CHAPTER 3

Classical calibration procedures

I argued in the introduction that calibration and uncertainty estimation are closely connected. Conceptual model construction and parameterisation are usually founded on a very weak data base. Values derived from literature and from personal experience may narrow the range of possible conceptual models and values of parameters and variables. Nevertheless, the variety of solutions may still be too large to provide acceptable predictions.

Calibration is the process of rating different alternative sets of conceptual models, parameters and variables according to the degree of fit between simulation and observations. Classical calibration procedures aim to find a single unique parameter set that corresponds to the global optimum with regard to the degree of fit.

The parameter estimating problem has traditionally been solved by trial-and-error techniques, with the hydrologist successively changing the unknowns until he/she believes that the solution is close to the optimal. For large complex models with many unknowns this trial-and-error calibration procedure is very difficult and time consuming. With the constant increase in computer power, a range of automatical calibration techniques becomes more and more interesting.

This chapter gives an introduction to objective functions within standard calibration procedures as well as to the statistics of calibrated parameters and predicted state variables. The issue of single versus multi-objective function is briefly discussed.

First, however, I look at the issue of parameterisation and the different types of observation data that can be used. The concept of ill-posedness is defined and

model non-linearity is discussed.

3.1 Parameterisation

In distributed physically-based groundwater models the domain under consideration is discretised into a finite number of elements/cells, each of which has an individual set of parameters - i.e. sub-surface cells have hydraulic conductivities in three directions, a storage coefficient, effective porosity, sink/source term, etc.

Parameterisation concerns the assignment of values to these parameters. In most model applications the scale of parameterisation is larger than the scale of the numerical cells, i.e. parameters are assumed to be constant within regions/zones that are larger than the cell scale. In relation to calibration and uncertainty estimation the issue of parameterisation is very important.

3.1.1 The continuum approach

The continuum approach is fundamental in groundwater modelling and introduces the first set of restrictions at the level of parameterisation. The flow process on the molecular/microscopic scale, Fig. 3.1, is very complicated and in practice inaccessible. First, it is impossible to determine the exact geometry of the soil under consideration and, secondly, it is impossible to solve the flow equations for the volumes normally considered in groundwater modelling problems. Using a cylindrical pipe for stationary flow, Henry Darcy found by experiment a linear relation between the head gradient and the flow.

$$q = \frac{Q}{A} = -\frac{\Delta\psi}{\Delta L}K \tag{3.1}$$

where q is the Darcy or filter velocity, $\Delta\psi$ is the head difference over the soil sample, ΔL is the length of the soil sample, A is the cross section area of the soil sample (similar to the pipe cross section area) and K is the proportional constant or hydraulic conductivity.

Darcy's law is based on a macroscopic scale and builds on the continuum approach, where the porous media are considered as a continuum for which representative average parameters can be found. By introducing Darcy's law we are prevented from describing flow on anything smaller than the macroscopic scale. For many problems the macroscopic scale is sufficient, but when it comes to considering transport processes the flow on the microscopic scale cannot be ignored because of the high degree of variation in the size of pore velocities. In order to compensate for the unknown variation of flow velocities and directions

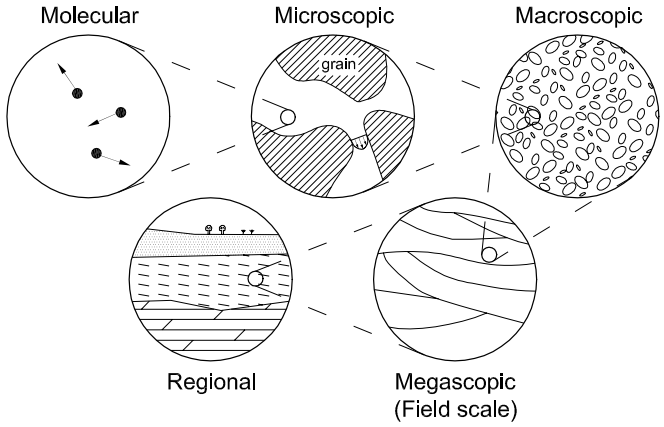

Figure 3.1: Definition of scales.

on the microscopic scale it is essential when modelling transport processes to introduce a dispersion process.

