0.00209	0.00072	0.00061	0.00068	0.00065	0.00061
$K_{Lay\ 1\ channel}$ (m/s)				0.001	0.002	0.0021	0.0056	0.0054
$K_{Lay\ 2\ right}$ (m/s)		= $K_{Lay\ 1}$	0.00051	0.00038	0.00046	0.00052	0.00049	0.00046
$K_{Lay\ 2\ left}$ (m/s)					0.00012	0.00014	0.00013	0.00011
$K_{V\ North}$ (m/s)		3.9E-07	7.5E-08	5.3E-08	3.1E-08	4.0E-10	5.1E-11	
$K_{V\ South}$ (m/s)						9.3E-08	8.0E-08	
$Cond_{Riverbed}$ (m^2/s)							0.06357	

From the above table we can tell that from all models, the best fit to the K_h observations is that of Model 7, followed by number 5. The observed mean value (last column of the above

table) is the average of the K_h values at the boreholes in the respective parameter K_h zone. However, this does not necessarily mean that Models 5 or 7 also provide the best fit to the h observations. A model fit to both K_h and h observation types is provided in the table and figure below.

The model selection evaluation of the seven models is summarized in the table below. With additional parameters, model fit improves while the penalty increases in linear (AIC and BIC) or almost linear (AIC_c) form, see table and figure below:

Table 16: Model selection analysis of candidate models calibrated to two types of observations

Model			AIC_c				AIC				BIC			
Nr.	Fit	K	Penalty	AIC_c	Δ_i	w_i	Penalty	AIC	Δ_i	w_i	Penalty	BIC	Δ_i	w_i
1	-54.9	2	4.1	-50.9	173.3	0.00	4	-50.9	173.9	0.00	10.6	-44.3	157.4	0.00
2	-111.2	3	6.1	-105.1	119.1	0.00	6	-105.2	119.6	0.00	15.9	-95.3	106.4	0.00
3	-161.3	4	8.2	-153.3	71.1	0.00	8	-153.3	71.4	0.00	21.2	-140.1	61.6	0.00
4	-217.5	5	10.3	-207.2	17.0	0.00	10	-207.5	17.3	0.00	26.5	-191.0	10.7	0.00
5	-232.3	6	12.4	-219.9	4.3	0.08	12	-220.3	4.5	0.07	31.8	-200.5	1.2	0.34
6	-238.8	7	14.6	-224.2	0.0	0.67	14	-224.8	0.0	0.66	37.1	-201.7	0.0	0.61
7	-239.0	8	16.8	-222.2	2.0	0.25	16	-223.0	1.8	0.27	42.4	-196.6	5.1	0.05

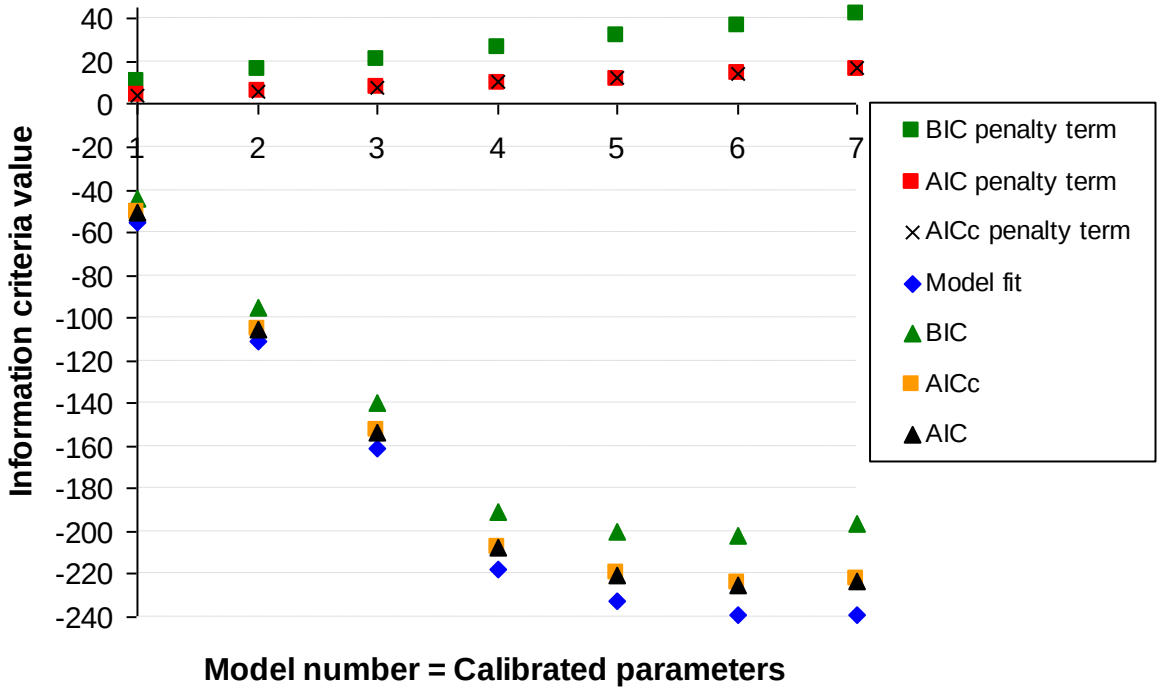

Figure 18: AIC_c, AIC, and BIC analysis of candidate models calibrated to 100 h and 100 K_h observations

The ratio n / K is 25 (= 200 / 8). The use of AIC_c is preferred to AIC, although, their behaviour is very similar, as we can discern from the above table and figure. All model selection

methods agree with approximately 2/3 of certainty that Model 6 is the best option (see w_i values in the above table). The second position is with approx. 30 % certainty reserved to Model 7 (AIC and AIC_c) or to Model 5 (BIC, approx. 30 %). This difference is due to the heavier penalties of BIC for adding parameters. BIC prefers as second option a simpler model, while AIC and AIC_c prefer a more complex one. The most parsimonious model (number 6) is not the one with the best fit to the K_h values (Model 7) and it is different to the one chosen as the best model in Section 4.4.1 (p. 46) using just one type of information (Model 2 and 7). Even though we have the same conceptual models, the same modeler and the same h observations, the ranking is different. However, this time we had additionally 100 K_h observations. In this example there might be enough data or enough data types to clearly support a complex model as the best parsimonious model. The effect of the number of observations in the model selection will be discussed in Section 4.5.1 (p. 54).

4.4.3 Model selection analysis with three types of observations

The same candidate models evaluated in the last two sections are now calibrated to all three data sets: 100 h, 100 K_h and 100 concentration observations.

The models are not calibrated with PEST or UCODE_2005 since they do not work with different types of normed observations values (as in Section 4.1, p. 31). These automated inverse codes work well with one type of observations. It is possible, for example, to calibrate the parameters of a groundwater flow model with the help of these codes using hydraulic pressure observations (as it was carried out in Section 4.4.1). It is also possible to use these codes to calibrate a transport model with just concentration data. The models here were calibrated by “trial and error” to the observed values of h (m), K_h (m/s) and concentration (mg/l). The unitless residuals were obtained as explained in Section 4.1 and the models were ranked according to their AIC_c value. “Trial and error” calibration was the first calibration technique in groundwater modeling and it is still preferred by many modelers. It is highly recommended to perform this type of calibration since the modeler develops a better “feeling” for the model and its assumptions (Kresic 1997:331).

More than 2000 “trials” were performed to find the best possible parameter values. The results obtained from calibration are listed in the following table:

Table 17: Parameter values obtained from calibration to 100 h, 100 K_h, and 100 concentration observations

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Observed mean
$K_{Lay\ 1\ without\ channel}$ (m/s)	0.00029	0.00052	0.00062	0.00048	0.00048	0.00048	0.00060	0.00061
$K_{Lay\ 1\ channel}$ (m/s)				0.0011	0.0011	0.0011	0.0016	0.0054
$K_{Lay\ 2\ right}$ (m/s)		= $K_{Lay\ 1}$	0.00055	0.00029	0.00032	0.00032	0.00037	0.00046
$K_{Lay\ 2\ left}$ (m/s)					0.00025	0.00024	0.00009	0.00011
$K_{V\ North}$ (m/s)		2.1E-07	2.8E-07	1.8E-07	2.1E-07	2.1E-07	7.0E-09	
$K_{V\ South}$ (m/s)						2.3E-07	3.1E-07	
$Cond_{Riverbed}$ (m^2/s)							0.00043	

Model 7 approaches the most the observed K_h mean values. In contrast to the calibrated values obtained with just one data set, the values obtained here lay within the same order of magnitude as the ones in the original model which generated the data.

The model selection results of all seven models are summarized in the table below. The model fit improves with every extra parameter. The penalty term of AIC_c is by 300 observations almost linear as the AIC one (red squares and crosses at the bottom of the figure below). The BIC penalty term is linear but more pronounced than that of AIC. BIC punishes the use of additional parameters stronger than AIC and AIC_c. AIC_c w_i confirms with a 100% of confidence that Model 7 is the best model supported by the data. The ratio n / K is almost 40, (300 / 8 = 37.5). At this ratio, AIC and AIC_c give practically the same result (see Figure 19). From these two criteria, just the AIC_c values are included in the table below to improve its readability.

Table 18: AIC_c and BIC evaluation of candidate models calibrated to 100 h, 100 K_h, and 100 concentration observations

Model		AIC_c				BIC			
Nr.	Fit	Penalty	AIC_c	Δ_i	w_i	Penalty	BIC	Δ_i	w_i
1	352.7	4.0	356.8	155.1	0.00	11.4	364.1	133.3	0.00
2	342.6	6.1	348.6	147.0	0.00	17.1	359.7	128.9	0.00
3	320.4	8.1	328.6	126.9	0.00	22.8	343.3	112.5	0.00
4	203.9	10.2	214.1	12.5	0.00	28.5	232.5	1.7	0.29
5	201.9	12.3	214.2	12.6	0.00	34.2	236.2	5.4	0.04
6	201.7	14.4	216.0	14.4	0.00	39.9	241.6	10.8	0.00
7	185.1	16.5	201.6	0.0	1.00	45.6	230.8	0.0	0.66

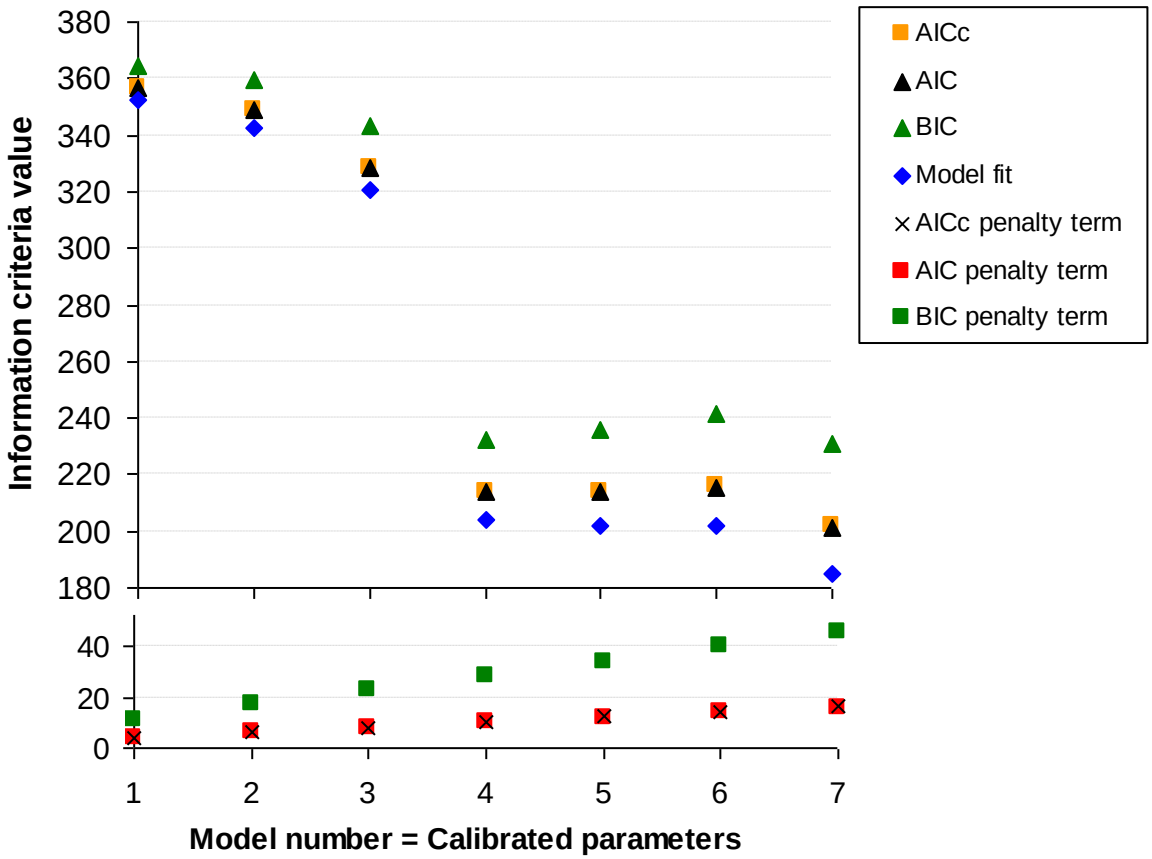

Figure 19: AIC_c, AIC, and BIC evaluation of candidate models calibrated to 100 h, 100 K_h, and 100 concentration observations

In this case, AIC_c selects the model which most resembles the synthetic data generating model. Model 7 has the lowest AIC_c value and also the highest number of parameters. In comparison to the other 6 models, Model 7 attains with 100 % certainty the best equilibrium between complexity and model fit. BIC selects Model 7 also as the best model, with 66 % of certainty (see BIC w_i, above table) followed by Model 4 with 29 % of probability of being the best model. BIC would even consider Model 5, however, with minor support (4 %). By considering the 3rd type of observations (concentrations) the model selection evaluation also reveals a big difference between models with channel (4 to 7) and those without channel (1 to 3). Since concentration data are considered in this case, including information about the channel is important. The channel controls transport behaviour to a much higher extent than flow behaviour. In this example there is enough data to support the most complex model as the best parsimonious model. However, this result might have been just the effect of a subjective selection of the order of parameters. If the channel (parameter $K_{Lay\ 1}$ differentiated in Model 4) and $Cond_{Riverbed}$ (added in Model 7) had been introduced before, the model selected would probably have been one with less parameters. The effect in model selection of the order in which the parameters are introduced will be discussed in Section 4.5.3 (p. 59).

4.5 DISCUSSION

4.5.1 Impact of number of observations

When considering 100, 200 or 300 observations in Section 4.4, AIC_c selected models of different complexity level. Not just the number of observations increased but also the types of information considered. By calibrating the models to the three types of information, the value of the estimated parameters (Table 17) resemble more the ones of the original data generating model (Figure 10, p. 35) in contrast to the parameter values estimated by the models calibrated with just one type of information (Table 13, p. 47). In order to see the impact of the number of observations in model selection, an evaluation with fewer observations than in previous examples is performed hereupon.

The candidate models are now calibrated to 20 h observations (data from the first 20 boreholes were selected, see Table 33, p. 99) instead of 100 h observations (Section 4.4.1, p. 46). 25% of the observations lay in the lower layer. The resulting calibrated parameter values are shown in the following table:

Table 19: Parameter values obtained from calibration to 20 h observations

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Observed mean
$K_{Lay\ 1\ without\ channel}$ (m/s)	0.00222	0.00457	0.01688	0.01051	0.02545	0.03158	0.04977	0.00037
$K_{Lay\ 1\ channel}$ (m/s)				0.0400	0.0818	0.0887	0.0705	0.0043
$K_{Lay\ 2\ right}$ (m/s)		= $K_{Lay\ 1}$	0.01437	0.01440	0.01902	0.02276	0.01365	0.00036
$K_{Lay\ 2\ left}$ (m/s)					0.00366	0.05183	0.02710	0.00010
$K_{V\ North}$ (m/s)		8.2E-08	1.1E-07	1.2E-07	3.1E-07	3.7E-07	2.1E-07	
$K_{V\ South}$ (m/s)						3.5E-07	1.8E-07	
$Cond_{Riverbed}$ (m^2/s)							0.12306	

The values of the parameters $K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$ and $Cond_{Riverbed}$ are as high as in Section 4.4.1, far away from those “observed.” The observed mean values (furthermost right column of the above table) are the K_h average values from the 20 boreholes in the respective hydraulic conductivity zones. $K_{Lay\ 1\ channel}$ and $K_{V\ North}$ are still one order of magnitude off and only $K_{V\ South}$ is similar to the one in the original data generating model (see Figure 10, p. 35).