3.1.2 Parameter scale

Darcy's law is the foundation for describing laminar flow in porous media. It is used directly in the derivation of the governing equations for groundwater flow, and in the laboratory for estimating the flow properties of a given soil sample. Darcy's law combined with a mass balance equation, together with the relevant dispersion relations, constitute the governing equations on a macroscopic scale. It is thus straightforward to discretise and solve the equations with related parameters on this scale. However, in most cases the discretisation scale is much coarser than the macroscopic scale, and there is no guarantee that parameter values on the measured scale are representative of the model scale. The term "effective parameter" is often used for parameters on the model scale, to suggest the fact that the parameters have no direct physical meaning. Figure 3.2 illustrates the typical evolution in a given parameter as a function of averaging volume. It can be seen that a number of different Representative Elementary Volumes (R.E.V.) can be found, depending on the scale under consideration.

In principle measured parameters can be used in the governing equations only on the scale on which they are found. So the question is: is the scale on which parameters are measured identical with the scale of the numerical models applied to a given problem? There is no general answer to this question, but in most cases the answer is negative.

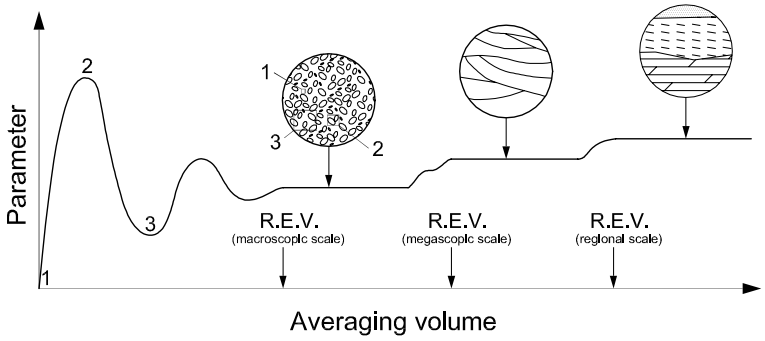

Figure 3.2: Parameter value versus averaging volume.

Figure 3.3 is an attempt to illustrate the parameter scale in relation to the method of measurement and the scale of application, which depends on which type of model is used.

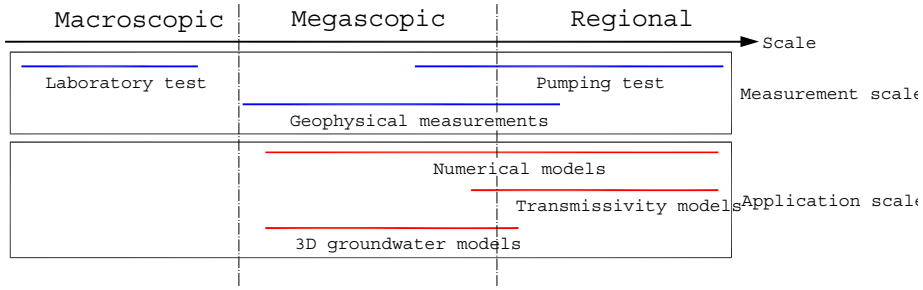

Figure 3.3: Measurement and application scales.

Geophysical measurements are in general used only to identify geological units, and it is very difficult to use them as the basis for estimating hydrogeological parameters. So long as interpretative assumptions are fulfilled, pumping tests are cabable of predicting aquifer properties on the megascopic and regional scales. Laboratory tests are typically performed on the macroscopic scale, and in the case of undisturbed soil samples the test might give a good estimate of parameter properties on the given scale.

From Figure 3.3 it can be seen that only geophysical measurements seem to have the scale on which 3D groundwater models are normally applied. Given the above mentioned limitations of geophysical measurements, no method of measurement is suited to giving precise parameter estimates on the megascopic scale on which 3D groundwater models are normally implemented.

Thus on the one hand the various methods of measurement may not yield precise parameter estimates while, on the other, the results may substantially reduce the possible parameter range.

3.2 Observation data

Calibration is performed with respect to past system behaviour. System behaviour is typically described by observations in time and space e.g. of water level, head potential and river discharge. But in principle all observations that in some way condition the parameter estimates are valuable. This section describes the different types of observations that can be incorporated into a calibration procedure and, where possible, the sources of mismatch between observed and computed values are described.

3.2.1 Head data

Head and water level observations originate from at least three types of surveys: a) logging in connection to newly established wells, b) more or less continuous logging of existing wells and c) synchronous observations of a large number of wells.

A mismatch between calculated and observed head data may originate from at least three sources: 1) observation errors, 2) scale errors and 3) errors due to time effects. The following description is inspired by Sonnenborg (2001).

Observations errors: The head observation errors in this description consist of errors directly associated with measurement, rather than all sources of mismatch between computed and observed heads, such as may be found in other literature. The contributions to head observation errors are:

- Finite precision of the measuring equipment. The expected error contribution is in the range of a **few millimetres** to a **few centimetres**.
- Manual reading of instruments and registration of the results. The expected error contribution is in the range of a **few millimetres** to a **few centimetres**.
- Incorrect or imprecise well reference level. References levels may be determined with considerable accuracy by using a levelling instrument or GPS. In this case the error contribution is typically in the range of a **few millimetres** to a **few centimetres**. Alternatively the reference level may be

determined from topographical maps and the error contribution may be several **metres**, depending of the quality and resolution of the maps.

- Atmospheric pressure. Head potential in confined aquifers fluctuates in proportion to fluctuations in the atmospheric pressure. The atmospheric pressure fluctuations are rarely taken into account in groundwater models and may lead to an error contribution of up to 0.1 m. The change in potential head due to a change in atmospheric pressure can be estimated from

$$\psi - \psi_0 = -\frac{BE}{\gamma}(p_a - p_{a0}) \tag{3.2}$$

  where γ [kg m s−2 m−3] is the gravitational body force, BE is the barometric efficiency defined as the ratio of water level change, $(\psi - \psi_0)$ [m], to the atmospheric pressure, $(p_a - p_{a0})$ [bar]. BE has been observed in the range of 0.25 to 0.75. (Bear 1972).

  A pressure difference of 20 mb (the passage of a storm depression) and a barometric efficiency of 0.5 yield a water level change of $\sim$ 0.1 m.

Scale errors: Scale errors originate partly from the finite discretisation of the computational cells and partly from the discretisation of the parameters involved. The main contributions to scale errors are:

- Horizontal discretisation. Head values are calculated at the centre of each computational cell and the observed heads may be located anywhere in the computational cell. A comparison of observed heads with the nearest calculated head value may result in an error contribution up to $0.5Jdx$, where J is the horizontal gradient of the hydraulic head and dx is the size of the computational cell. If the computed head is interpolated into the location of the observed head the error contribution is significantly smaller.

- Vertical discretisation. Observed head values are representative of head values ranging from the minimum to the maximum head in the filter section. Typically the head in the high yield formations in a filter section will have a strong influence on what is measured. Simulated heads are average head in the computational layer. Given that the filter section is identical with the computational layer, the maximum error contribution is of the magnitude $0.5J_v dz$, where J_v is the vertical hydraulic gradient over the filter section, and dz is the computational layer thickness. Filter information may in many cases be incomplete and discussions as to which geological unit has actually been measured may occur. The error contribution arising from an imperfect match between filter levels and computational layers is difficult to quantify, since information on vertical gradients is rare. The

error contribution from monitoring the wrong geological unit (e.g. monitoring the secondary aquifer believing that is is the primary aquifer) may be up to **several metres**.

- Topographical variations. Topographical variations within a computational cell may result in variations in the potential head that will not be represented by the numerical model. The error contributions may be significant in near-surface reservoirs, and can be assumed to be proportional to the topographical variations, and inversely proportional to the depth below ground surface (Sonnenborg 2001).

- Hydrogeological heterogeneity. Parameter heterogeneity cannot be resolved at a finer scale than the resolution of the numerical grid. Usually heterogeneity within a geological unit is ignored and the parameters are constants within regions much larger than the grid. The ignored heterogeneity will contribute to the error between computed and observed heads where the standard error, s_h, may be formulated as a function of the hydraulic gradient J, the standard deviation of the log-transformed (natural) hydrological conductivity, $\sigma_{\ln K}$, and the length scale, α_l. The error contribution may be approximated as (Gelhar 1986).

$$s_h = \sqrt{\frac{1}{3} J^2 \sigma^2_{\ln K} \alpha_l^2} \tag{3.3}$$

  The length scale, α should be chosen as the minimum of: a) the correlation length of the conductivity field, b) extension of the region with constant parameters. In the case of fully distributed parameters (with a different value in each cell) the grid dimension should be used.

  Errors due to hydrogeological heterogeneity may alternatively be approximated from head observation in closely positioned wells. The difference in observed heads, subtracting all other sources of error, may represent the errors due to hydrogeological heterogeneity (Sonnenborg 2001).

Errors due to time effects: The number of head observations in time and space is often limited, and all available observations have to be used. Typically the set of chosen head observations is a mixture of equidistant time series, non-equidistant time series (few observations) and single observations. The data set may be incomplete in a variety of ways, and this may contribute to error. The description below focuses on the error contribution in relation to stationary models.