A summary of the result of the model selection evaluation is given in the following table and graph:

Table 20: AIC evaluation of candidate models calibrated to 20 h observations

Model			AIC_c				AIC				BIC			
Nr.	Fit	K	Penalty	AIC_c	Δ_i	w_i	Penalty	AIC	Δ_i	w_i	Penalty	BIC	Δ_i	w_i
1	-15.8	2	4.7	-11.1	39.8	0.00	4	-11.8	42.0	0.00	6.0	-9.8	39.6	0.00
2	-58.4	3	7.5	-50.9	0.0	0.69	6	-52.4	1.4	0.16	9.0	-49.4	0.0	0.50
3	-58.5	4	10.7	-47.8	3.1	0.15	8	-50.5	3.4	0.06	12.0	-46.5	2.9	0.12
4	-59.6	5	14.3	-45.3	5.6	0.04	10	-49.6	4.2	0.04	15.0	-44.7	4.8	0.05
5	-65.6	6	18.5	-47.1	3.8	0.10	12	-53.6	0.2	0.29	18.0	-47.6	1.8	0.20
6	-65.7	7	23.3	-42.4	8.5	0.01	14	-51.7	2.1	0.11	21.0	-44.7	4.7	0.05
7	-69.9	8	29.1	-40.8	10.2	0.00	16	-53.9	0.0	0.33	24.0	-45.9	3.6	0.09

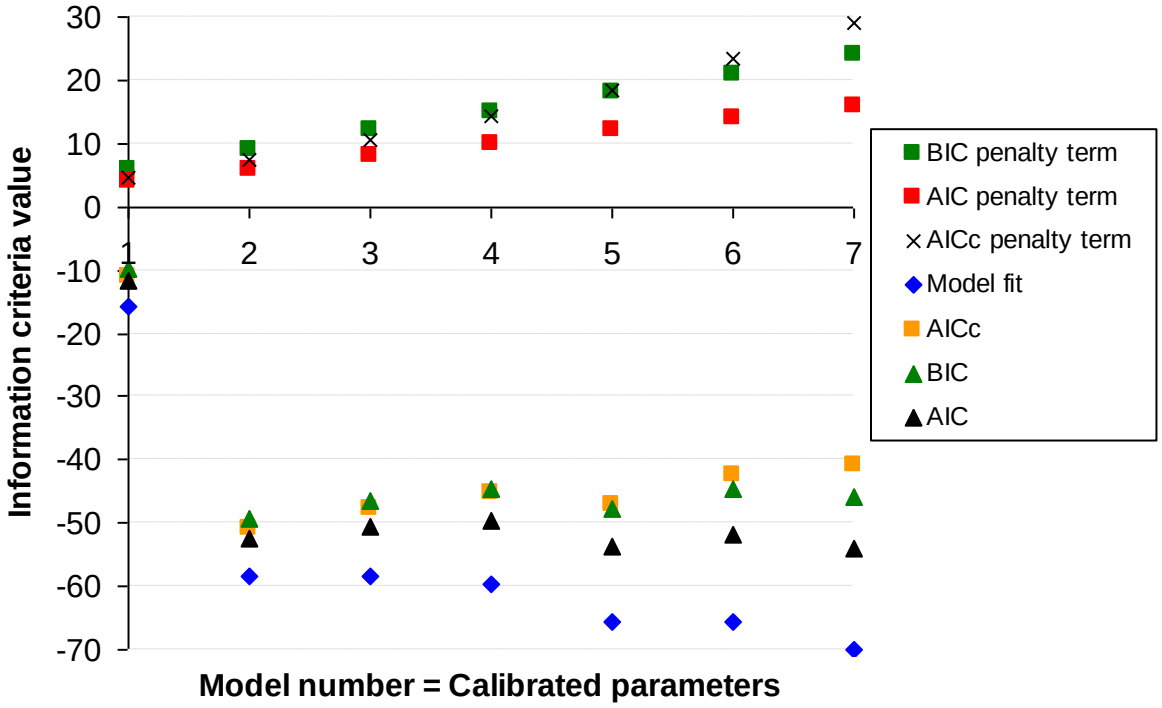

Figure 20: AIC_c, AIC, and BIC analysis of candidate models calibrated to 20 h observations

The ratio n / K is 2.5 (= 20 / 8). At this ratio, the AIC_c penalty term differentiates itself clearly from both AIC and BIC penalty terms. As mentioned in Section 3.5, AIC_c and KIC are better options than AIC and BIC respectively, if few observations are present. Still, these two last ones are calculated here for comparison purposes. In Section 4.4.1 (with 100 observations), the best models are numbers 2 and 7 (each with approx. 30 % probability). In contrast, the AIC_c weights evidence in this case with almost 70 % certainty that Model 2 represents the best option. Model 1 was improbable to be chosen, even with 20 observations, as observed already in Section 4.4.1; the fit to one layer model with just h observations is extremely poor. AIC_c recognizes that for this case a two-layer model describes the data better than a 1 layer model. However, for more model complexity, the data is certainly (70 %) not enough.

The seven candidate models are now calibrated to just the first 19 h and 19 K_h observations[12]. The normalized data is shown on Appendix B.5 (p. 101). The parameter values obtained through calibration are shown in the following table:

Table 21: Parameter values obtained from calibration to 19 h and 19 K_h observations

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Observed mean
$K_{Lay\ 1\ without\ channel}$ (m/s)	0.0007	0.0008	0.0007	0.00054	0.00037	0.00041	0.00042	0.00037
$K_{Lay\ 1\ channel}$ (m/s)				0.0055	0.0043	0.0046	0.0047	0.0043
$K_{Lay\ 2\ right}$ (m/s)		= $K_{Lay\ 1}$	0.0003	0.00029	0.00036	0.00041	0.00041	0.00036
$K_{Lay\ 2\ left}$ (m/s)					0.00010	0.00012	0.00012	0.00010
$K_{V\ North}$ (m/s)		1.0E-07	3.9E-08	3.8E-08	3.1E-08	4.0E-10	1.2E-10	
$K_{V\ South}$ (m/s)						9.3E-08	9.5E-08	
$Cond_{Riverbed}$ (m^2/s)							0.09916	

Model 5 fits again the observed K_h mean values very well as it did previously in Section 4.4.2. With more parameters, the values deviate from the mean; with less, the observed mean does not apply. The values in the last column of Table 21 are valid just for the differentiated K_h zones. However, K_h observations account for just half of the calibration data. A summary of the model selection results for the seven models is given in the following table and graph:

Table 22: AIC evaluation of seven candidate models calibrated to 19 h and 19 K_h observations

Model			AIC_c				AIC				BIC			
Nr.	Fit	K	Penalty	AIC_c	Δ_i	w_i	Penalty	AIC	Δ_i	w_i	Penalty	BIC	Δ_i	w_i
1	-6.4	2	4.3	-2.1	21.1	0.00	4	-2.4	22.7	0.00	7.3	0.9	17.7	0.00
2	-22.1	3	6.7	-15.4	7.7	0.01	6	-16.1	9.0	0.00	10.9	-11.2	5.6	0.03
3	-26.3	4	9.2	-17.1	6.0	0.02	8	-18.3	6.7	0.01	14.6	-11.8	5.1	0.05
4	-35.0	5	11.9	-23.2	0.0	0.48	10	-25.0	0.0	0.33	18.2	-16.8	0.0	0.57
5	-37.1	6	14.7	-22.4	0.8	0.32	12	-25.1	0.0	0.33	21.8	-15.2	1.6	0.26
6	-38.4	7	17.7	-20.6	2.5	0.14	14	-24.4	0.7	0.24	25.5	-12.9	3.9	0.08
7	-38.4	8	21.0	-17.4	5.7	0.03	16	-22.4	2.7	0.09	29.1	-9.3	7.6	0.01

The resulting ratio (n / K) for this assessment is 4.75 (= 38 / 8), which is much lower than 40. In Figure 21 we can discern a clear difference between AIC_c and AIC. The last parameter (riverbed conductance) brought no improvement in model fit. A constant head boundary (without resistance to flow) provides the same fit as a river condition without (or just minimal) flow resistance. For this case, the flow resistance property of parameter 7 seems not to be needed to attain a better fit to the 38 observations. The best model according to AIC_c is number 4 with 48 % of certainty, followed by number 5 with 32 % probability. BIC also ranks Model 4 as the best (57 %) followed again by Model 5 (26 %).

[12] These were the first calibrated models of the dissertation. The number of observations was arbitrarily chosen. Afterward, I decided to calibrate models with round numbers such as 100 or 20 observations.

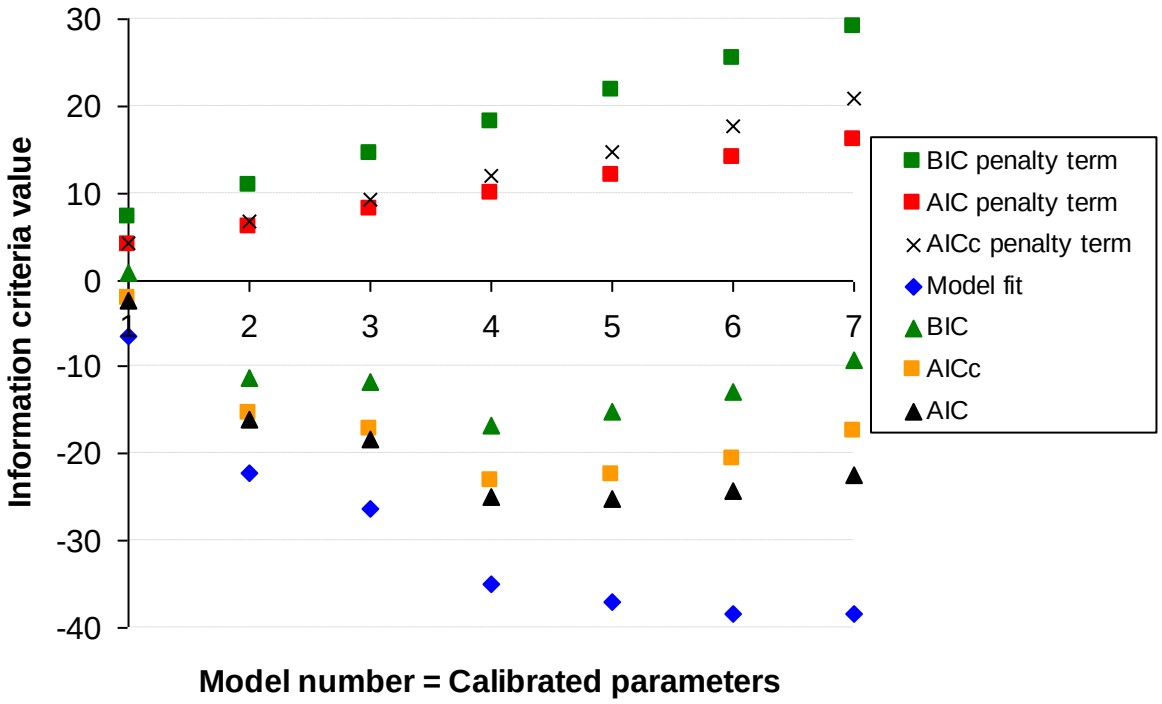

Figure 21: AIC_c, AIC, and BIC analysis of models calibrated to 19 h and 19 K_h observations

The AIC_c plotted side by side for all four evaluated cases is presented in the following figure:

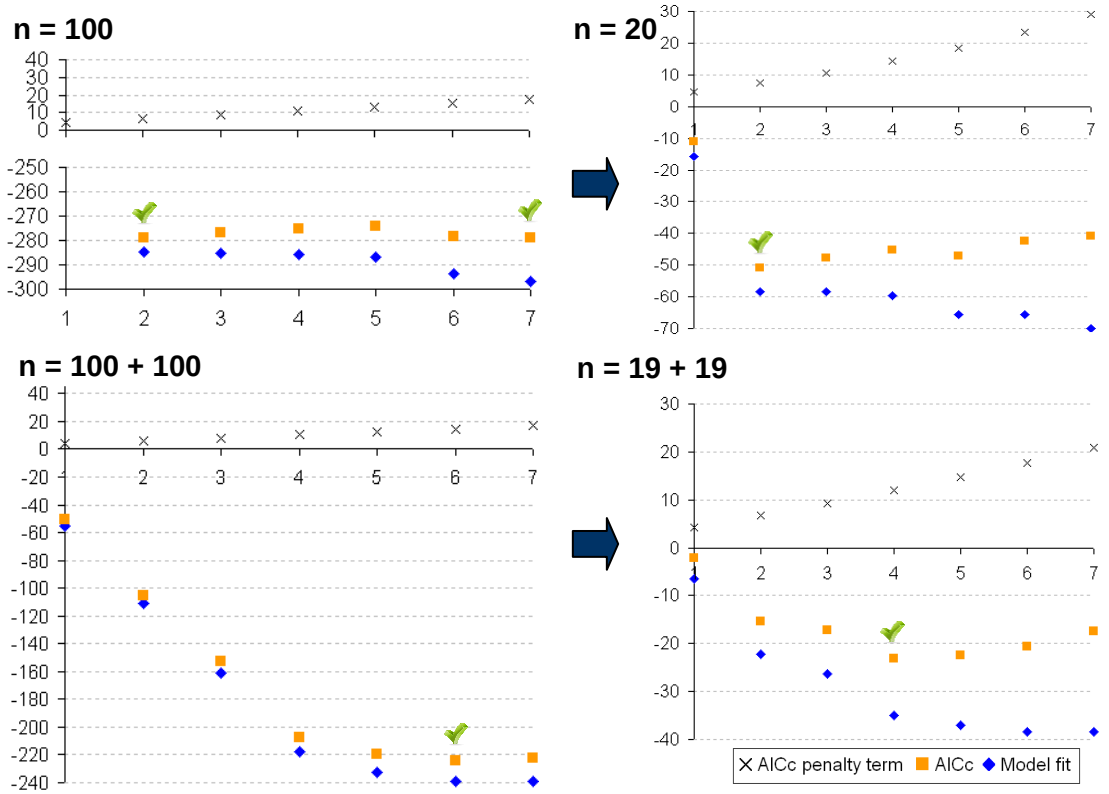

Figure 22: AIC_c of candidate models calibrated to different number and type of observations: 100 h, 20 h, 100 h + 100 K_h, and 19 h + 19 K_h observations. The green symbol indicates the best model(s) selected by AIC_c

When reducing the number of observations from 100 to 20 for models calibrated to just heads (upper graphs on above figure), the AIC_c selected with certainty the simplest (Model 2) of both possible best options (Model 2 and 7) as the optimal. The number of observations were reduced from 100 h plus 100 K_h in Section 4.4.2 (lower left graph on above figure) to 19 h plus 19 K_h observations (lower right graph on above figure) while everything else remained the same. As a result, we observed that the model selection criteria chose a model with four parameters instead of one with 6 as the optimal model. Even though reality is more complex, just a certain level of complexity can be attained with a limited amount of data. The more parameters one implements, the more the limited information spreads throughout them. Furthermore, the fewer the number of observations, the harder the AIC_c penalizes an increase in complexity. This causes AIC_c to prefer simpler models when fewer observations are available.

4.5.2 Impact of type of observations

The values of the calibrated parameter values did not change substantially by considering different numbers of observations (see table below). When calibrating to 100 or 20 heads, the values remained mostly on the same order of magnitude. The same occurred when calibrating to 100 h + 100 K_h or to just 19 % of these observations. The following table gives an overview of the resulting order of magnitude for the calibrated parameters when considering different type and number of observations, detailed values can be found in Table 13, 15, 19, 21, and in Figure 10.

Table 23: Order of magnitude of parameters calibrated to different types and numbers of observations and those of the data generating model for comparison

	20 h	100 h	19 h + 19 K_h	100h+100K_h	100h+100K_h + 100 Conc.	Data generating model
$K_{Lay\ 1\ without\ channel}$ (m/s)	10^{-2}	10^{-2}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
$K_{Lay\ 1\ channel}$ (m/s)	10^{-2}	10^{-2}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
$K_{Lay\ 2\ right}$ (m/s)	10^{-2}	10^{-2}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
$K_{Lay\ 2\ left}$ (m/s)	10^{-2}	10^{-2}	10^{-4}	10^{-4}	10^{-5}	10^{-5} - 10^{-6}
$K_{V\ North}$ (m/s)	10^{-7}	10^{-7}	10^{-8}	10^{-10}	10^{-9}	10^{-6}
$K_{V\ South}$ (m/s)	10^{-7}	10^{-7}	10^{-8}	10^{-8}	10^{-7}	10^{-7}
$Cond_{Riverbed}$ (m^2/s)	-	-	-	-	10^{-4}	10^{-3}

The calibrated parameter values which approximate the data generating model values more closely are those of the model which used all three types of information (penultimate column

of the above table), followed by the two sets of values which use two types of information. Farthest from the original values were both of the calibrated models which just use head values for calibration.

The type of information influences model structure. Every type of information reveals important model features. Head information clearly pointed out that a two layer model is needed, independently of either 20 or 100 observations were considered. However, this information would never have suggested that a paleochannel is a required model feature. In contrast the model of Section 4.4.3, where concentration information was available, obviously supported the presence of a paleochannel. In the same manner, the models calibrated to both: head and K_h observations, were able to identify the two layer models with K-zonations as the optimal models, independently of 100 h + 100 K_h or just 19 of each observation type were considered. As we have seen, the number of observations using these two sets of information would just affect the decision of how many K-zones are supported by the data.

Therefore, in order to improve the quality of the calibrated parameters, the diversity of information type is more important than a mere increase in number of a given observation type. Furthermore, information diversity also provides additional support for model complexity.

4.5.3 Impact of parameter order

In Section 4.4.3 (p. 51), Model 7 was ranked best with 100 % of certainty. However, the model fit curve (Figure 19) does not resemble that of Figure 2, where the biggest improvements occur at the beginning and then gradually diminish. In Section 4.4.3, the most important improvement happens first at Model 4 (new parameter: $K_{Lay\ 1\ channel}$) and there is no significant improvement until Model 7's new parameter $Cond_{Riverbed}$ appears. In Section 4.4.1 (p. 46), we have again a similar model fit behaviour (Figure 17, p. 48) where after an initial amelioration, no significant improvement occurs until the most complex models. This contrasts to the analysis made in 4.4.2 (p. 48) where there is a model fit curve (Figure 18, p. 50), similar to that of Figure 2 (p. 14) in which the improvement is gradually decreasing with model complexity. This suggests that the order in which the parameters are introduced may not be trivial. As a matter of fact, the important parameters, those for which the models react more sensitive to changes in parameter value, should come first. Therefore, a sensitivity analysis for the more complex model (where all parameters are calibrated) is performed and a ranking of parameters according to their sensitivities is obtained. With this information it

can be evaluated whether introducing first the most influential parameters provides a more parsimonious model selection ranking.

Sensitivity is a measure of the change in output (calculated values) with respect to a change in input (parameter values). To evaluate this measure, we can calculate sensitivity coefficients for each parameter at every observation value. These coefficients are formed by calculating the partial derivative of the output value (a model result like hydraulic pressure or concentration) with respect to the partial derivative of the parameter (Mc Elwee 1982, 1987):

$$sensitivity\ coefficien\ t = \frac{\partial\, output}{\partial\, parameter} \tag{52}$$

The sensitivity coefficient can be approximated with the forward, backward, or central difference methods (Hill and Tiedeman 2007a:47). In these methods, the parameter is changed by a small amount, which is as a rule of thumb between 1 and 5 % (Zheng and Bennett 2002:345). The sensitivity coefficient is the difference in model result between a reference parameter and either a slightly higher value (forward difference), a lower one (backward difference), or the average of these two (central difference) divided by the difference in parameter value. For example, the sensitivity coefficient approximated by forward difference method is calculated as:

$$sensitivity\ coefficien\ t \approx \frac{\Delta output}{\Delta\, parameter} = \frac{\hat{y}(p_{i+1}) - \hat{y}(p_i)}{p_{i+1} - p_i} \tag{53}$$

where $i = 1,...,$ number of parameter changes; $\hat{y}(p_i)$ and $\hat{y}(p_{i+1})$ are calculated values obtained at current parameter position i and at parameter p varied by a small amount $i+1$.

The units are those of the output values divided by those of the parameter. For comparison purposes the sensitivity coefficients can be normalized as follows so that their units are the same as the output values (Zheng and Bennett 2002:343):

$$rel.sensitivity\ coefficien\ t_i \approx \frac{\hat{y}(p_{i+1}) - \hat{y}(p_i)}{p_{i+1} - p_i} \cdot p_{reference} \tag{54}$$

The relative sensitivity coefficients can be further normalized to dimensionless relative sensitive coefficients. This can be achieved by multiplying the sensitivity coefficients by the ratio reference parameter to reference output value as follows:

$$rel.sensitivity\ coefficien\ t_i \approx \frac{\hat{y}(p_{i+1}) - \hat{y}(p_i)}{p_{i+1} - p_i} \cdot \frac{p_{reference}}{\hat{y}(p_{reference})} \quad 55$$

The reference values can be for example, the calibrated parameter and the calibrated model result. The above equation provides dimensionless relative sensitivity coefficients for every parameter at every observation. However, it can be more helpful to have a single coefficient as indication of the sensitivity of a given parameter value. This unification can be achieved with the sum of squares of all relative sensitivity coefficients for a given parameter value. The summation is performed over all observations. The reverse mathematical actions (to divide by the number of observations and take the square root) are also implemented for consistency. In this way, we obtain just one relative sensitivity coefficient for every parameter value, instead of *n* different ones:

$$rel.\ sensitivity\ coefficien\ t = \sqrt{\frac{\sum_{j=1}^{n} \left(rel.\ sensitivity\ coefficien\ t_j\right)^2}{n}} \quad 56$$

where *n* = number of observations, and *j* counts from 1 until *n*.