In stationary models time-average head conditions are simulated, and present head observations may not represent stationary conditions. This may lead to errors. The size of these errors can be approximated by analysing the time series for neighbouring wells. The aim of this analysis is to establish yearly fluctuations

and trends, seasonal fluctuations and more rapid fluctuations. Yearly fluctuations and trends represent errors due to adapting head data from one year in the calibration of another year. Seasonal fluctuations represent the errors arising from presenting a seasonal representative observation as the average yearly head. The rapid fluctuations represent errors due to ignored small time-scale variations e.g. in infiltration and abstraction. Useful statistics could be: a) average weekly standard deviation of daily values, b) average seasonal standard deviation of weekly averages, c) average yearly standard deviation of seasonal averages, d) trends in yearly average. Sonnenborg (2001) has suggested $\Delta H/2$, where ΔH is the difference between the minimum and maximum value in the time series, as a simple measure of the error contribution.

Table 3.1 summarises the different contributions to the mismatch between observed and computer head. Values and intervals are approximate and based on Danish conditions.

Table 3.1: Standard deviation of error contributions. Inspired by Sonnenborg (2001)

	Error contribution	Standard deviation [m]	Reference
Observation error	Measuring equipment	0.03	
	Reading and bookkeeping	0.05	
	Reference level	0.05 - 2.0	
	Atmospheric pressure	0.0 - 0.15	
Scale error	Horizontal discretisation	$0.5\Delta x J$ [1)]	Sonnenborg 2001
	Vertical discretisation	$0.5 J_v dz - 2.0$ [2)]	
	Topographical variations	σ_{topo}/d [3)]	Sonnenborg 2001
	Heterogeneity	$\sqrt{\frac{1}{3} J^2 \sigma^2_{ln\,K} \alpha^2_l}$ [4)]	Gelhar 1986
Time effects	Non-stationarity	$\Delta H/2$ [5)]	Sonnenborg 2001
Total error		$\sqrt{\sum \sigma^2}$	

1. Δx is the horizontal discretisation and J is the hydraulic gradient.
2. Δy is the vertical discretisation and J_v is the vertical gradient.
3. σ_{topo} is the standard deviation of the topography within a computational grid. d is the depth from ground surface to groundwater.
4. $\sigma_{\ln K}$ is the standard deviation of the log-transformed (natural) hydrological conductivity. α is the minimum of the correlation length for $\ln K$ and extension of the region with constant parameters.
5. ΔH is the difference between the minimum and maximum value in the time series.

3.2.2 River discharges

Measurements of river discharge are regularly carried out in major streams and rivers. This is done by measuring water levels and deriving the actual flow from these measurements, using a mathematical relation between water level and flow (Q-ψ relation). These point flow velocity measurements are then transformed into discharges. The observation data originate from a number of continuously logged gauging stations and/or from synchronous measurements of the flow in a large number of cross sections.

As in the case of head data, the mismatch between calculated and observed river discharges may originate from observation errors, scale errors and errors due to time effects.

Observation errors: Observation errors originate from the registration of water levels, the registration of flow velocities and the subsequent transformation of these measurements into discharges.

- Discharge derived from water level registration. Water level measurements are transformed into discharges by using a Q-ψ relation, describing the relationship between water level and discharge. The Q-ψ relation for a given cross section is dynamic and will change with the amount of vegetation and possible erosion. The uncertainty of the discharge estimate is in the order of 10% (Blicher 1991).

- Discharge derived from flow velocity measurements. When discharge is estimated from a number of point velocity measurements in well defined cross sections, the level of uncertainty is relatively low, probably in the order of 5%.

Scale errors: Scale errors in relation to river modelling are closely related to numerical and parameter discretisation.

- The level of detail. The level of detail at which the geometry of a river can be described depends on numerical discritisation. River branches smaller than grid dimension cannot be described.

- Topographical variations. The interaction between river and groundwater may depend to a considerable degree on small-scale variations in the topography: e.g. seepage flow in local topographical depressions.

- Hydrogeological heterogeneity. Small-scale heterogeneity may be determined for the interaction between groundwater and river.

Definitions

Median value of annual minimum flow is the daily median value of the annual minimum flow.

Error due to time effects: Synchronous observations of summer discharge are performed for the purpose of estimating the base flow contribution to the river. The results may be converted to another year or years from a reference gauging point, on the assumption that the synchronous observation points behave in the same manner as the reference station. Errors may occur, depending on the degree to which this assumption is fulfilled. (Sonnenborg 2001)

Time series data are used to estimate the median value of annual minimum flow. Bjarnov (1987) has established the relation between base flow, q_b $[l/s/km^2]$, and the standard deviation on the median value of annual minimum flow, s_{q_b} $[l/s/km^2]$ reported in Fig. 3.4, for ten Danish stationary gauging stations.

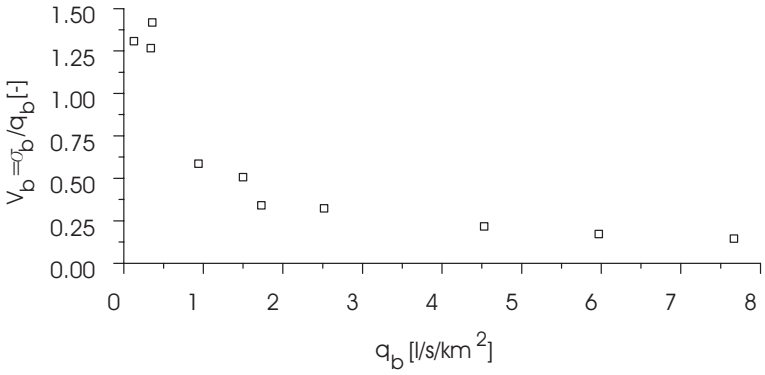

Figure 3.4: Relation between the median value of annual minimum flow, q, and the coefficient of variation of the median value of annual minimum flow, s_q for 10 Danish stationary gauging stations. Estimation based on a 65-year time series

3.2.3 Concentration data

Concentration data are available from regular testing in abstraction wells and from surveys of polluted sites. Concentration data can be valuable as tracer data in cases where the source and leakage period are known, for example in the case of gasoline pollutions with MTBE spill. MTBE was added to gasoline in the mid-eighties and the source location is often well described. The calibration of concentration levels from point pollutions is in general associated with large

errors due to: 1) the density of the concentration data in time and space in comparison with the extension of the pollution, 2) great uncertainty in the description of degradation and sorption, 3) hydrological heterogeneity. The effects of 1) and 3) may be averaged out if area sources, such as nitrate or pesticides, are considered.

3.2.4 Tracer data (natural and artificial)

Natural tracer data such as CFC gas may be very useful in determining groundwater age. The uncertainty of the age estimate depends on the degree of mixture. The water in abstraction wells is typically a mixture of water from large regions with different travel time, and for this reason estimates of its age may be very uncertain. In observation wells with narrow filter intervals one can assume a lower level of uncertainty, since the only mixture to occur here is natural.

Because groundwater motion is usually slow, artificial tracers are often used over short distances, normally in order to constrain the flow velocities in a given deposit. The uncertainty of flow velocity estimates is due to at least two factors: the finite number of observation points and the existence of small-scale hydrological heterogeneity.

3.2.5 Subjective observations

Observations by local farmers and citizens offer an alternative form of data where hydrological conditions are otherwise ungauged. A local farmer may for example observe that the "creek on his land dries out every summer" or that his "spring yields water only in wet years". It is challenging to incorporate subjective observations into an automatic calibration procedure, but such information may be very valuable.

3.2.6 Weighting of observations

If the aim in calibration is to incorporate different kinds of observations (e.g. head and river discharge) and/or observations of the same kind with varying levels of expected error, the observations have to be mutually weighted. The standard procedure is to weight the observations according to the inverse of the estimated error covariance matrix, $\mathbf{V} = \mathbf{C}^{-1}$.

The correlation between observations is usually very difficult to determine and in general observations are assumed to be uncorrelated. Positive correlation may, however, be a useful tool in ensuring certain trends. In many situations, for

example, it will be equally crucial to ensure a low average mean value of the absolute head residuals and at the same time a high level of agreement between the observed head gradients and simulated head gradients. A high degree of positive correlation will ensure weighting of the gradients.

3.3 Ill-posedness

Well-posedness is a fundamental requirement in applying traditional calibration procedures to a calibration problem. A well-posed calibration problem is defined as one that has a parameter solution that is **identifiable**, **unique** and **stable**.

The solution is identifiable if it can be found from the observation set, and it is unique if only one such solution can be found.

The solution is thus unique if one and only one set of models, parameters and variables can be established from the set of observations. A necessary, but not sufficient, requirement for uniqueness is that the number of models, parameters and variables to be estimated is less or equal to the number of observations. The observations have to be spatially distributed and preferably of different types (e.g. head, stream flow, tracer, etc.).