The results of the sensitivity analysis are not absolute, but influenced by the choice of variation percentage in parameter ($\Delta parameter$) and by the selection of the reference parameter values (Zheng and Bennett 2002:353). It is, however, an analysis tool which provides useful information for further uncertainty analysis, model development, and calibration.

To illustrate the sensitivity analysis procedure and its importance to the parameter order when building model complexity, we take the example of Section 4.4.1 (p. 46) where seven models were calibrated to 100 h observations with PEST assistance. For Model 7, we calculated the sensitivity coefficients of all of the seven parameters. The parameters are evaluated for a variation of 5 % and the calibrated values (see Table 13, p. 47) are taken as a reference.

The symbols in the graph below represent the relative sensitivity coefficients for all parameters of Model 7 (calculated with Equation 56). The higher the values, the more influential the parameters are. The points at 1.0 in the x-axis represent the sensitivities at calibrated values. The sensitivity coefficients are calculated in a 5 % step change in calibrated value until the range from 50 to 150 % of the calibrated values is covered.

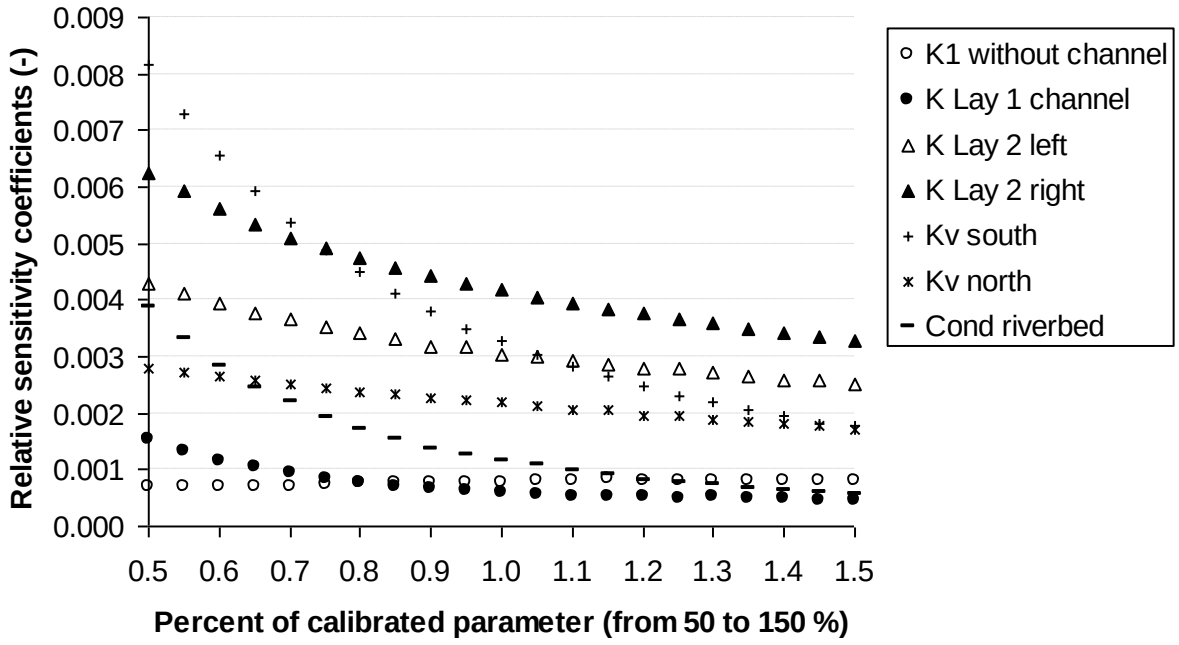

Figure 23: Relative sensitivity coefficients of Model 7 (100 h observations)

The above figure shows that the parameters $K_{Lay\ 2}$ (triangles) and K_V (+ and *) are more influential than $Cond_{Riverbed}$ (-) and $K_{Lay\ 1}$ (circles). Sometimes it is difficult to generalize which parameter is more important, since it depends on which range the values lay. In Model 7, the parameter $K_{V\ South}$ (+) gains on importance with smaller values than the calibrated value, while for higher values the $K_{Lay\ 2\ left}$ and $K_{Lay\ 2\ right}$ (triangles) are the most influential parameters. However, we can say that for most of the analyzed range, the most influential parameters are K_V and $K_{Lay\ 2}$ and the less influential parameter is $K_{Lay\ 1}$.

In the table below, we find an overview of the arrangements of calibration parameters in the sequence in which they were considered. In the first column, we have the original setup of Section 4.4.1 (p. 46). In the second column, the arrangement suggested by the sensitivity analysis (most influential parameters first, less influential at the end) is presented. Finally, the third column shows the best order which was found after trying several different combinations.

Table 24: Order of parameters: original (1st column), sensitivity analysis parameter ranking (middle column), optimal (last column); newly introduced parameters are marked bold

Model	Calibration parameters: $K_{Lay\ 1}$, $K_{Lay\ 2}$, K_V and $Cond_{Riverbed}$	Calibration parameters: K_V, $K_{Lay\ 2}$, $Cond_{Riverbed}$ and $K_{Lay\ 1}$	Calibration parameters: K_V, $Cond_{Riverbed}$, $K_{Lay\ 2}$ and $K_{Lay\ 1}$
1	**$K_{Lay\ 1}$**	**$K_{Lay\ 1}$**	**$K_{Lay\ 1}$**
2	$K_{Lay\ 1}$ and **K_V** ($K_{Lay\ 2}$ tied to $K_{Lay\ 1}$)	$K_{Lay\ 1}$ and **K_V** ($K_{Lay\ 2}$ tied to $K_{Lay\ 1}$)	$K_{Lay\ 1}$ and **K_V** ($K_{Lay\ 2}$ tied to $K_{Lay\ 1}$)
3	$K_{Lay\ 1}$, **$K_{Lay\ 2}$** and K_V	$K_{Lay\ 1}$, **$K_{V\ North}$** and **$K_{V\ South}$**	$K_{Lay\ 1}$, **$K_{V\ North}$** and **$K_{V\ South}$**
4	**$K_{Lay\ 1\ without\ channel}$**, **$K_{Lay\ 1\ channel}$**, $K_{Lay\ 2}$ and K_V	$K_{Lay\ 1}$, $K_{V\ North}$, $K_{V\ South}$ and **$K_{Lay\ 2}$**	$K_{Lay\ 1}$, $K_{V\ North}$, $K_{V\ South}$ and **$Cond_{Riverbed}$**
5	$K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 1\ channel}$, **$K_{Lay\ 2\ right}$**, **$K_{Lay\ 2\ left}$** and K_V	$K_{Lay\ 1}$, $K_{V\ North}$, $K_{V\ South}$, **$K_{Lay\ 2\ right}$** and **$K_{Lay\ 2\ left}$**	$K_{Lay\ 1}$, $K_{V\ North}$, $K_{V\ South}$, $Cond_{Riverbed}$ and **$K_{Lay\ 2}$**
6	$K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 1\ channel}$, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$, **$K_{V\ North}$** and **$K_{V\ South}$**	$K_{Lay\ 1}$, $K_{V\ North}$, $K_{V\ South}$, **$Cond_{Riverbed}$**, $K_{Lay\ 2\ right}$ and $K_{Lay\ 2\ left}$	$K_{Lay\ 1}$, $K_{V\ North}$, $K_{V\ South}$, $Cond_{Riverbed}$, **$K_{Lay\ 2\ right}$** and **$K_{Lay\ 2\ left}$**
7	$K_{Lay\ 1\ without\ channel}$, $K_{Lay\ 1\ channel}$, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$, $K_{V\ North}$, $K_{V\ South}$ and **$Cond_{Riverbed}$**	**$K_{Lay\ 1\ without\ channel}$**, **$K_{Lay\ 1\ channel}$**, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$, $K_{V\ North}$, $K_{V\ South}$ and $Cond_{Riverbed}$	**$K_{Lay\ 1\ without\ channel}$**, **$K_{Lay\ 1\ channel}$**, $K_{Lay\ 2\ right}$, $K_{Lay\ 2\ left}$, $K_{V\ North}$, $K_{V\ South}$ and $Cond_{Riverbed}$

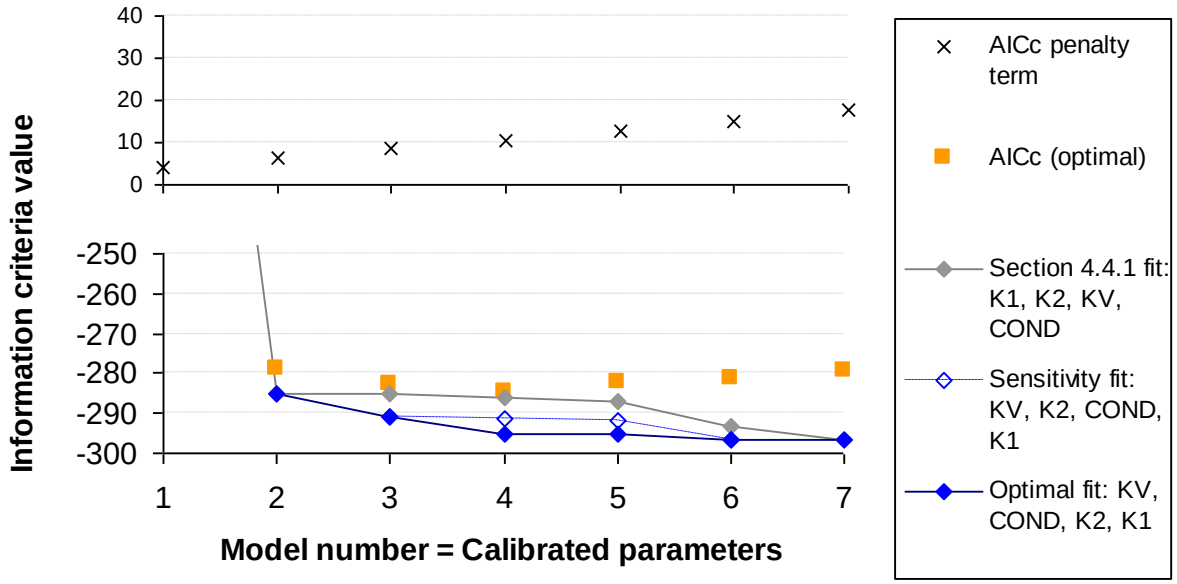

Figure 24: Model selection analysis of different parameter arrangements for models calibrated to 100 h observations (The AIC_c refers exclusively to the bottom blue diamonds)

In the model fit of Section 4.4.1 (gray diamonds in above figure), it is evident that there is no significant model fit improvement after Model 2 until Model 6 (where K_V is divided into zones) is introduced. $K_{V\ South}$ is the most influential parameter below 80% of the calibrated value (see "+" symbols in Figure 23). This suggests that K_V zonation should be introduced earlier as a calibrating parameter. The second arrangement of models (developed by taking into account the sensitivity analysis ranking) differentiates as suggested, first in Model 3, between the K_V parameters (instead of $K_{Lay\ 1}$), then, between the $K_{Lay\ 2}$ parameters, followed by adding $Cond_{Riverbed}$, and finally, discerning between $K_{Lay\ 1}$ zones. Models 1 and 2 are the same for all three arrangements. The model fit improves significantly (compare blue diamond at model number 3 with the gray one in the above figure). However, Model 4 (new parameter $K_{Lay\ 2}$) and Model 5 ($K_{Lay\ 2}$ zonation) bring just little improvement (white diamonds with blue outline on the above figure). Nevertheless, it improved after adding $Cond_{Riverbed}$ in Model 6. Adding this last parameter earlier (into Model 4; third arrangement) brought the lowest (best) AIC_c result (orange squares in the above figure). This best model has just one value of horizontal conductivity, two vertical conductivities and a flow resistance at the South boundary. It does not consider zonations at either upper or lower layer. The order of parameters plays an important role for the model ranking. By rearranging the parameters in such a way that the biggest improvements in model fit occur first, we could obtain a much lower value of AIC_c than the one previously obtained in Section 4.4.1 (p. 46), thus, a more parsimonious model.

Unexpectedly, the sensitivity analysis arrangement did not obtain the best results. A reason for this is that the sensitivity of Model 7 is not transferable to the simpler models. The sensitivity of a parameter depends also on the interaction with other parameters. For example, $K_{Lay\ 2}$ might just be very influential in presence of $Cond_{Riverbed}$. Also, the simpler (homogeneous) models have other parameters, since they lack zonations. For example, Model 2 has just one horizontal hydraulic conductivity parameter for both layers. In Model 7, this parameter is covered by four different hydraulic conductivity parameters. Another reason is that, as mentioned on page 61, the sensitivity value is not absolute and maybe an adequate interval or reference value was not chosen for a certain parameter. Hence, we were not looking at the meaningful value range.

The goal of Akaike's Information Criterion is to find the model having the best result with as few parameters as possible. In order to truly achieve this goal, we have to build up complexity with an optimal parameter arrangement, as accomplished here with the third arrangement (blue diamonds in above figure). However, the parameter ranking of the sensitivity analysis of Model 7 did not forecast the optimal complexity order. Trying all parameter combinations to find the optimal arrangement is not a practical approach, since

they are too many of them and have to be calibrated each time. Calibrating the models is usually a very difficult task in contrast to performing sensitivity analysis where just repeated forward simulation runs are performed while varying parameter values. It is our task to find the most influential parameters for each model since AIC does not find them for us. Burnham and Anderson (2002:62) note:

> *AIC is useful in selecting the best model in the set; however, if all the models are very poor, AIC will still select the one estimated to be best, but even that relatively best model might be poor in an absolute sense. Thus, every effort must be made to ensure that the set of models is well founded*.

The sets of models can be optimised by a sensitivity analysis which can help to design the candidate models already with optimal parameters. I suggest that a short sensitivity analysis of all parameters (or as many as possible) should be made at the initial stage of designing the candidate models. Since the models at this initial stage are not calibrated, we can use literature parameters or an initial guess with their respective model results as reference values for the Equations 54 and 55 (p. 60 and 61). Two runs for parameter changes (simply + or – 5%) might be enough to give us an idea about parameter sensitivity. It is not wasted time since such an analysis provides useful information which proves very valuable for calibration purposes. While performing calibration, it is important to have a “feeling” of how the model reacts to changes in parameters. This preliminary analysis will improve the quality of our candidate models. Despite sensitivity analysis the arrangement of the candidate models may not be optimised, this would mean that improvements do not decrease gradually. In this case, one can slip back the parameters that cause significant improvements and consider them within the less complex models as it was accomplished in Figure 24. Another approach is to apply a “stepwise” calibration. Several parameters that according to sensitivity analysis and modeler experience come into consideration are first tried out before selecting a new parameter; the one which provides the biggest improvement is chosen. In this manner, an optimal arrangement is usually guaranteed without trying all possible parameter combinations.

In this chapter, it was shown how to norm the different types of information using their respective observed standard deviations. Then, the synthetic data generating model and the candidate models were presented. Three different sets of synthetic data were generated, annexed in Appendix B.4 (p. 99). It was demonstrated that AIC can be applied to groundwater models using different types of information and depending on the amount and type of data selected for calibration, it chooses a model with certain complexity. Furthermore,

the impact of number of observations, types of observations, and order of parameters on model selection results was discussed. In the frame of this discussion, an evaluation using models calibrated with much less observations was performed. It was corroborated that AIC_c prefers simpler models when fewer observations are available, and the opposite is true for more data availability. It was also possible to prove that the use of different types of information for calibration also provides additional support for model complexity. Moreover, it was seen that for the quality of calibrated parameters, the diversity of information type is more important than the number of observations. Also the order of parameters plays an important role in model assessment. By rearranging the parameters in such a way that the biggest improvements in model fit occur first, we could obtain much lower values of AIC and AIC_c than with the original arrangement in Section 4.4.1, thus, a more parsimonious model was achieved.

In the next chapter we will apply AIC to real data using two different types of observations and we will optimally arrange the calibrated parameters to find the most parsimonious model.

5

MODEL ASSESSMENT WITH DIFFERENT TYPES OF OBSERVATIONS USING REAL DATA

Synthetic data can be biased since it is especially designed for a given purpose. A test with real data is thus necessary. In this chapter, the AIC is applied to different field measured data on a lysimeter model parametrized by Bärschneider (2008) in her hydrological-meteorological student project report for the Institute of Hydrology and Meteorology of the Technische Universität Dresden. Lysimeters are columns of soils which are carefully monitored to investigate the water mass balance at a given location. In order to accomplish this task, the water which infiltrates through the soil column is collected at the bottom of the column and measured. The column has an inbuilt scale which records a change in weight due to precipitation or evapotranspiration.

5.1 REAL DATA

The real data were measured at the lysimeter 10/2 located in Etzdorf (approx. 40 km westward from Dresden, Germany), which is part of the Brandis lysimeter station operated by the Saxon State Agency for Environment, Agriculture and Geology (LfULG). For model evaluation the lysimeter data set consists basically of two time series of daily values of actual evapotranspiration (ET) and seepage water (RU), shown in the figure below. The six months time-series start on October 1st, 1987 which corresponds to the start of the sowing period.