The solution is stable if small changes in the observations produces small changes in the parameter solutions. Instability often arises from the fact that the parameter solution is non-identifiable, or only poorly so. It is generally associated with objective functions that are flat or nearly flat in the region around the parameter optimum. Unstable problems can result in parameter solutions that are very sensitive to the starting point of the search (Carrera and Neuman 1986b).

A thorough description of (in)stability, (non-)uniqueness and (non-)identifiability can be found in Carrera and Neuman (1986b) and Yeh (1986).

As the above indicates, ill-posedness thus depends on parameterisation - any model can be rendered well-posed by reducing the number of parameters. One should however bear Fig. 3.5 in mind.

Figure 3.5 shows the contributions to the total modelling error as a function of the number of parameters involved. As the number of parameters increases the error contribution from model error decreases, and the contribution from parameter error increases. When the model in simplified in order to make the calibration problem unique, there is a corresponding increase in the contribution from model error.

Example 3.1 *If we consider example 1.1 on page 7 again, and assume that only*

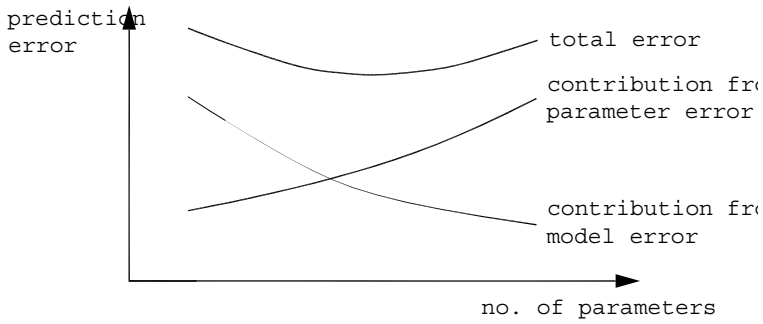

Figure 3.5: Contributions to the total modelling error as a function of the number of parameters

head observations are present, this leaves us with two observations, h_2^* and h_3^*, and two unknowns, q and T. The first requirement of uniqueness is fulfilled (no. of observations is greater than or equal to the no. of unknowns). However, the problem is non-unique because only one type of observation (head) is present. An infinite number of combinations of q and T will result in the same head distribution - only the ratio between q and T can be found. The problem becomes unique if the river inflow estimate, Q_r, is included. Alternatively we could consider q as being deterministic and take only T as the target of calibration. By assuming that q is known, we increase the amount of model error.

3.4 Non-linearity and discontinuity

Groundwater problems are in general non-linear. This non-linearity results from: 1) changing water levels in free reservoirs, leading to changes in cross section area/transmissivity, 2) discontinuous or non-linear sink-source terms, 3) spatial variations in the hydrogeological parameters in the flow domain and 4) threshold-dependent processes such as groundwater pumping, drainage flow or surface flow. Figure 3.6 illustrates a schematic aquifer system and the response of the head potential in the lower aquifer as a function of the change of the hydrological conductivity in the lower aquifer.

At low conductivities only a small amount of water will leave the model through the lower aquifer. The model will instead generate surface flow, drain flow and horizontal flow in the upper aquifer. As the conductivity increases the model predicts a drop in the head potential in the lower aquifer and for large conductivities horizontal flow will only appear in the lower aquifer. In some conductivity regions the head response (response surface) may be close to linear, while in oth-

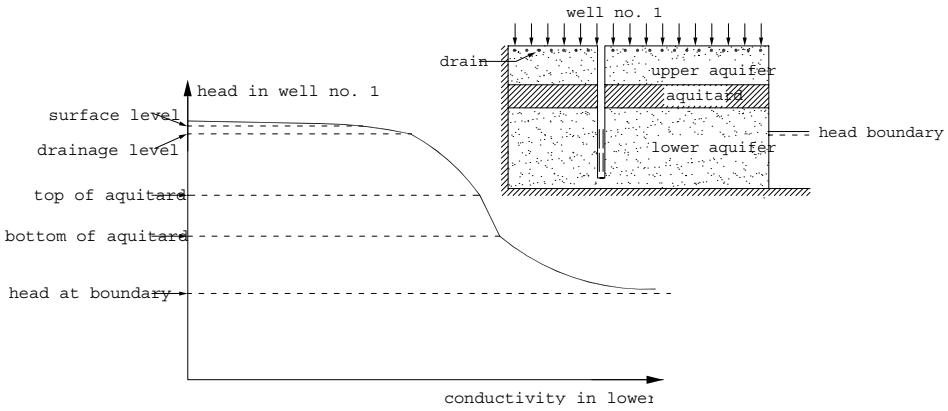

Figure 3.6: Example of head development in an aquifer due to changes in the conductivity - effect of threshold-dependent processes, discontinuous sink-source terms, spatially varied hydrogeological parameters and transmissivity.

ers it will be strongly non-linear. A plateau and a valley in the response surface are found at respectively low and high conductivity values.

This imaginary example illustrates a response surface that has a plateau, a valley and varying degree of non-linearity. Furthermore the response surface is non-differentiable in a number of points. These circumstances play an important role in estimating parameters. Plateaus, valleys and a rapid change in the gradient of the response surface may result in problems when using gradient search methods for optimisation, and models are often linearised in order to stabilise the optimisation: see e.g. Christensen et al. (1998). The subsequent estimation of parameter statistics and the prediction of uncertainties are often based on linearised estimates, which may be misleading for strong non-linear models.

3.5 Least square method

Non-linear least square methods are widely used for parameter estimation in groundwater models (Cooley 1977; Cooley 1979; Hill 1992). The standard least square parameter estimate results from a minimisation of

$$J(\boldsymbol{\theta}) = \left[\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})\right]^T \mathbf{C}_{\boldsymbol{\psi}^*}^{-1} \left[\boldsymbol{\psi}^* - \boldsymbol{\psi}(\boldsymbol{\theta})\right] \tag{3.4}$$

where $\boldsymbol{\psi}^*$ is the vector of observed state variables, $\boldsymbol{\psi}(\boldsymbol{\theta})$ is the vector of computed state variables given the parameter set $\boldsymbol{\theta}$, and $\mathbf{C}_{\boldsymbol{\psi}^*}$ is the expected error covariance of the observed state variables.

If any prior parameter information exists Eq. 3.4 is expanded to

$$J(\boldsymbol{\theta}) = (\mathbf{z}^* - \mathbf{z})^T \, \mathbf{C}_{z^*}^{-1} \, (\mathbf{z}^* - \mathbf{z}) \tag{3.5}$$

where

$$\begin{aligned}
\mathbf{z}^* &= (\boldsymbol{\psi}^*, \boldsymbol{\theta}^*)^T \\
\mathbf{z} &= (\boldsymbol{\psi}, \hat{\boldsymbol{\theta}})^T \\
\mathbf{C}_{z^*} &= \begin{pmatrix} \mathbf{C}_{\boldsymbol{\psi}^*} & 0 \\ 0 & \mathbf{C}_{\boldsymbol{\theta}^*} \end{pmatrix}
\end{aligned}$$

$\boldsymbol{\theta}^*$ is the expected value of the prior parameter vector, $\mathbf{C}_{\boldsymbol{\theta}^*}$ is the covariance of the prior parameter estimate and $\hat{\boldsymbol{\theta}}$ the parameter estimate.

Example 3.2 *In example 1.1, page 7, we want to estimate q and T from the observations ψ_2^*, ψ_3^* and Q_r^* including the prior information on q^* and T^*. From the various sources of uncertainty described in section 3.2 the expected error of ψ_2^*, ψ_3^* and Q_r^* is estimated to $\sigma_{\psi_2^*}$, $\sigma_{\psi_3^*}$ and $\sigma_{Q_r^*}$. From example 1.1 we have the prior information on q^* and T^* $(\mu_q^*, \sigma_q^*, \mu_T^*, \sigma_T^*)$. The transmissivity T is log-transformed before estimation. If the observations and parameters are assumed to be uncorrelated, the components of Eq. 3.5 are given as*

$$\begin{aligned}
\mathbf{z}^* &= (\psi_2^*, \psi_3^*, Q_r^*, \mu_q, \mu_{\ln T})^T \\
\mathbf{z} &= (\psi_2(\hat{q}, \hat{T}), \psi_3(\hat{q}, \hat{T}), Q_r(\hat{q}, \hat{T}), \hat{q}, \hat{T})^T \\
\mathbf{C}_{z^*} &= \begin{pmatrix}
\sigma_{\psi_2^*} & 0 & 0 & 0 & 0 \\
0 & \sigma_{\psi_3^*} & 0 & 0 & 0 \\
0 & 0 & \sigma_{Q_r^*} & 0 & 0 \\
0 & 0 & 0 & \sigma_q & 0 \\
0 & 0 & 0 & 0 & \sigma_{\ln T}
\end{pmatrix}
\end{aligned}$$