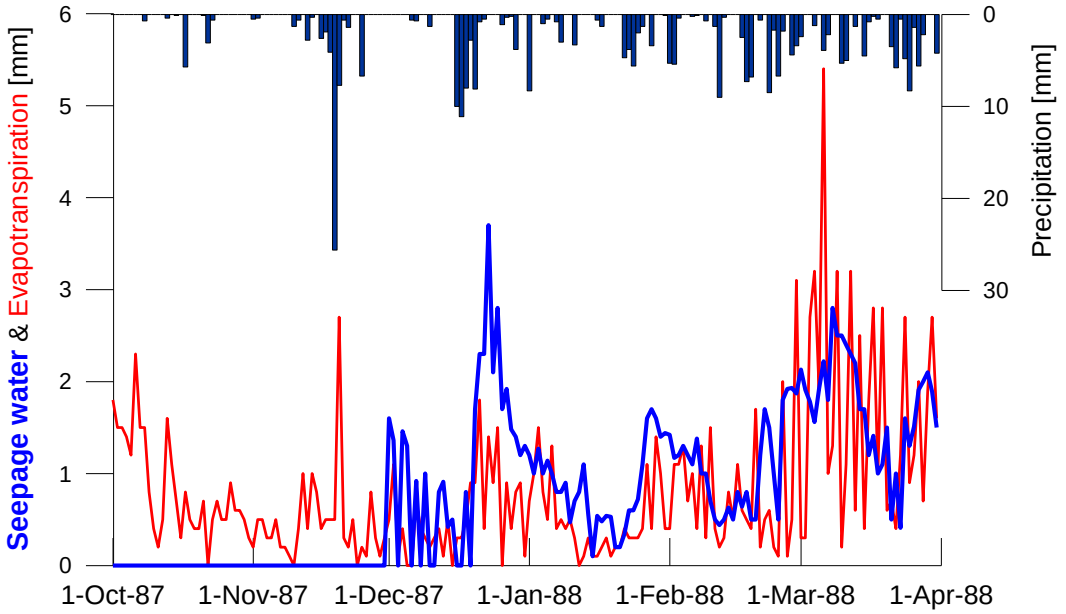

Figure 25: Real data time series of actual evapotranspiration, seepage water, and precipitation at lysimeter 10/2 from October 1st, 1987 until March 31st, 1988

Additional measured data were also available to Bärschneider for parametrization, such as precipitation values registered near the ground; reference potential evaporation; the lysimeter balance records, which register the change in weight of the soil column; the lysimeter soil description; land use schedule (with crop type, sowing and harvesting dates); and the crop leaf area index.

5.2 LYSIMETER MODEL

In order to simulate the lysimeter data, Bärschneider (2008) used the Richards equation (1931) without hysteresis to describe the one dimensional flow of water through a partially saturated porous medium:

$$\frac{\partial \theta}{\partial time} = \frac{\partial}{\partial x}\left[K_{saturated}(x) \cdot K_{relative}(h)\left(\frac{\partial h}{\partial x} + 1\right)\right] - Sink \qquad 57$$

and parametrized the soil hydraulic properties according to van Genuchten (1980) based on Mualem's model:

$$\theta(h) = \begin{cases} \theta_r + \dfrac{\theta_s - \theta_r}{\left[1 + |\alpha h|^n\right]^m} & h < 0 \\ \theta_s & h \geq 0 \end{cases} \quad 58$$

$$K_{relative}(h) = \theta_e^{Tortuosity}\left[1 - \left(1 - \theta_e^{1/m}\right)^m\right]^2 \quad 59$$

$$\theta_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} \quad 60$$

$$m = 1 - 1/n; \quad n > 1 \quad 61$$

where θ is water content [L^3L^{-3}]; x is the spatial coordinate [L] (positive upward); $K_{relative}$ is the relative [-] and $K_{saturated}$ the saturated hydraulic conductivity [LT^{-1}]; h is water pressure (head) [L]; θ_r and θ_s are the residual and saturated water content [-]; α, m and n are empirical parameters [1/L], [-], [-] and θ_e stands for effective water content [-].

To simulate the data, Bärschneider used the public domain program Hydrus 1-D version 4.03 (Simunek et al. 2005-2008), which numerically solves the Richards equation using the finite element method.

The lysimeter is 3 m long and with a 1 m^2 surface which during the period of interest was cultivated with winter wheat. The soil column is discretized in 101 nodes, which are 3 cm apart from each other. In the figure below the resulting 100 “cells” are outlined at the side of the column. In 1994, the lysimeter was cut and analyzed revealing three horizons of soil material: silty loam (material 1), mid-clayey silt (material 2) and sand (material 3). The first two horizons are mostly silt while the third one consisted mostly of sand.

Figure 26: Lysimeter 10/2 geometry and spatial discretization

For each of the three soil materials the following six parameters of the Mualem - van Genuchten hydraulic model must be parametrized:

- residual water content (θ_r);
- saturated water content (θ_s);
- the empirical parameters of the soil water retention function (α, n);
- saturated hydraulic conductivity (K_s); and
- tortuosity (l).

Bärschneider (2008) used a dynamically linked library (dll) included in Hydrus-1D named Rosetta to get estimations of K_s and water retention parameters (θ_r, θ_s, α, n; see following table). Rosetta, which was developed by Marcel Schaap at the U.S. Salinity Laboratory implements pedotransfer functions (PTFs) to predict these parameters. The PTFs were calibrated with soils from northern USA and northern Europe (Bärschneider 2008). The tortuosity parameter l is estimated as 0.5 by Mualem (1976).

Table 25: Mualem - van Genuchten parameters generated with Rosetta dll for the 3 soil materials

Material	θ_r [-]	θ_s [-]	α (mm^{-1})	n [-]	K_s (mm/d)	l [-]
silty loam	0.0630	0.4918	0.00036	1.7831	1871.9	0.5
mid-clayey silt	0.0833	0.5594	0.00053	1.6726	853.0	0.5
sand	0.0492	0.5429	0.00554	2.0581	5836.3	0.5

The initial value of the pressure head was estimated by Bärschneider (2008) as -100 mm at the top and gradually increased to 0 mm at the bottom.

The upper boundary condition is atmospheric pressure with surface water layer. This condition permits water to build up on the surface: *The height of the surface water layer increases due to precipitation and reduces because of infiltration and evaporation* (Simunek et al. 2008). The upper boundary condition is specified daily according to the values of precipitation and potential evapotranspiration. The lower boundary condition is Seepage Face (h = 0), which is used when the bottom of the lysimeter is exposed to the atmosphere, as it is the case here.

The variable boundary conditions are parametrized with daily data of precipitation, potential evaporation, and potential transpiration values. The absolute value of the minimum allowed pressure head at the soil surface (hCritA) was set by Bärschneider to 10 000 mm. The precipitation values are measured and the potential evaporation and transpiration were calculated by Bärschneider with the following equations from Mailhol et al. (1997):

$$E = (1 - C_P) ET_0 \quad (62)$$

$$T = C_P \cdot K_C \cdot ET_0 \quad (63)$$

$$C_P = 1 - e^{-\alpha_w \cdot LAI} \quad (64)$$

where E is the soil water potential evaporation [LT^{-1}], T is potential transpiration [LT^{-1}], C_P is a partition coefficient also known as surface cover fraction [-], ET_0 is the reference potential evaporation measured at the lysimeter station [LT^{-1}], LAI is the measured leaf area index [-], K_C is the crop coefficient [-] (see Table 26), and a_w an empirical crop parameter with a value of 0.75 [-] taken from Mailhol et al. (1997). Hydrus simulates the actual evaporation and transpiration according to water availability based on the given potential E and T (Simunek et al. 2008).

Table 26: K_C coefficient for winter wheat in Europe (Allen et al. 1998:Chapter 6)

	Start phase	Growth phase	Maturation phase	Final phase
Days	160	75	75	25
K_C	0.70	0.70 - 1.15	1.15	0.40

Most of the simulation time performed by the model (182 days) comprises just the K_C start phase which lasts 160 days and begins on October 16th, 1987. When analysing the cut lysimeter in 1994, the maximal depth of the roots was found to be 160 cm. The root distribution gradually increased attaining its maximum at 120 - 150 cm depth. Bärschneider (2008) selected the root water uptake model according to Feddes et al. (1978) and a critical stress index for water uptake of 1. The advantage of choosing the Feddes et al. model is that Hydrus-1D includes for this model a database of water uptake parameters for different crops. The parameters for winter wheat are shown in the following table:

Table 27: Root water uptake values for winter wheat after Wesseling et al. (1991)

Parameters	(mm)
Pressure head below which roots start to extract water (P0)	0
Pressure head below which roots start to extract water at maximum rate (P0pt)	-10
Limiting pressure head, below which roots cannot longer extract water at max. rate, assuming r2H potential transpiration rate (r2H below in this table) (P2H)	-5000
Limiting pressure head, below which roots cannot longer extract water at max. rate, assuming r2L potential transpiration rate (r2L at the end of this table) (P2L)	-9000
Pressure head below which root water uptake ceases (wilting point) (P3)	-160000
Potential transpiration rate per day (r2H)	5
Potential transpiration rate per day (r2L)	1

The values for time discretization, iteration criteria, time step control, and the interval for internal interpolation tables are shown in Table 28. The time step control parameters and the pressure head tolerance are those recommended and given as default by Hydrus-1D. The initial time step value of 0.001 d is used while the number of iterations remains in the optimal iteration range (between 3 to 7 iterations; table below). Out of this range, the initial step is increased or decreased as needed by multiplying it by the lower or upper time step multiplication factor (Simunek et al. 2008).

Table 28: Values for time discretization, iteration criteria, time step control, and internal interpolation tables interval

Time discretization (days)	
Initial Time	0
Final Time	183
Initial Time Step	0.001
Minimum Time Step	1×10^{-5}
Maximum Time Step	5
Iteration Criteria	
Maximum Number of Iterations [-]	10
Water Content Tolerance [-]	0.001
Pressure Head Tolerance (mm)	10
Time Step Control [-]	
Lower Optimal Iteration Range	3
Upper Optimal Iteration Range	7
Lower Time Step Multiplication Factor	1.3
Upper Time Step Multiplication Factor	0.7
Internal Interpolation Tables (mm)	
Lower Limit of the Tension Interval	1×10^{-5}
Upper Limit of the Tension Interval	100 000

The internal interpolation tables (last two parameters of above table) define a range of pressure heads in which the hydraulic properties are internally interpolated from tables. The tables are of water contents, hydraulic conductivities, and specific water capacities from the specified set of hydraulic parameters which Hydrus generates for each soil type at the beginning of a numerical simulation (Simunek et al. 2008). Bärschneider (2008) set an interval big enough to encompass the entire range of pressure heads since the interpolation is much faster than calculating each time the values with hydraulic functions.

The uncalibrated model parametrized by Bärschneider provides the following time series (plotted against original data for comparison):

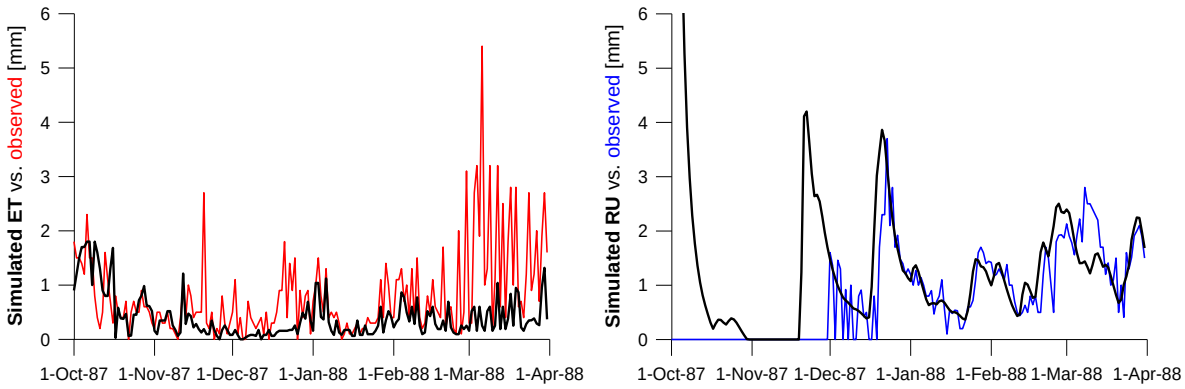

Figure 27: Uncalibrated model results of ET and RU (black line) vs. observed values (in color)

As can be observed in the above figure, only the ET values for the first six weeks are very similar to the observed values. The rest of the time series reveals a need for calibration. On the other hand, the seepage water (RU) simulation results follow closely the observed values with exception of the first two months of the time series. During this period, the simulated RU results are strongly affected by the initial values of pressure head distribution, which were assumed to be evenly distributed from -100 mm at the surface to 0 mm at the bottom of the lysimeter. A model with optimal initial values saves simulation time since it needs less iterations to converge. Therefore, it is recommendable to provide better initial pressure head values. The final pressure head distribution ranged from -2851.354 mm at the surface to 0 mm at the bottom. Since the initial values are not exactly the final values, a reasonable guess is to set the initial head values evenly distributed from -3000 mm at the surface to 0 mm at the bottom of the lysimeter. This provided the following simulated time series:

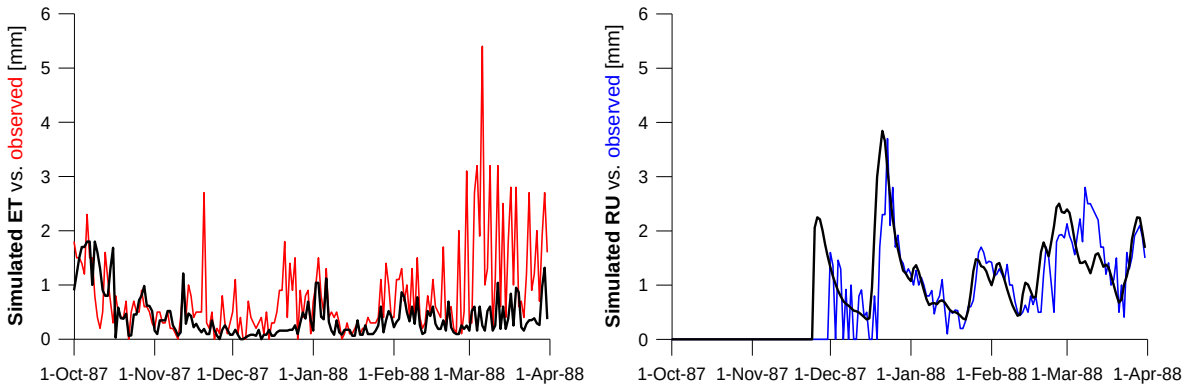

Figure 28: Uncalibrated model results of ET and RU (black line) vs. observed values (in color) with initial head values evenly distributed from -3000 mm at the surface to 0 mm at the bottom of the lysimeter

As it can be seen in Figure 28, there is no change in ET simulated values. However, the change in initial head pressure distribution positively impacted the simulated RU values during the first two months. The simulation time decreased by 80 % (from 11.50 s to 2.25 s).

5.3 SENSITIVITY ANALYSIS

The uncalibrated model presented in the previous section may not be parsimonious (providing good results while keeping the number of calibrated parameters as low as possible). To find out whether it is parsimonious, we have to gradually introduce some complexity and evaluate the improvement with AIC. It was demonstrated in Section 4.5.3 that the order in which the parameters are considered influence the choice taken by AIC. The sensitivity analysis provides hints on which could be the best order of complexity. 30 calibratable parameters were found in the model. They are listed in the following table:

Table 29: Calibratable parameters of the lysimeter model

Category	Parameters
Discretization	Time step
	Number of nodes
Initial conditions	Initial pressure head
Boundary cond.	Absolute value of the minimum allowed pressure head at the soil surface (hCritA)
Soil	Water retention parameters (θ_r, θ_s, α, n) and K_s for each of the three soil materials
	Tortuosity
Plant	K_C
	a_w
	Critical stress index for water uptake
	P0, P0pt, P2H, P2L, P3, r2H, r2L

Equation 54 (p. 60) can be used to obtain relative sensitivity coefficients for every observation since both types of information have the same units. To apply Equation 54, forward runs with the parametrization of Bärschneider as reference, and with parameter variations of plus and minus 5 % were performed. To summarize all of the resulting coefficients into just one single coefficient for each parameter, we use Equation 56 (p. 61). The results are plotted on the graph below. A difficulty was encountered with parameter P0 which has a value of zero. Since we cannot use zero as reference value in Equation 54 to norm the sensitivity coefficients or to determine a small variation, the value of a very close

parameter was taken instead. The related parameter is P0pt with a value of -10 mm, which makes a 5 % variation of -0.5 mm.

Along the y-axis in Figure 29, the relative sensitivity coefficients derived from the difference in results by a 5 % parameter change (x-axis) are plotted. The first values along the x-axis represent the relative sensitivity coefficient considering a reduced parameter value of 95 % of the original value. The values in the middle of the x-axis analyze the impact by using 105 % of the original parameter value. Finally, the last values in the x-axis are the relative sensitivity coefficient by considering a 5 % increment to a parameter value of 110 %.

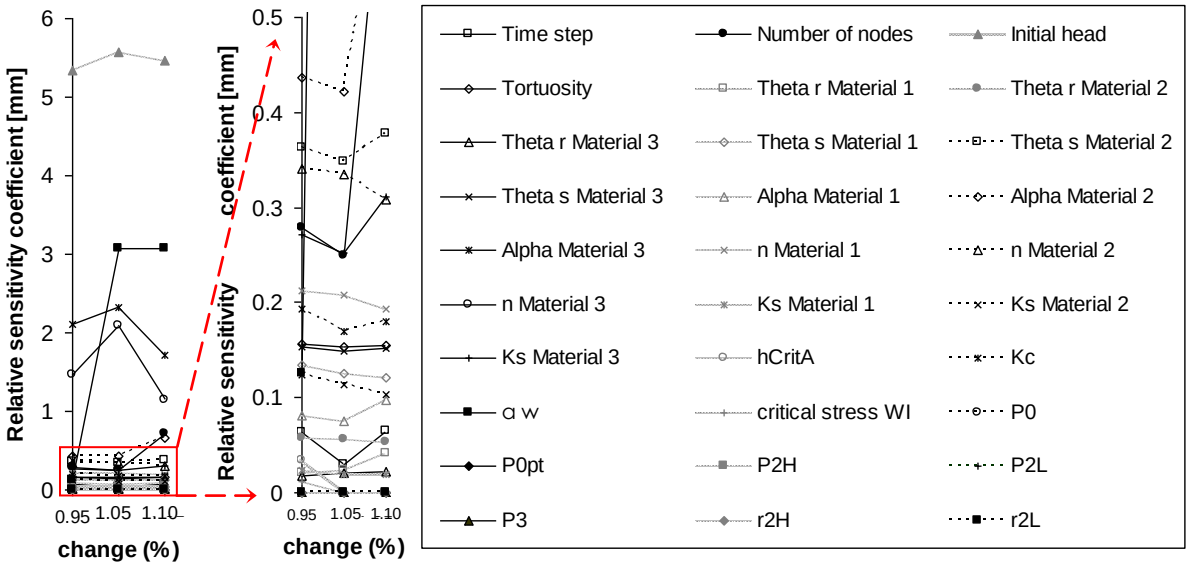

Figure 29: Relative sensitivity coefficients of the lysimeter model parameters

There are four very influential parameters which are (in order of sensitivity): initial head, a_w, $\alpha_{material\ 3}$, and $n_{material\ 3}$. They have relative sensitivity coefficients greater than 1 mm. Not too influential are the five parameters which range from 0.25 to 1mm. These are $\alpha_{material\ 2}$, $\theta_{s\ material\ 2}$, $n_{material\ 2}$, number of nodes, and $K_{s\ material\ 3}$. The rest of the 21 parameters have by a 5 % change of the model value not much impact on model results. From these, just six parameters ($n_{material\ 1}$, K_C, tortuosity, $\theta_{s\ material\ 3}$, $\theta_{s\ material\ 1}$, and $K_{s\ material\ 2}$) have minor relative sensitivity coefficients in the range of 0.1 to 0.25 mm. The time step and five soil parameters (θ_r of all three materials, K_s, and α of material 1) were almost unimportant with relative sensitivity coefficient values lower than 0.1 mm. The seven root water uptake parameters, the critical stress index for water uptake, and the hCritA were unimportant by a 5 % change of the model value.

5.4 MODEL ASSESSMENT

In order to calculate Akaike's Information Criterion considering different model results as it is the case here, the data have to be normed as already explained in Section 4.1. By norming the observed and calculated values of evapotranspiration and infiltration with their standard deviation, we obtain values which are unitless and within the same order of magnitude. The evapotranspiration standard deviation (σ_{ET}) norms the different sets of observed and calculated evapotranspiration. The infiltration standard deviation (σ_{RU}) does the same with the observed and calculated infiltration data sets. The two standard deviation values (σ_{ET} and σ_{RU}) are calculated from their respective observed data sets centered at zero. The residuals obtained from the normed values can be then analyzed as a single data set.

The ratio n / K is 61 (= 366 / 6). A ratio bigger than 40 indicates that we can rank the models using AIC and still provide the same result as with the slightly more complex corrected version (AIC_c). The results are summarized in the following graph and table:

Table 30: AIC evaluation results by adding complexity as suggested by the sensitivity analysis

Model				AIC			
Nr.	Param.	Fit	K	Penalty	AIC	Δ_i	w_i
0		-72.6	1	2	-70.6	173.3	0.00
1	ini. h	-129.2	2	4	-125.2	118.7	0.00
2	$+\alpha_w$	-129.3	3	6	-123.3	120.5	0.00
3	$+\alpha_3$	-161.2	4	8	-153.2	90.7	0.00
4	$+\alpha_2$	-186.4	5	10	-176.4	67.5	0.00
5	$+K_c$	-255.8	6	12	-243.8	0.0	1.00

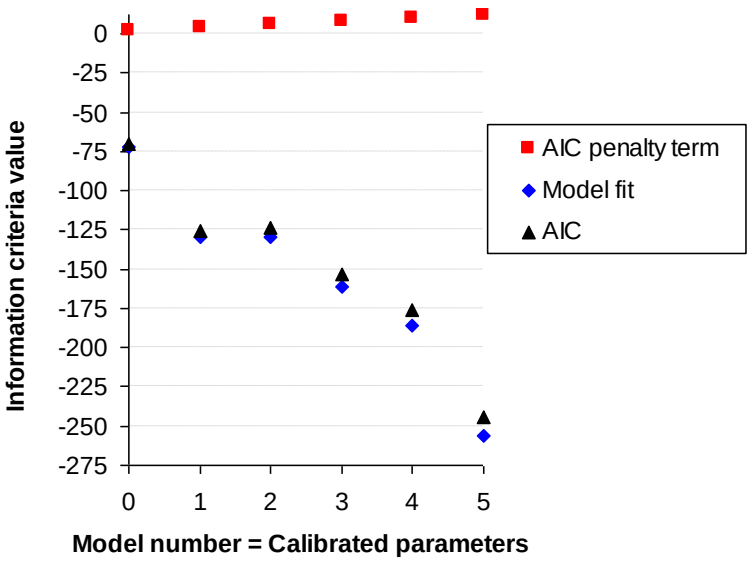

Figure 30: AIC evaluation adding model complexity as suggested by the sensitivity analysis

By calibrating the model using the parameters in the order of the sensitivity analysis, we get the biggest improvement using all 5 parameters together. Out of this 5 calibration scenarios, we can tell that number 5 is chosen by AIC as the best option with 100% certainty. However, the calibration with two parameters provided only little improvement. Also, the improvements by adding the soil parameters α_3 and α_2 are not as big as in the last calibration where the plant parameter was added. These discrepancies are an indication that the different parameter combinations are not optimally chosen. To find the optimal combinations of parameters, we try several parameters first. The one which provides the biggest improvement is chosen. We repeat this for all of the parameters and by doing so we obtain the following results:

Table 31: AIC evaluation results by adding complexity stepwise

Model				AIC			
Nr.	Param.	Fit	K	Penalty	AIC	Δ_i	w_i
0		-72.6	1	2	-70.6	177.0	0.00
1	$+K_c$	-186.1	2	4	-182.1	65.4	0.00
2	$+\alpha_2$	-240.1	3	6	-234.1	13.4	0.00
3	$+\alpha_w$	-255.6	4	8	-247.6	0.0	0.64
4	$+\alpha_3$	-255.8	5	10	-245.8	1.8	0.26
5	ini. h	-255.8	6	12	-243.8	3.8	0.10

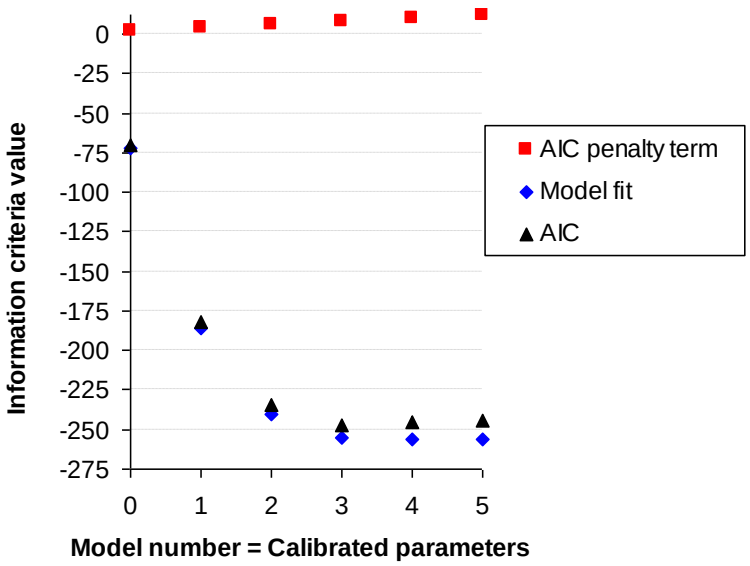

Figure 31: AIC evaluation by adding complexity stepwise

The most parsimonious model is achieved by calibrating stepwise, in this case with 3 parameters: first with the plant parameter K_c, then with the soil parameter α_2 and finally, using the plant parameter α_w. Calibrating with more parameters does not improve the AIC.

Model 3 results can be seen in the figure below. Observed evapotranspiration is in red; observed infiltration is in blue; and the simulation results are in black:

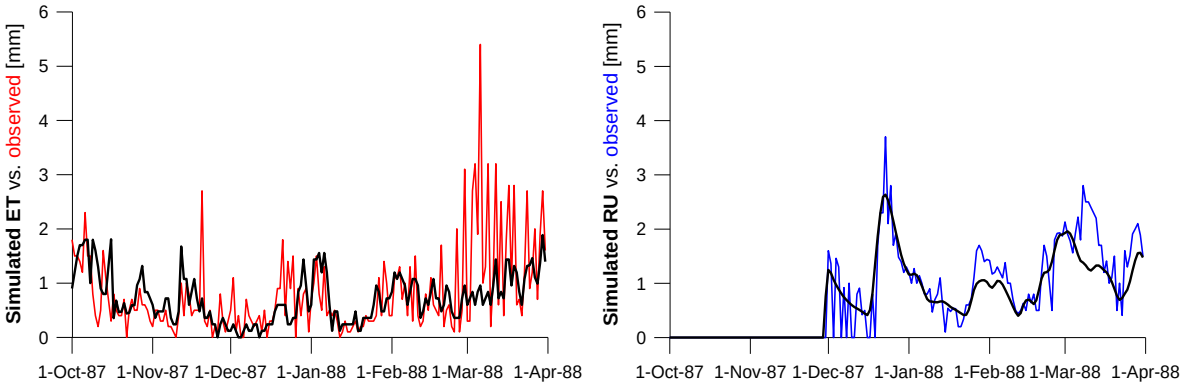

Figure 32: Calibrated optimal fit of the parsimonious model

The parsimonious model did not exactly match the observed values. Nevertheless, the calculated ET values fit the observed data much better than the uncalibrated results (Figure 28, p. 74), especially during winter and spring.

5.5 DISCUSSION AND CONCLUSION

In this chapter, the AIC application to real data using different kinds of observations proves the viability of this method to address uncertainty assessment on hydrogeological models. The real data and the model are not biased since data were not created by the modeler. Likewise, the model itself was neither created nor parametrized by the model selection evaluator.

Complexity was slowly built in, one parameter at a time, calibrating first the most influential parameters. As occurred in Section 4.5.3, the sensitivity analysis did not reveal the optimal order of the parameters. Nevertheless, it was useful to distinguish between parameters which were probably worth trying from those which were not.

By calibrating the model stepwise, for the first parameter the highest improvement was achieved with the plant parameter K_c. I would like to point out that both plant parameters (K_c and α_w) brought a better fit as first choice parameter than any other tested soil parameter. This is probably due to the fact that evapotranspiration values have a worst fit to observed data than infiltration values; thus, having more potential to improve. A soil parameter, α_2, brought the best improvement when selected as second parameter. In this calibration step, all of the tested soil parameters achieved a bigger improvement in model fit as the plant

parameter α_w. Yet, as a third parameter, α_w had more impact than the rest of the soil parameters. This calibrating behaviour reveals that more important than choosing a given parameter is to choose a parameter from a certain group, for example, after calibrating a plant parameter it is recommended for the next calibration step to select a soil parameter.

The most efficient way to find a parsimonious model, which provides the best fit with the lowest amount of calibrated parameters, is to perform a stepwise calibration trying at each step influential parameters of different parameter groups.

6

CONCLUSION AND OUTLOOK

One of the biggest uncertainties in hydrogeological modeling, which can be minimized by applying AIC, is the choice of the conceptual model itself. The uncertainty assessment based on Information Theory (AIC, explained in Section 3.3) ranks models calibrated to the same data by favouring them for fitting observed values and penalizing them for building additional model complexity. Hence, AIC can be regarded as a balance equation between closeness to the truth and model complexity. The information criterion developed by Akaike has gained on popularity due to its practical simplicity. Yet, it is based on deep statistical theories originated from the Information Theory. AIC can estimate a relative information loss between the unknown truth and a given conceptual model based on the maximum likelihood function.

It is not always easy to apply model selection methods to hydrogeological models, in particular, when using different sources of information (see implementation errors documented in Section 3.8). Different types of observations have different order of magnitudes and types of units. However, it is possible to norm observed and simulated values by their respective observed standard deviations, as described in Section 4.1. This provides a single data set of normed observations which can be analyzed by AIC.

The application of AIC was successfully tested on groundwater models using three different sets of synthetic data: hydraulic pressure (h), horizontal hydraulic conductivity (K_h), and

tracer concentration. Several conceptual models were proposed to simulate these data, each of them with increasing number of parameters to be calibrated (Section 4.3). The conceptual models were calibrated first to one data set (h), then to two data sets (h plus K_h), and finally, to all of the three data sets. The results of the three model analyses (Section 4.4) revealed a need for investigation of the influence of number of observations, type of information considered, and order of calibrated parameters, in model selection results (Section 4.5).

The investigation on data availability (Section 4.5.1) revealed that the number of observations determine how complex a model can be. The fewer the number of observations, the harder the AIC penalizes an increase in complexity. This causes AIC to prefer simpler models when fewer observations are available.

Also information diversity allows for further complexity in parsimonious models. Every type of information reveals important model features. For example, the models calibrated with just heads (Sections 4.4.1 and 4.5.1) clearly pointed out the need of a two layer model, disregarding whether 100 or just 20 observations were considered. Models calibrated additionally to K_h values (Sections 4.4.2 and 4.5.1) provided information about K-zonations also independent of the number of observations considered. Concentration data supported the presence of a Paleochannel structure (Section 4.4.3). Hence, more complex models can be selected by AIC if different types of observations are used for calibration. Furthermore, the quality of the calibrated parameters improved, the more different types of observations were used for calibration (Section 4.5.2).

The analysis made on the order of the calibrated parameters revealed that the order in which complexity is built up is essential to attain a truly parsimonious model (Section 4.5.3). Calibrated parameters that provide bigger improvements in model fit should be first taken into consideration. This can be achieved, by either trying all possible combinations of parameters or by combining a sensitivity analysis with a stepwise calibration. However, trying all parameter combinations to find the optimal arrangement is not a practical approach, since there are too many of them and they have to be calibrated each time. In contrast, performing sensitivity analysis does not require calibration, therefore, it is a much easier procedure which provides valuable information for an effective calibration. The stepwise calibration consists of several steps in which several parameters are tried out and the one which provides the biggest improvement is selected. The procedure is repeated with the rest of the parameters until a desired complexity level is attained or AIC has been minimized.

As synthetic data was created for the specific purpose of testing the previous models and as this step was performed by the same person who also developed these models, it can be argued that the procedure is biased. Applying AIC with different types of information was thus also verified with unbiased lysimeter data. The model was parametrized by Bärschneider (2008); the real data (actual evapotranspiration and seepage water) was measured from the Brandis lysimeter station operated by the Saxon State Agency for Environment, Agriculture and Geology (LfULG) near Dresden, Germany. Using this model and these data, it was demonstrated that AIC can be applied in real hydrogeological studies using different types of information, hence, reducing uncertainty. AIC found an optimal lysimeter model using a stepwise calibration which slowly built up model complexity. A sensitivity analysis provided hints on which parameters may be worth trying to obtain the biggest model fit improvements happening first. Although, the resulting parsimonious model did not exactly matched the observed values, calibration with only three parameters substantially improved the original model fit.

To further demonstrate the advantages of conceptual model selection analysis, it would be necessary to test AIC using different types of observations in more real cases. This will also motivate modelers to apply the here presented analyses to their hydrogeological models. Moreover, the implementation of this uncertainty assessment as automated or semi-automated algorithms will contribute to simplify its practical application and extend its use among the hydrogeological modeling community. These algorithms should calculate normalized relative sensitive coefficients and also be able to calibrate several parameters in automated form to different types of observations. Furthermore, implementing the stepwise calibration with AIC_c model selection ranking as an automated algorithm would provide a very useful model selection tool.

Modelers are divided in two groups: first, those who see in stochastic modeling the tool for optimal aquifer complexity description and proper model prediction and second, those who favor the use of parsimonious models. At the session of “complexity versus simplicity” of the MODFLOW2003 conference, a discussion between both groups took place and they concluded that both approaches had their advantages and disadvantages (Gómez-Hernández 2006). With help of the above mentioned automated or semi-automated tools, an approach could be developed to minimize conceptual model uncertainty by applying AIC using different types of information to stochastic resulting models. With such a parsimonious stochastic model selection procedure, it may be possible to reconcile these two groups of modelers.

This research proved the benefits of combining the application of AIC in hydrogeological models with the stepwise calibration using several types of observations. It is hoped that the presented dissertation will stimulate hydrogeologists to address conceptual model uncertainty using this innovative uncertainty assessment method based on Information Theory.

APPENDIX

A. BASIC MATHEMATICAL ASPECTS TO UNDERSTAND AIC

Akaike estimated the “K-L information minus a truth dependent constant” from given data. He arrived to a result involving the likelihood function. This was possible by using the maximum likelihood method to estimate the model parameters during the derivation of the solution. Therefore, in order to have a better understanding of Akaike’s Information Criterion, it is also important to draw attention to the maximum likelihood estimation method and the likelihood function.

A.1 ESTIMATING PARAMETERS WITH THE MAXIMUM LIKELIHOOD ESTIMATION METHOD

Maximum likelihood estimation (MLE) is a standard parameter estimation method, commonly applied in statistical inference. It estimates coefficients from a given data set, assuming a certain probability distribution function.

In natural sciences, data is in the best of cases just a random sample taken out of all of the existing values. According to Myung (2003), all of these existing values could be represented

statistically with a probability distribution which has some parameters associated with it. If the values of these parameters are changed, a different probability distribution is generated. All of these probability distributions together are called a *model*; and a certain value of the parameters gives the most likely approximation to the data.

To understand MLE, it is important to recall the *probability* concept. For this purpose, we take the easiest example which is tossing a coin. For an unbiased coin, the probability to get head or tail is 0.5. For this case, the model, the data, and the parameters are given. The interest lies on finding the probability of observing a certain event. Burnham and Anderson (2002) remind us that the "model" of this problem is the binomial probability function $g($data: n, k | parameter: p; model: *binomial)*,

$$g(n,k \mid p; binomial) = \binom{n}{k} p^k (1-p)^{(n-k)} \qquad \text{where } \binom{n}{k} = \frac{n!}{(n-k)!k!} \qquad 65$$

where n is the number of trials, k the number of successes, and parameter p is the probability of success. For example, for the case of getting 3 heads out of 10, the calculation is as follows

$$g(3 \mid 10,\ 0.5; binomial) = \left(\frac{10!}{7!3!}\right) 0.5^3 \cdot 0.5^7 = 0.117 \qquad 66$$

There is approx. 12 % probability of getting 3 heads out of 10 coin flippings. Let us vary the number of success k = 0, 1, …, 10; n and p remain fix. The results are shown in Figure 33:

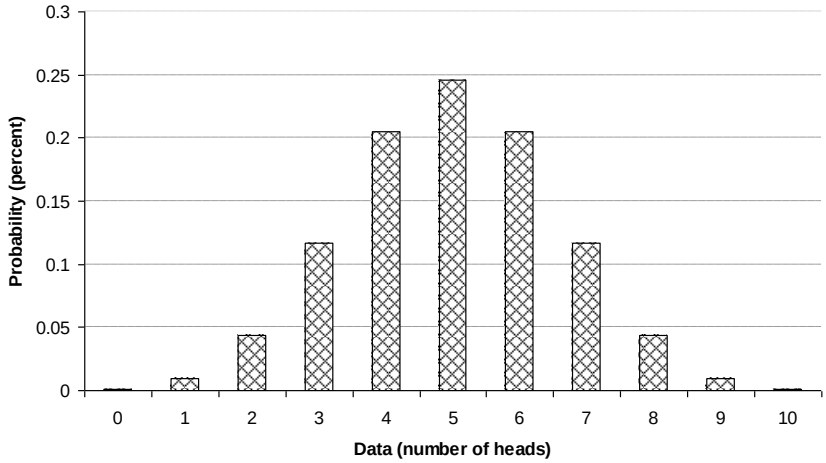

Figure 33: Probability of having a certain number of head outcomes out of 10 coin flips

The probability is on the y axis for the number of heads obtained out of 10 coin flips (x axis). It is easy to see that the highest probability is 5 heads out of 10 and that 3 heads out of 10 has a probability of 0.117 (approx. 12 %).

Now, in the case of the *likelihood* $L(p \mid n, k; binomial)$, the model and the data are also given. The interest lies in estimating the parameters. Usually, in natural sciences we observe some data, assume a model and try to estimate the parameters. The formula as Massmann (2004) remarks is the same as the probability function, just our assumption changes. Before, we calculated the probability of an event given parameters and data. Now, we calculate the likelihood of the parameters given the data and event. For example, assuming a fix data of 3 heads out of 10 coin flips, which would be the most likely value of the parameter p? The answer is given by the likelihood function of the binomial model:

$$L(p \mid 10,\ 3; binomial) = \left(\frac{10!}{7!3!}\right) p^3 (1-p)^7 \qquad (0 \le p \le 1) \qquad 67$$

This function is plotted in the following diagram:

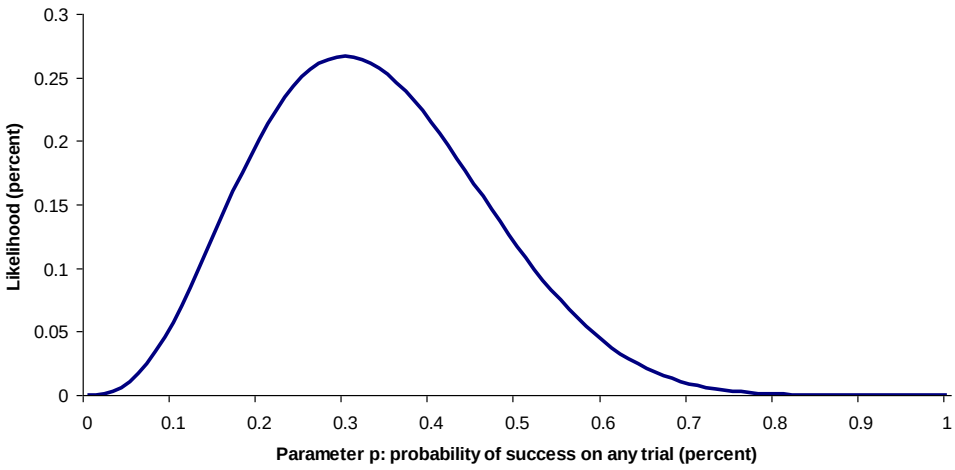

Figure 34: Likelihood function of having 3 heads out of 10 coin flips

The likelihood function plotted above tells us that the most probable value for the parameter p (to get the result of 3 out of 10 events) would be 0.3, meaning an unbalanced coin which tends to give head just 30 % of the times. It also shows that if the parameter had a 0.5 probability of success, our result of 3 heads out of 10 coin flips would have approx. 12 % of probabilities to occur. As Myung (2003) points out, both graphs are not comparable since they have different x axis. In the probability function, we are varying the data, while in the likelihood function we vary the parameter. Notice that the graph is a simple curve. If we would be estimating two parameters, then the plot would be a 2D surface. If there were x number of parameters to

estimate, then the result would be an x-dimensional geometrical figure sitting above an x-dimensional hyperplane spanned by the parameter vectors $p = (p_1, \ldots, p_x)$ (Myung 2003).

The exact value of the parameter is found by maximizing the likelihood function. The likelihood function provides values between 0 and 1. *Zero* means it is impossible that the data is created by the model while *one* means a complete certainty that the data is created by the model. This might be simple, straightforward and clear but seldom applied. As Hélie (2006) explains, the result of multiplying numbers bonded between [0, 1] never increases. Hence, the likelihood is always a very small number which can result in an underflow of modern computers (the smallest representable number is usually 1×10^{-7}). The solution to the underflow is to use the logarithmic transformation. The logarithm is a monotonic increasing function. Products in the logarithm become sums. The ln-likelihood function is bounded between $[-\infty, 0]$ where minus infinity represents certainty that the model did not generate the data and zero representing absolute certainty that the model generates the data. This seems more practical for computers. The value is always a negative number which could be seen as counterintuitive for an error function. Therefore, the negative of the ln-likelihood is often used. This function is bounded from $[0, \infty]$; zero representing absolute certainty that the model generates the data and infinity representing certainty that the model did not generate the data (Hélie 2006). Let us then apply the natural logarithm to our likelihood example to avoid underflow,

$$\ln L(p \mid 10,\ 3; binomial) = \ln\left(\frac{10!}{7!3!}\right) + 3\ln\ p + 7\ln(1-p) \qquad (0 \le p \le 1) \qquad 68$$

The natural logarithm of our likelihood function is described in the following graph:

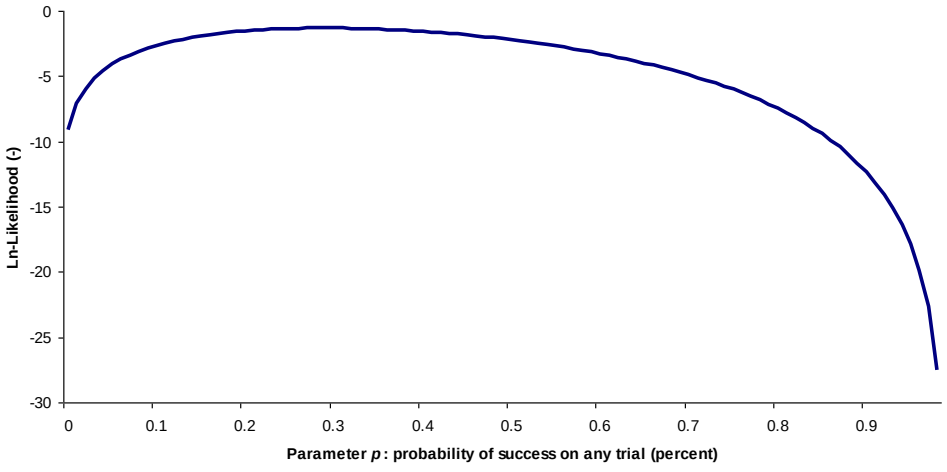

Figure 35: Ln-likelihood of having 3 heads out 10 coin flips

The curve in the figure has another form but the peak remains at the same position. Maximizing this function (finding the peak) is the same as maximizing the function without logarithmic transformation. We can find the maximum likelihood estimate by solving the derivative with respect to the parameter of the ln-likelihood function when it equals zero,

$$\frac{\partial \ln L(p \mid n,k)}{\partial p} = 0 \tag{69}$$

$$\frac{\partial \ln L(p \mid 10,3)}{\partial p} = \frac{3}{p} - \frac{7}{1-p} \tag{70}$$

Evaluated at zero p_{MLE} = 0.3. The notation MLE as subscript denotes the Maximum Likelihood Estimate. To make sure that the solution represents a maximum and not a minimum, the second derivative has to be calculated and evaluated at $p = p_{MLE}$ (Myung 2003),

$$\frac{\partial^2 \ln L(p \mid n,k)}{\partial p^2} = -\frac{3}{p^2} - \frac{7}{(1-p)^2} < 0 \tag{71}$$

$$\frac{\partial^2 \ln L(p \mid 10,3)}{\partial p^2} = -\frac{3}{0.3^2} - \frac{7}{(1-0.3)^2} = -47.62 < 0 \tag{72}$$

The negative value confirms that we have indeed found a maximum. In practice, Myung (2003) remarks that it is not always possible to find an analytical solution, especially when several parameters must be estimated and a non linear problem is present. This has to be calculated by using non linear optimization algorithms. These algorithms play a kind of *hotter - colder* game to get to the answer, instead of conducting an exhausting search through the entire domain. Depending on the choice of the initial parameter, the algorithm could give us the maximum likelihood estimate p_{MLE}, *global maximum*, or a *local maximum*.

On Figure 36, Point C is called global maximum while Point A, B, and D are local maxima. Selecting x2 as the initial guess of parameter *p* will lead to the global maximum, while choosing x1 leads to Point A. Selecting other starting point could lead to other suboptimal answer such as Point B, or in the case of choosing x3, to Point D.

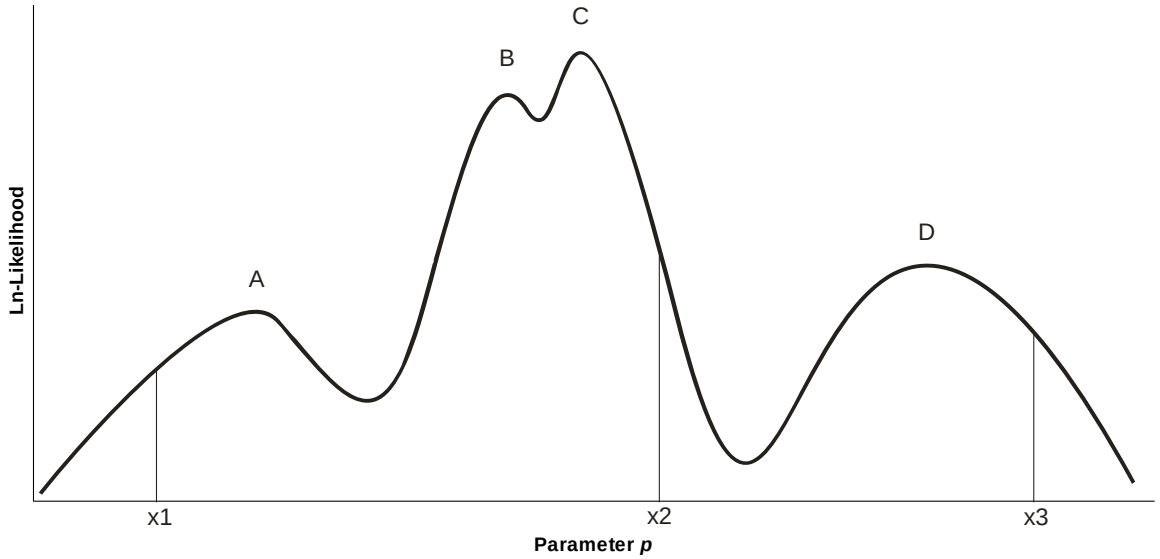

Figure 36: Schematic representation of a Ln-likelihood function of a 1 parameter model with several local maxima and a global maximum at C. Parameters x1, x2, and x3 are initial guesses

Unfortunately, there is no general solution to overcome the local maximum problem efficiently. Even though, a variety of techniques can be used to try to avoid this problem. For example, one could try to choose different starting values and if the same parameter value is obtained, one could have some confidence that the value represents a global maximum. A stochastic optimization algorithm known as simulated annealing (Kirkpatrick et al. 1983) can theoretically overcome the local maxima problem, however, the algorithm may not be a practicable option as it may take an unrealistically long time to find the solution (Myung 2003).

For the case of easy examples of a few parameters, it is possible to use an analytical solution for the estimation of parameters. The solution will be to derivate the natural logarithmic likelihood function with respect to the parameter, as it was exemplified. The likelihood function is in fact the probability function, yet, treating the parameters as variables. In practice, we use iterative maximum-likelihood-based numerical algorithms to estimate parameters.

For other examples, see Myung (2003). He also provides the example of how to estimate the value of two parameters (a, b > 0) for the exponential model $y = a\ exp(-b\ x)$ and for the power model $y = a\ x^{-b}$ using a MATLAB numerical algorithm which is also provided.

A.2 MAXIMUM LIKELIHOOD ESTIMATE FOR THE NORMAL DISTRIBUTION CASE

In Burnham and Anderson (2002), we find the derivation of the likelihood function of normally distributed residuals. Consider a multiple linear regression model where a variable y is function of r variables x, with β_r regression coefficients. The residuals ε_i of the n observations are assumed to be independent and normally distributed with a constant variance of σ^2:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_r x_r + \varepsilon_i \qquad i = 1, \ldots, n \qquad 73$$

The probability distribution model for the observation i is given by,

$$g(\varepsilon_i \mid \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2} \qquad 74$$

where θ represents the unknown parameters $(\beta_0, \beta_1, ..., \beta_r, \sigma)$. The number of parameters in linear regressions equals the unknown regression coefficients + 1, unknown standard deviation σ.

Since we want to work with all of the observations, the probability distribution model for all independent observations is simply the product of each probability distribution from i = 1, …, n.

$$g(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n \mid \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma}\right)^2} \qquad 75$$

If we interpret the above probability function as a function of: the parameters θ given the model and the data x and having the same properties of the residuals ε_i, as assumed before (normal distribution, independence, and constant variance σ^2), the likelihood function is

$$L(\theta \mid x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2} \qquad 76$$

Notice that the right term of the equation is the same as in the probability distribution, just the interpretation at the left side changed. Instead of laying interest on the probability of getting a certain residual, we are interested on which is the most probable parameter according to x data.

The variance normally has to be estimated, so it gets a "hat," and a logarithmic transformation is applied to obtain a computer friendly likelihood function:

$$\ln L(\hat{\theta} \mid x) = -n \ln(\sqrt{2\pi\hat{\sigma}^2_{ML}}) - \frac{1}{2\hat{\sigma}^2_{ML}} \sum_{i=1}^{n} \varepsilon_i{}^2 \qquad 77$$

where

$$\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \qquad 78$$

denotes the estimated variance in the maximum likelihood method. This variance is biased, but there is no theoretical justification to replace it with the unbiased least squared variance $\hat{\sigma}^2_{LS}$ (Ye et al. 2008)

$$\hat{\sigma}^2_{LS} = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n - K} \qquad 79$$

When n is much greater than K, the difference between the two variances is negligible.

By replacing $\hat{\sigma}^2_{ML}$ in the above likelihood or ln-likelihood function we obtain the maximized likelihood function of the normal distribution

$$L(\hat{\theta} \mid x) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2_{ML}}} \right)^n e^{-\frac{1}{2}n} \qquad 80$$

and the maximized Gaussian Ln-likelihood function

$$\ln(L(\hat{\theta} \mid x)) = -\frac{1}{2} n \ln(\hat{\sigma}^2_{ML}) - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \qquad 81$$

These resulting three terms can be easily calculated and are the ones which are normally substituted in model selection methods like in the general AIC equation (Equation 12) or into BIC or KIC, which were developed under a Bayesian context.

A.3 BAYESIAN INFERENCE

Bayesian inference is basically statistical inference using the Bayes Theorem. Its goal is to update the probability of an event A by taking into account additional information gained by observing event B. For example, consider the existence of God[13] (event A). Either God could exist or not, statistically the probability is ½. However, if we consider the creation and order of the universe with its stars, galaxies, and different celestial objects (event B), the probability might change. With the Bayes theorem, it is possible to calculate a new probability which takes this information into account. The Bayes theorem is (Soong 2004:25)

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)} \qquad 82$$

where

$P(A|B)$ is the posterior probability or conditional probability of A given B;

$P(A)$ is the prior or marginal probability of A, regardless of information about B;

$P(B|A)$ is the likelihood or conditional probability of B given A;

$P(B)$ is the prior or marginal probability of B defined as

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i) \qquad 83$$

where A_i represents a set of mutually exclusive and exhaustive events. In this case only two events apply, either God exists or not.

Let us put some numbers on our very simple example: the probability that God exists without additional knowledge is $P(A)$ = ½, while the probability that He does not is $P(A')$ = ½. We guess or assume that the creation of the universe (event B) is more probable if God exists, $P(B|A)$ = ⅔; than if He does not, $P(B|A')$ = ⅓.

[13] The calculation with the Bayes Theorem of the probability that God exists was the cover article of the P.M. magazine in December 2006.

$$P(God\ exists \mid creation\ of\ the\ universe) = \frac{0.5 \times 0.666}{0.5 \times 0.666 + 0.5 \times 0.333} = 0.666 \qquad 84$$

Bayes theorem can be expressed in different forms. For example, by dividing both the numerator and denominator in Equation 82 by $P(B|A')$, which is the conditional probability of B given an alternative to A (such as "not A"), we obtain

$$P(A \mid B) = \frac{P(A)\frac{P(B \mid A)}{P(B \mid A')}}{P(A)\frac{P(B \mid A)}{P(B \mid A')} + P(A')} \qquad 85$$

This formula was rewritten for simplicity by Vašek (2006) as follows:

$$P_{after} = \frac{P_{before} \cdot GI}{P_{before} \cdot GI + 1 - P_{before}} \qquad 86$$

where P_{after} replaces $P(A|B)$; P_{before} substitutes $P(A)$; 1 – P_{before} is used instead of $P(A')$; and GI, the God existence indicator, is the ratio $\frac{P(B \mid A)}{P(B \mid A')}$ for which Vašek (2006) suggests values from 10 to 0.1, see following table:

Table 32: Meaning of several *GI* values

GI value	meaning	$\frac{P(B \mid A)}{P(B \mid A')}$
10	most likely when God exists	$\frac{0.90\underline{90}}{0.09\underline{09}}$
2	likely when God exists	$\frac{0.666\underline{6}}{0.333\underline{3}}$
1	neutral	$\frac{0.5}{0.5}$
0.5	likely when God does not exist	$\frac{0.333\underline{3}}{0.666\underline{6}}$
0.1	most likely when God does not exist	$\frac{0.09\underline{09}}{0.90\underline{90}}$

The probability that God exists increased from ½ to ⅔ when considering the creation of the universe. We can continue to calculate the probability of God existence considering other

criteria such as war, hunger, rapes, murder, and criminality. These calamities are most likely to occur in a world without God; therefore, we assign a GI of 0.1

$$P_{after} = \frac{0.66\underline{6} \times 0.1}{0.66\underline{6} \times 0.1 + 1 - 0.66\underline{6}} = 0.166 \qquad 87$$

The probability for the existence of God drops to around 17 %. One could continue in this manner to recalculate the probability considering other events which may affect the result.

Other examples can be found in most books of Statistics, for example in Soong (2004:25) or Koch (2000:5).

B. MODEL SELECTION ISSUES

Relevant model selection issues referred to in the main text are annexed here, as well as the synthetic data generated by the model presented in Section 4.2.

B.1 TAKEUCHI'S INFORMATION CRITERION (TIC, TAKEUCHI 1978)

Although Akaike was the first to estimate the K-L information, he was not the only one. Takeuchi (1976) developed a model selection criterion based on an estimate of the K-L distance. He made a very general derivation without assuming that reality exists as a model or that such a model is in the set of candidate models. Takeuchi's information criterion (TIC) is given as (Burnham and Anderson 2002:65):

$$TIC = -2\ln[L(\theta \mid y)] + 2 \cdot tr[J(\theta)I(\theta)^{-1}] \qquad 88$$

where "*tr*" stands for the matrix trace function. $J(\theta)$ and $I(\theta)$ are matrices of K x K dimensions. Burnham and Anderson (2002:65) do not recommend the use of TIC since it is very hard to estimate proper values for the matrices. These values can hardly be estimated unless n is extremely big. Even if the matrix elements are properly estimated and we have good approximating models, the result is almost equal to AIC. When the approximating model is "good" (meaning low K-L information), the second term in TIC, $2 \cdot tr[J(\theta)I(\theta)^{-1}]$ is ALMOST EQUAL to $2K$; this makes AIC an excellent approximation of TIC (Burnham and Anderson 2002:65, 368, 385). According to Shibata (1989) the best estimator of $tr[J(\theta)I(\theta)^{-1}]$ is probably K. For these reasons, Burnham and Anderson (2002:65)

conclude that AIC is a reliable estimator of TIC. This proves that AIC is as unbiased as TIC, which would mean that the truth does not need to be present in the set of models. Ye et al. (2008) disagree on this last issue and point out that Burnham and Anderson (2002:368) also state that if the "true" model is considered in the set of candidate models then $tr[J(\theta)I(\theta)^{-1}]$ EQUALS K. This would mean that AIC assumes that the truth is included. More information on the mathematical derivation of the TIC can be found in Shibata (1989), Konishi and Kitagawa (1996), and Burnham and Anderson (2002:368).

B.2 MULTIMODEL INFERENCE

The development of the Akaike weights for model selection made possible the incursion into another area of data analysis: multimodel inference. Instead of only selecting a model with AIC, modelers use the best ranked models together to make inference. In practice, this is achieved by averaging model parameters. Burnham and Anderson (2002) experimented in ecology and wildlife models using the average procedure and highly recommend the method. In their experience, the predictions are better and the method considers the other good models in addition to just the best one. In the case that the best model is not the absolute winner with Akaike weights (w) bigger than 0.9, then model average could be carried out with help of these weights as follows:

$$\hat{\bar{y}} = \sum_{i=1}^{R} w_i \hat{y}_i \qquad 89$$

where $\hat{y}_i$ is the estimated result of each model i and $\hat{\bar{y}}$ is the averaged estimated result,

$$\hat{\bar{\beta}} = \sum_{i=1}^{R} w_i \hat{\beta}_i \qquad 90$$

where $\hat{\beta}_i$ is the estimated parameter value of each model i and $\hat{\bar{\beta}}$ stands for the averaged estimated parameter.

If the parameter is not present in all of the models, there are two possibilities to average it. Either we simply assume a value of zero when the parameter is absent, $\hat{\beta} = 0$ in Equation 90, or we just consider the models which include the parameter being averaged (Massmann

2004:16). Notice that for this last option, we have to recalculate the Akaike weights to also sum 1 for the subset of models considered (Poeter and Anderson 2005).

However, Poeter and Anderson (2005) do not recommend parameter averaging in groundwater modeling; in their own words: "*Parameter averaging is rarely useful for ground water modeling because use of an average parameter value in a particular model construct is not appropriate*."

B.3 HILL AND TIEDEMAN'S (2007A) METHOD TO OBTAIN THE VALUE OF AIC FOR DIFFERENT TYPES OF OBSERVATIONS

Hill (1998) implemented the AIC in hydrogeology to models calibrated with different kinds of observations. However, this procedure is not an easy task and severe implementation errors (discussed in Section 3.8) were made.

AIC is given by Hill and Tiedeman (2007a:99) as

$$AIC(b') = S'(b') + NP \times 2 \qquad 91$$

where $S'(b')$ is the maximum likelihood function and NP is the number of estimated parameters.

Hill and Tiedeman's (2007a:375) $S'(b')$ is based on the normal distribution likelihood function found on Brockwell and Davis (1987:247). By assuming normally distributed residuals, multiplying this likelihood function by -2, and taking the natural logarithm of it, Hill and Tiedeman obtained the following maximum likelihood objective function:

$$S'(b') = n\ln(2\pi) + n\ln\hat{\sigma}^2 - \ln|\omega| + e^T(\omega/\hat{\sigma}^2)e \qquad 92$$

where n is number of observations, $|\omega|$ is the determinant of a weight matrix with dimension [n x n], and e is a vector of residuals of the form [y-y'] where y is an observation vector and y' the simulated value vector. All vectors have a dimension equal to the number of observed values.

Hill and Tiedeman (2007a:29) define the weight matrix such that the common error variance of the residuals equals one. Substituting the unit variance in Equation 92 causes the leading

term of the AIC Gaussian form $n\ln(\hat{\sigma}^2)$ to drop and the maximum likelihood function is simplified to:

$$S'(b') = (ND + NPR)\ln(2\pi) - \ln|\omega| + e^T \omega e \quad 93$$

where (*ND*+*NPR*) is the number of observations plus number of prior information. Prior information is another type of data or "observation," for example a conductivity value. In this dissertation there is no distinction between them and both are treated as *n* = number of observed values.

This objective function, adopted from Hill (1998) does not appear in any model selection criteria developed by other authors. Model selection criteria based on the maximum likelihood function usually have the same goodness of fit term: $n\ln(\hat{\sigma}^2)$, and the criteria differ just on the penalty term. Hill and Tiedeman's book corrections (Hill and Tiedeman 2007b) state that AIC, AIC_c, and BIC are more commonly calculated with Equation 94 below and not with Equation 93 above.

$$S'(b') = n \times \ln\left[\left(e^T \omega e\right)/ n\right] \quad 94$$

where *n* stands for the number of observations plus number of prior information.

Substituting 94 in 91 results in the accepted AIC formula (Equation 16, p. 13), whereas substituting 93 in 91 results in the AIC formula used by Hill and Tiedeman (2007a:29, 99)

$$AIC_{Hill\ and\ Tiedeman} = n\ln(2\pi) - \ln|\omega| + e^T \omega e + 2K \quad 95$$

where $e^T \omega e$ is the weighted Residual Sum of Squares, *RSS*.

AIC does not provide an absolute value. The value is a relative value which only provides information when compared with AIC values of other candidate models. Therefore, we are not interested in the value itself and are able to drop constant terms. By dropping the terms which are invariable for all candidate models, we find the underlying equation which actually ranks the models of Hill and Tiedeman:

$$AIC_{Hill\ and\ Tiedeman} = e^T \omega e + 2K \quad 96$$

B.4 SYNTHETIC DATA

The following data were obtained from the "data generating model" documented in Section 4.2. It consists of three data sets, each of 100 observations. Hydraulic pressure (h) and conductivity (K) observations where observed once, since the flow model is steady state. The concentration observations were observed in selected boreholes every 28 days.

Table 33: Synthetic data: observed hydraulic pressure (h), hydraulic conductivity (K), and concentration

Borehole nr.	Coordinates X	Coordinates Y	h (m)	K (m/s)	Borehole nr.	Time (s)	Concentration (g/m^3)
1	392.29	855.35	37.59	6.05E-04	1	2419200	3.04E-06
2	223.81	744.65	37.53	1.80E-04	1	4838400	2.51E-05
3	218.78	588.68	38.05	2.14E-04	1	7257600	5.72E-05
4	148.37	523.27	37.16	9.44E-05	1	9676800	8.38E-05
5	1002.36	224.89	36.89	1.49E-03	1	12100000	1.01E-04
6	246.44	203.77	36.88	4.32E-03	1	14520000	1.10E-04
7	402.35	334.59	37.44	2.99E-05	1	16930000	1.13E-04
8	500.42	586.16	37.88	1.63E-04	1	19350000	1.13E-04
9	628.67	548.43	37.36	1.57E-03	1	21770000	1.10E-04
10	691.53	802.52	37.54	3.40E-04	1	24190000	1.05E-04
11	606.04	228.93	36.83	2.03E-04	1	26610000	1.00E-04
12	772.00	314.47	36.91	2.55E-04	1	29030000	9.42E-05
13	955.57	314.47	37.11	2.50E-04	1	31449200	8.84E-05
14	1020.96	561.01	37.07	2.33E-03	6	26610000	0
15	1174.35	845.28	37.44	2.72E-04	6	29030000	1.74E-07
16	1390.61	822.64	38.12	5.16E-04	6	31449200	4.73E-07
17	1254.82	425.16	37.02	4.70E-05	8	31449200	0
18	1352.89	193.71	36.82	1.95E-04	22	21770000	0
19	1121.54	98.11	36.75	8.39E-04	22	24190000	2.94E-07
20	123.00	847.24	37.66	2.43E-04	22	26610000	7.94E-07
21	66.45	377.65	37.00	3.80E-03	22	29030000	1.75E-06
22	234.68	437.06	37.07	6.96E-03	22	31449200	3.31E-06
23	585.30	694.48	37.42	1.08E-02	23	31449200	0
24	421.30	138.61	36.79	1.61E-03	40	16930000	4.45E-07
25	600.85	379.07	37.07	5.28E-04	40	19350000	1.37E-06
26	852.50	932.11	37.59	4.60E-05	40	21770000	3.19E-06
27	916.12	792.08	37.40	8.21E-04	40	24190000	6.08E-06
28	875.12	609.62	37.24	1.26E-03	40	26610000	9.97E-06
29	1036.29	397.45	36.95	7.65E-03	40	29030000	1.46E-05
30	1396.80	554.46	37.31	4.96E-05	40	31449200	1.96E-05
31	1238.45	243.28	36.90	1.67E-03	41	14520000	4.94E-07
32	333.65	241.87	37.18	2.30E-05	41	16930000	1.92E-06
33	1229.97	848.66	38.39	3.62E-04	41	19350000	5.33E-06
34	1187.56	613.86	37.72	6.60E-04	41	21770000	1.16E-05
35	1401.04	438.47	37.57	6.37E-03	41	24190000	2.10E-05
36	1104.15	377.65	37.40	2.32E-03	41	26610000	3.32E-05
37	928.84	425.74	37.32	3.03E-04	41	29030000	4.75E-05
38	926.01	212.16	36.99	2.45E-04	41	31449200	6.26E-05
39	1080.11	275.81	37.02	3.35E-04	42	9676800	8.45E-07
40	370.39	590.41	37.32	4.61E-03	42	12100000	4.36E-06
41	342.75	623.62	37.32	5.79E-03	42	14520000	1.36E-05
42	377.76	688.19	37.36	2.01E-03	42	16930000	3.05E-05
43	460.69	656.83	37.39	2.64E-02	42	19350000	5.56E-05

44	416.46	619.93	37.36	2.39E-02	42	21770000	8.68E-05
45	462.53	592.25	37.36	6.80E-03	42	24190000	1.22E-04
46	383.29	433.58	37.18	1.26E-03	42	26610000	1.57E-04
47	764.74	621.77	37.30	2.17E-03	42	29030000	1.90E-04
48	571.25	898.52	37.58	2.50E-02	42	31449200	2.18E-04
49	66.34	706.64	37.54	3.09E-04	43	21770000	2.62E-07
50	73.71	73.80	36.81	1.19E-03	43	24190000	5.28E-07
51	528.87	435.42	37.21	7.14E-04	43	26610000	8.83E-07
52	221.13	278.60	36.98	1.04E-02	43	29030000	1.30E-06
53	302.21	178.97	36.83	4.90E-03	43	31449200	1.74E-06
54	427.52	282.29	36.97	2.89E-03	44	16930000	1.82E-07
55	451.47	311.81	37.01	3.06E-03	44	19350000	5.83E-07
56	545.45	46.13	36.68	3.84E-03	44	21770000	1.33E-06
57	232.19	933.58	37.68	7.83E-04	44	24190000	2.46E-06
58	473.59	881.92	37.58	6.20E-03	44	26610000	3.88E-06
59	963.76	70.11	36.65	1.05E-03	44	29030000	5.49E-06
60	136.36	190.04	36.90	7.01E-03	44	31449200	7.13E-06
61	410.93	787.82	37.51	2.43E-04	45	31449200	1.34E-07
62	316.95	793.36	37.59	4.55E-04	46	29030000	1.81E-07
63	239.56	461.25	37.09	3.35E-03	46	31449200	3.35E-07
64	574.94	147.60	36.74	2.03E-03	52	26610000	1.66E-07
65	674.45	459.41	37.18	1.41E-04	52	29030000	4.78E-07
66	164.00	619.93	37.37	3.74E-04	52	31449200	1.12E-06
67	326.17	306.27	36.99	2.91E-03	58	31449200	0
68	175.06	306.27	36.99	4.97E-03	61	2419200	1.28E-04
69	228.50	92.25	36.78	4.88E-02	61	4838400	7.06E-04
70	493.86	522.14	37.32	3.47E-03	61	7257600	1.29E-03
71	816.34	848.71	38.99	3.86E-05	61	9676800	1.64E-03
72	781.33	627.31	37.66	1.39E-04	61	12100000	1.78E-03
73	961.92	695.57	37.86	2.27E-04	61	14520000	1.78E-03
74	1087.22	610.70	37.70	6.94E-04	61	16930000	1.71E-03
75	1085.38	466.79	37.48	2.90E-03	61	19350000	1.60E-03
76	1247.54	396.68	37.49	4.03E-03	61	21770000	1.48E-03
77	1310.20	278.60	37.42	1.32E-03	61	24190000	1.35E-03
78	1315.73	178.97	37.28	2.07E-04	61	26610000	1.23E-03
79	1435.50	88.56	37.09	8.87E-05	61	29030000	1.11E-03
80	1083.54	62.73	36.85	7.54E-05	61	31449200	1.00E-03
81	1190.42	60.89	36.84	4.30E-05	62	31449200	0
82	1199.63	178.97	37.16	2.81E-04	63	24190000	2.27E-07
83	1253.07	317.34	37.43	8.28E-04	63	26610000	6.15E-07
84	1225.43	507.38	37.54	2.90E-03	63	29030000	1.37E-06
85	1393.12	625.46	37.72	4.81E-03	63	31449200	2.60E-06
86	1304.67	741.70	37.92	1.50E-03	67	31449200	0
87	1136.98	793.36	38.16	5.16E-04	68	29030000	1.58E-07
88	1177.52	905.90	38.73	4.39E-04	68	31449200	4.17E-07
89	1374.69	922.51	38.71	2.45E-04	70	31449200	0
90	1442.88	955.72	39.03	5.75E-05	92	31449200	1.67E-07
91	528.87	387.45	37.55	1.29E-04	94	9676800	1.53E-07
92	348.28	555.35	37.90	6.71E-04	94	12100000	2.74E-07
93	263.51	732.47	38.80	3.41E-04	94	14520000	4.22E-07
94	364.86	845.02	39.09	2.15E-04	94	16930000	5.90E-07
95	125.31	915.13	39.27	6.28E-04	94	19350000	7.70E-07
96	95.82	562.73	38.08	9.61E-05	94	21770000	9.55E-07
97	119.78	287.82	37.42	4.24E-04	94	24190000	1.14E-06
98	281.94	97.79	36.95	4.77E-05	94	26610000	1.31E-06
99	504.91	123.62	36.79	9.08E-06	94	29030000	1.47E-06
100	674.45	119.93	36.81	7.55E-05	94	31449200	1.61E-06

B.5 NORMALIZED SYNTHETIC DATA

The following data consists of two data sets: hydraulic pressure (h) and conductivity (K). We consider the first 19 observations of the full synthetic data set of 100 observations provided in the previous section. The observed values are normalized by their observed standard deviation. In order to calculate the standard deviation, the values are first centered at zero.

Table 34: Normalized values of observed hydraulic pressure (h) and hydraulic conductivity (K)

Borehole nr.	Observed h (m)	Centered at zero	Normalized by St.dev. (-)	Observed K (m/s)	Log transformed	Centered at zero	Normalized by St.dev. (-)
1	37.59	0.3	91.0772	6.05E-04	-3.2	0.3	-5.9546
2	37.53	0.3	90.9326	1.80E-04	-3.7	-0.3	-6.9285
3	38.05	0.8	92.1780	2.14E-04	-3.7	-0.2	-6.7890
4	37.16	-0.1	90.0374	9.44E-05	-4.0	-0.6	-7.4469
5	36.89	-0.4	89.3687	1.49E-03	-2.8	0.6	-5.2297
6	36.88	-0.4	89.3363	4.32E-03	-2.4	1.1	-4.3756
7	37.44	0.2	90.6930	2.99E-05	-4.5	-1.1	-8.3722
8	37.88	0.6	91.7661	1.63E-04	-3.8	-0.3	-7.0068
9	37.36	0.1	90.5094	1.57E-03	-2.8	0.7	-5.1888
10	37.54	0.3	90.9514	3.40E-04	-3.5	0.0	-6.4182
11	36.83	-0.5	89.2340	2.03E-04	-3.7	-0.2	-6.8317
12	36.91	-0.4	89.4194	2.55E-04	-3.6	-0.1	-6.6475
13	37.11	-0.2	89.9169	2.50E-04	-3.6	-0.1	-6.6633
14	37.07	-0.2	89.8154	2.33E-03	-2.6	0.8	-4.8721
15	37.44	0.2	90.6981	2.72E-04	-3.6	-0.1	-6.5957
16	38.12	0.8	92.3451	5.16E-04	-3.3	0.2	-6.0821
17	37.02	-0.3	89.6778	4.70E-05	-4.3	-0.9	-8.0072
18	36.82	-0.5	89.1918	1.95E-04	-3.7	-0.2	-6.8623
19	36.75	-0.5	89.0371	8.39E-04	-3.1	0.4	-5.6918
Mean	37.28	0.0			-3.5	0.0	
St.dev.		0.41				0.54	

REFERENCES

Akaike H (1969) **Fitting Autoregressive Models for Prediction** *Annals of The Institute of Statistical Mathematics* 21(2):243-247.

Akaike H (1973) **Information theory and an extension of the maximum likelihood principle** In: Kotz S, Johnson NL, editors. (1992) Breakthroughs in Statistics, vol. 1: Foundations and basic theory. New York, USA: Springer-Verlag. pp. 610-624 ISBN: 0387975-667.

Allen DM (1974) **The Relationship between Variable Selection and Data Augmentation and a Method for Prediction** *Technometrics* 16(1):125-127.

Allen RG, Pereira LS, Raes D, Smith M (1998) **Crop evapotranspiration - Guidelines for computing crop water requirements** FAO Irrigation and drainage paper 56. Rome, Italy: FAO - Food and Agriculture Organization of the United Nations. ISBN 9251042195 Available at http://www.fao.org/docrep/x0490e/x0490e00.htm#Contents.

Barron AR (1984) **Predicted squared error: a criterion for automatic model selection** In: Farlow SJ, editor. Self-Organizing Methods in Modeling: GMDH Type Algorithms, vol. 54. New York, USA: Marcel Dekker. pp. 87-103.

Bengtsson T, Cavanaugh JE (2006) **An improved Akaike information criterion for state-space model selection** *Computational Statistics and Data Analysis* 50(10):2635-2654 DOI: 10.1016/j.csda.2005.05.003.

Bärschneider H (2008) **Modellierung der Bodenwasserbewegung in Lysimetern mit Hydrus 1D** Hydrologisch-Meteorologische Projektarbeit, Institut für Hydrologie und Meteorologie, Technische Universität Dresden.

Bhansali RJ, Downham DY (1977) **Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion** *Biometrika* 64(3):547-551 DOI: 10.1093/biomet/64.3.547.

Boltzmann LE (1877) **Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respective den Sätzen über das Wärmegleichgewicht** *Wiener Berichte* 76:373-435.

Bozdogan H (1987) **Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions** *Psychometrika* 52:345-370.

Bozdogan H (2000) **Akaike's information criterion and recent developments in information complexity** *Journal of Mathematical Psychology* 44(1):62-91 DOI: 10.1006/jmps.1999.1277.

Bozdogan H, Haughton DMA (1998) **Informational complexity criteria for regression models** *Computational Statistics and Data Analysis* 28(1):51-76.

Brockwell JA, Davis RA (1987) **Time Series, Theory and Methods** New York, USA: Springer-Verlag. ISBN: 0387964061.

Browne MW (2000) **Cross-validation methods** *Journal of Mathematical Psychology* 44(1): 108-132 DOI: 10.1006/jmps.1999.1279.

Brutsaert W (2005) **Hydrology: An Introduction** Cambridge University Press. ISBN: 05218-24796, 9780521824798.

Burnham KP, Anderson DR (1999) **Understanding information criteria for selection among capture-recapture or ring recovery models** *Bird Study* 46(Supp/1):14-21.

Burnham KP, Anderson DR (2001) **Kullback-Leibler information as a basis for strong inference in ecological studies** *Wildlife Research* 28(2):111-119 DOI: 10.1071/WR 99107.

Burnham KP, Anderson DR (2002) **Model selection and multimodel inference: A practical information-theoretic approach** 2nd ed. New York, USA: Springer-Verlag. ISBN: 978-0-387-95364-9.

Burnham KP, Anderson DR (2004) **Multimodel inference: Understanding AIC and BIC in model selection** *Sociological Methods and Research* 33(2):261-304 DOI: 10.1177/0049 124104268644.

Carrera J, Neuman SP (1986) **Estimation of aquifer parameters under transient and steady-state conditions** *Water Resources Research* 22(2):199-242.

Cavanaugh JE (2004) **Criteria for Linear Model Selection Based on Kullback's Symmetric Divergence** *Australian & New Zealand Journal of Statistics* 46(2):257-274 DOI: 10.1111/ j.1467-842X.2004.00328.x.

Cavanaugh JE, Neath AA (1999) **Generalizing the derivation of the Schwarz information Criterion** *Commun. Stat. Theory Methods* 28:49-66.

Chiang W-H (2005) **3D-Groundwater Modeling with PMWIN: A Simulation System for Modeling Groundwater Flow and Transport Processes** 2nd ed. Berlin Heidelberg, Germany: Springer-Verlag. ISBN: 103-540-27590-8.

Chiang W-H, Kinzelbach W (1991-2001) **Processing Modflow (PMWIN), Pre- and postprocessors for the simulation of flow and contaminant transport in groundwater system with MODFLOW, MODPATH and MT3D** Version 5.3.0. Available at http://www.pmwin.net.

Chiang W-H, Kinzelbach W (1994-2001) **PMPATH for Windows** Version 6.1.0. Available at http://www.pmwin.net.

Craven P, Wahba G (1979) **Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation** *Numerische Mathematik* 31(4):377-403 DOI: 10.1007/BF01404567.

de Luna X (1998) **An Improvement of Akaike's FPE Criterion to Reduce its Variability** *Journal of Time Series Analysis* 19(4):457-471 DOI: 10.1111/1467-9892.00103.

deLeeuw J (1992) **Introduction to Akaike (1973) Information theory and an extension of the maximum likelihood principle** In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics, vol. 1: Foundations and basic theory. New York, USA: Springer-Verlag. pp. 599-609 ISBN: 0387975667.

Doherty J (2006) **PEST Model-Independent Parameter Estimation, V10.1**, Bethesda, USA: Papadopulos SS, Inc. Available at http://www.sspa.com/pest.

Efron B (1979) **Bootstrap Methods: Another Look at the Jackknife** *The Annals of Statistics* 7(1):1-26 DOI: 10.1214/aos/1176344552.

Feddes RA, Kowalik PJ, Zaradny H (1978) **Simulation of Field Water Use and Crop Yield** Wageningen, the Netherlands: PUDOC. ISBN: 0470264632

Foglia L, Mehl SW, Hill MC, Perona P, Burlando P (2007) **Testing Alternative Ground Water Models Using Cross-Validation and Other Methods** *Ground Water* 45(5):627-641 DOI: 10.1111/j.1745-6584.2007.00341.x.

Frenzel H (1995) **A field generator based on Mejia's algorithm** Institut für Umweltphysik, University of Heidelberg, Germany.

Geisser S (1975) **The Predictive Sample Reuse Method with Applications** *Journal of the American Statistical Association* 70(350):320-328.

Geweke J, Meese R (1981) **Estimating Regression Models of Finite but Unknown Order** *International Economic Review* 22(1):55-70.

Gómez-Hernández JJ (2006) **Complexity** *Ground Water* 44(6):782-785 DOI: 10.1111/j.1745-6584.2006.00222.x.

Grünwald P (2000) **Model selection based on minimum description length** *Journal of Mathematical Psychology* 44:133-152.

Han D, Yan W, Moghaddamnia A (2009) **Model Input Data Selection by the Gamma Test** EGU General Assembly *Geophysical Research Abstracts* 11:EGU2009-9711-2.

Hannan EJ, Quinn BG (1979) **The Determination of the Order of an Autoregression** *Journal of the Royal Statistical Society. Series B (Methodological)* 41(2):190-195.

Harbaugh AW (2002) **A data input program (MFI2K) for the U.S. Geological Survey modular ground-water model (MODFLOW-2000)** *U.S. Geological Survey Open-File Report*:02-41.

Harbaugh AW, Banta ER, Hill MC, McDonald MG (2000) **MODFLOW-2000, The U.S. Geological Survey modular ground-water model - User guide to modularization concepts and the ground-water flow process** *U.S. Geological Survey Open-File Report*:00-92.

Hélie S (2006) **An introduction to model selection: Tools and algorithms** *Tutorials in Quantitative Methods for Psychology* 2(1):1-10.

Hernández AF, Neuman SP, Guadagnini A, Carrera J (2006) **Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media** *Water Resources Research* 42:W05425 DOI: 10.1029/2005WR004449.

Hill MC (1998) **Methods and guidelines for effective model calibration** *U.S. Geological Survey Water-Resources Investigations Report*:98-4005.

Hill MC, Banta ER, Harbaugh AW, Anderman ER (2000) **MODFLOW-2000, The U.S. Geological Survey modular ground-water model - User guide to the observation, sensitivity, and parameter-estimation processes** *U.S. Geological Survey Open-File Report*: 00-184.

Hill MC, Tiedeman CR (2007a) **Effective groundwater model calibration - with analysis of data, sensitivities, predictions, and uncertainty** Hoboken, USA: John Wiley & Sons. ISBN: 978-0-471-77636-9.

Hill MC, Tiedeman CR (2007b) **Corrections and additional references (Nov/5/2007)** http://wwwbrr.cr.usgs.gov/projects/GW_ModUncert/hill_tiedeman_book/documentation/ Accessed February 2008.

Hill MC, Tiedeman CR (2007c) **Effective Groundwater Model Calibration, with Analysis of Data, Sensitivities, Predictions, and Uncertainty - Answers to Problems Posed in the Exercises (24 January 2007)** http://wwwbrr.cr.usgs.gov/projects/GW_ModUncert/hill_tiedeman_book/ Accessed February 2008.

Hölting B (1992) **Hydrogeologie: Einführung in die Allgemeine und Angewandte Hydrogeologie** 4th ed. Stuttgart, Germany: Enke. ISBN: 3-432-90794-X.

Hurvich CM, Tsai C-L (1989) **Regression and time series model selection in small samples** *Biometrika* 76:297-307.

Kashyap RL (1982) **Optimal choice of AR and MA parts in autoregressive moving average models** *IEEE Trans. Pattern Anal. Machine Intell.* 4(2):99-104.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) **Optimization by simulated annealing** *Science* 220:671-680.

Koch K-R (2000) **Einführung in die Bayes-Statistik** Springer-Verlag. ISBN: 3540666702, 9783540666707.

Kolditz O (1997) **Strömung, Stoff- und Wärmetransport im Kluftgestein** Berlin - Stuttgart, Germany: Gebrüder Borntraeger. ISBN: 3443010350.

Konishi S, Kitagawa G (1996) **Generalized information criteria in model selection** *Biometrika* 83:875-890 DOI: 10.1093/biomet/83.4.875.

Kresic N (1997) **Quantitative Solutions in Hydrogeology and Groundwater Modeling** Boca Raton, USA: CRC Press LLC. ISBN: 0566702194.

Kullback S (1959) **Information theory and statistics** New York, USA: John Wiley and Sons.

Kullback S, Leibler RA (1951) **On information and sufficiency** *Annals of Mathematical Statistics* 22(1):79-86.

Larsen J, Hansen LK (1994) **Generalization performance of regularized neural network models** In: Vlontzos J, Hwang J-N, Wilson E, editors. Piscataway, USA: Proceedings of the 1994 IEEE Workshop on Neural Networks for Signal Processing IV. pp. 42-51 ISBN: 0-7803-2026-3.

Lebreton J-D, Burnham KP, Clobert J, Anderson DR (1992) **Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies** *Ecological Monographs* 62(1):67-118 DOI: 10.2307/2937171.

Liedl R (2007) **Grundwassermodellierung – ein Riese auf tönernen Füßen?** Berufungsverfahren „Hydrogeologie“ Tübingen, 26. November 2007. Institut für Grundwasserwirtschaft, Technische Universität Dresden, Germany.

Linhart H, Zucchini W (1986) **Model Selection** New York, USA: John Wiley & Sons. ISBN: 0471837229.

Mailhol JC, Olufayo AA, Ruelle P (1997) **Sorghum and sunflower evapotranspiration and yield from simulated leaf area index** *Agricultural Water Management* 35(1-2):167-182 DOI: 10.1016/S0378-3774(97)00029-2.

Mallows CL (1973) **Some Comments on Cp** *Technometrics* 15(4):661-675.

Massmann C (2004) **Simplicity vs. complexity in tracer test modelling** MSc Thesis, Center for Applied Geoscience, University of Tübingen, Germany.

Mc Elwee CD (1982) **Sensitivity analysis and the ground-water inverse problem** *Ground Water* 20(6):723-735.

Mc Elwee CD (1987) **Sensitivity analysis of groundwater models** In: Bear J, Corapcioglu MY, editors. Advances in Transport Phenomena in Porous Media. Dordrecht, The Netherlands: NATO Advanced Study Institute Series. pp. 751-817 ISBN: 9024735335.

Meyer PD, Ye M, Rockhold ML, Neuman SP, Cantrell KJ (2007) **Combined estimation of hydrogeologic conceptual model, parameter, and scenario uncertainty with application to uranium transport at the Hanford Site 300 area** *NUREG/CR-6940 (PNNL-16396)* .

Moody JE (1991) **Note on generalization, regularization and architecture selection in nonlinear learning systems** In: Juang BH, Kung SY, Kamm CA, editors. Neural Networks for Signal Processing: Proceedings of the 1991 IEEE Workshop. Los Alamitos, USA: IEEE. pp. 1-10.

Moody JE (1992) **The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems** In: Moody JE, Hanson SJ, Lippmann RP, editors. Advances in Neural Information Processing Systems 4. San Mateo, USA: Morgan Kaufmann. pp. 847-854.

Mualem Y (1976) **A new model for predicting the hydraulic conductivity of unsaturated porous media** *Water Resources Research* 12(3):513–522 DOI:10.1029/WR012i003p00513.

Myung IJ (2003) **Tutorial on maximum likelihood estimation** *Journal of Mathematical Psychology* 47(1):90-100 DOI: 10.1016/S0022-2496(02)00028-7.

Neuman SP (2003) **Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models** *Stochast. Environ. Res. Risk Assess.* 17(5):291-305 DOI: 10.1007/s0047-003-0151-7.

NIST/SEMATECH (2003) **e-Handbook of statistical methods** Available at http://www.itl.nist.gov/div898/handbook/ Accessed June 2007.

Parzen E (1974) **Some recent advances in time series modeling** *IEEE Transactions on Automatic Control* 19(6):723-730.

Parzen E (1977) **Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes** In: Krishnaiah PR, editor. Multivariate Analysis–IV. Amsterdam, the Netherlands. pp. 283-295.

Poeter EP, Anderson DR (2005) **Multimodel ranking and inference in ground water modeling** *Ground Water* 43(4):597-605 DOI: 10.1111/j.1745-6584.2005.0061.x.

Poeter EP, Hill MC (2007) **MMA, a computer code for multimodel analysis** *U.S. Geol. Surv. Tech. Methods* :TM6-E3.

Poeter EP, Hill MC, Banta ER, Mehl S, Christensen S (2005) **UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation** *U.S. Geol. Surv. Tech. Methods* :6-A11.

Reddy JR (2005) **A Text Book of Hydrology** Firewall Media. ISBN: 8170080991.

Regonda SK, Rajagopalan B, Lall U, Clark M, Moon YI (2005) **Local polynomial method for ensemble forecast of time series** *Nonlinear Processes in Geophysics* 12(3):397-406 DOI: 1607-7946/npg/2005-12-397.

Richards LA (1931) **Capillary conduction of liquids through porous mediums** PhD Thesis, Cornell University.

Richards SA (2008) **Dealing with overdispersed count data in applied ecology** *Journal of Applied Ecology* 45(1):218-227 DOI: 10.1111/j.1365-2664.2007.01377.x.

Rissanen J (1978) **Modeling by shortest data description** *Automatica* 14(5):465-471 DOI: 10.1016/0005-1098(78)90005-5.

Sawa T (1978) **Information criteria for discriminating among alternative regression models** *Econometrica* 46(6):1273-1291 DOI: 10.2307/1913828.

Schwarz G (1978) **Estimating the dimension of a model** *Annals of Statistics* 6(2):461-464.

Seghouane A-K, Amari S-I (2007) **The AIC criterion and symmetrizing the Kullback-Leibler divergence** *IEEE Transactions on Neural Networks* 18(1):97-106 DOI: 10.1109/TNN.2006.882813.

Seghouane A-K, Bekara M (2004) **A small sample model selection criterion based on the Kullback symmetric divergence** *IEEE Trans. Signal Process* 52(12):3314-3323.

Sewell M (2008) **Model Selection** http://www.modelselection.org. Accessed April 2009.

Shibata R (1981) **An Optimal Selection of Regression Variables** *Biometrika* 68(1):45-54 DOI: 10.1093/biomet/68.1.45.

Shibata R (1989) **Statistical aspects of model selection** In: Willems JC, editor. From data to model. London, UK: Springer-Verlag. pp. 215-240.

Simunek J, Sejna M, Saito H, Sakai M, van Genuchten MT (2008) **The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Version 4.0** *Hydrus Series* 3,

Department of Environmental Sciences, University of California Riverside, Riverside, USA.

Simunek J, Sejna M, van Genuchten MT (2005-2008) **Code for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably Saturated Porous Media** Available at http://www.hydrus2d.com

Soong TT (2004) **Fundamentals of probability and statistics for engineer**s West Sussex, England: John Wiley & Sons, Ltd. ISBN: 0470868139 (HB), 0470868147 (PB).

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) **Bayesian measures of model complexity and fit** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4):583-639 DOI: 10.1111/1467-9868.00353.

Stone M (1974) **Cross-Validatory Choice and Assessment of Statistical Predictions** *Journal of the Royal Statistical Society. Series B (Methodological)* 36(2):111-147.

Sugiura N (1978) **Further analysis of the data by Akaike's Information Criterion and the finite corrections** *Communications in Statistics – Theory and Methods.7*(1):13-26 DOI: 10.1080/03610927808827599.

Takeuchi K (1976) **Distribution of informational statistics and a criterion of model fitting** *Suri-Kagaku (Mathematical Sciences)* 153:12-18 [In Japanese].

van Genuchten MT (1980) **A closed-form equation for predicting the hydraulic conductivity of unsaturated soils** *Soil Science Society of America Journal* 44(5):892–898. Available at http://hydro.nevada.edu/courses/gey719/vg.pdf.

Vapnik VN (1979) **Estimation of Dependences Based on Empirical Data** Moscow, USSR: Nauka [In Russian] (English translation Springer Verlag, 1982).

Vapnik VN, Chervonenkis AY (1968) **On the uniform convergence of relative frequencies of events to their probabilities** *Doklady Akademii Nauk USSR* 181(4) [In Russian].

Vapnik VN, Chervonenkis AY (1971) **On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities** *Theory of Probability and Its Applications* 16(2): 264-280 DOI: 10.1137/1116025.

Vapnik VN, Chervonenkis AY (1974) **Theory of Pattern Recognition: Statistical Problems of Learning** Moscow, USSR: Nauka. 415 pp. [In Russian] (German translation as Wapnik and Tscherwonenkis, 1979. Theorie der Zeichenerkennung. Berlin: Akademie Verlag).

Vašek T (2006) **Die Gottesformel** *P.M. Magazin* (Dezember):12-22. Available at http://www.pm-magazin.de/de/specials/artikel.asp?id=145.

Wallace CS, Boulton DM (1968) **An Information Measure for Classification** *Computer Journal* 11(2):185-194.

Walther W, Pätsch M, Konrad C (2008) **Dynamik des unterirdischen Wassers** Lecture script. Institut für Grundwasserwirtschaft, Technische Universität Dresden, Germany.

Wei CZ (1992) **On Predictive Least Squares Principles** *Annals of Statistics* 20(1):1-42 DOI: 10.1214/aos/1176348511.

Wel J (1975) **Least squares fitting of an elephant** *Chemtech* (February):128-129.

Wesseling JG, Elbers JA, Kabat P, van den Broek BJ (1991) **SWATRE: instructions for input**, Internal Note, Winand Staring Centre, Wageningen, the Netherlands.

Wherry RJ (1931) **A New Formula for Predicting the Shrinkage of the Coefficient of Multiple Correlation** *The Annals of Mathematical Statistics* 2(4):440-457 DOI: 10.1214/aoms/1177732951.

Wilks DS (1995) **Statistical Methods in the Atmospheric Sciences: An Introduction** International Geophysics Series, vol. 59. 1st ed. San Diego, USA: Academic Press. ISBN: 0127519653

Ye M, Meyer PD, Neuman SP (2008) **On model selection criteria in multimodel analysis** *Water Resources Research* 44(3):W03428 DOI: 10.1029/2008WR006803.

Ye M, Neuman SP, Meyer PD (2004) **Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff** *Water Resources Research* 40:W05113 DOI: 10.1029/2003WR002557.

Ye M, Neuman SP, Meyer PD, Pohlmann KF (2005) **Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff** *Water Resources Research* 41:W12429 DOI: 10.1029/2005WR004260.

Zheng C, Bennett GD (2002) **Applied Contaminant Transport Modelin**g 2nd ed. New York, USA: John Wiley & Sons. ISBN: 0471384771.