when inserted in Eq. 3.5 this gives

$$\begin{aligned}
J(\hat{q}, \hat{T}) = & \frac{1}{\sigma_{\psi_2^*}} (\psi_2^* - \psi_2(\hat{q}, \hat{T}))^2 + \frac{1}{\sigma_{\psi_3^*}} (\psi_3^* - \psi_3(\hat{q}, \hat{T}))^2 + \frac{1}{\sigma_{Q_r^*}} (Q_r^* - Q_r(\hat{q}, \hat{T}))^2 \\
& + \frac{1}{\sigma_g} (\mu_q - \hat{q})^2 + \frac{1}{\sigma_{\ln T}} (\mu_{\ln T} - \ln \hat{T})^2
\end{aligned} \tag{3.6}$$

The estimate of $\hat{q}$ and $\hat{T}$ are found by minimising Eq. 3.6.

3.6 Maximum likelihood method

The maximum likelihood estimate is the conditioned density

$$p_{\boldsymbol{\psi}^*|\boldsymbol{\theta}}\left(\boldsymbol{\psi}^*|\boldsymbol{\theta}\right) = f_{\boldsymbol{\psi}^*|\boldsymbol{\theta}}\left(\boldsymbol{\psi}^*|\boldsymbol{\theta}\right) \tag{3.7}$$

which is commonly called the likelihood function. $f_{\boldsymbol{\psi}^*|\boldsymbol{\theta}}\left(\boldsymbol{\psi}^*|\boldsymbol{\theta}\right)$ is the joint probability density function of the observations. The maximum likelihood estimate is a maximisation of Eq. 3.7. In some cases it is convenient to derive the maximum likelihood estimate from a minimisation of the log-likelihood criterion

$$-2\ \ln\ p_{\boldsymbol{\psi}^*|\boldsymbol{\theta}}\left(\boldsymbol{\psi}^*|\boldsymbol{\theta}\right) \tag{3.8}$$

If the residuals are assumed to be Gaussian distributed the likelihood function becomes

$$f_{\boldsymbol{\psi}^*|\boldsymbol{\theta}}\left(\boldsymbol{\psi}^*|\boldsymbol{\theta}\right) = (2\pi)^{\frac{N}{2}}\ |\ C_{\boldsymbol{\psi}^*}\ |^{-\frac{1}{2}}\ e^{\left(\frac{1}{2}(\boldsymbol{\psi}^*-\boldsymbol{\psi}(\boldsymbol{\theta}))^T C_{\boldsymbol{\psi}^*}^{-1}(\boldsymbol{\psi}^*-\boldsymbol{\psi}(\boldsymbol{\theta}))\right)} \tag{3.9}$$

if prior parameter information is included Eq. 3.9 becomes

$$f_{\boldsymbol{\psi}^*|\boldsymbol{\theta}}\left(\mathbf{z}^*|\boldsymbol{\theta}\right) = (2\pi)^{\frac{N}{2}}\ |\ C_{\mathbf{z}^*}\ |^{-\frac{1}{2}}\ e^{\left(\frac{1}{2}(\mathbf{z}^*-\mathbf{z})^T C_{\mathbf{z}^*}^{-1}(\mathbf{z}^*-\mathbf{z})\right)} \tag{3.10}$$

The parameter set found from minimising 3.10 corresponds to the parameter set found from minimising the least square objective function.

3.7 Single versus multi-objective parameter estimation

Some criticism has been raised against using a single objective function in the estimation of parameters. Yapo et al. (1998) state that *"Practical experiences with model calibration suggests that any single-objective function, no matter how carefully chosen, may not adequately measure the ways in which the model fails to match the important characteristics of the observed data."*

In order to overcome this problem the calibration problem can be formulated as a multi-objective optimisation problem of $F(\boldsymbol{\theta})$

$$F(\boldsymbol{\theta}) = f_1(\boldsymbol{\theta}), f_2(\boldsymbol{\theta}), ..., f_m(\boldsymbol{\theta}) \tag{3.11}$$

where $f_1(\boldsymbol{\theta}), ..., f_m(\boldsymbol{\theta})$ are objective functions to be simultaneously minimised with respect to the parameters $\boldsymbol{\theta}$. $f_i(\boldsymbol{\theta})$ might e.g. be the least square or maximum likelihood parameter estimate with respect to head observations and $f_j(\boldsymbol{\theta})$

might be the parameter estimate with respect to river discharge observations. A minimisation of the individual objective functions may result in parameter solutions that are not unique and accordingly it is not possible to find the best solution using objective methods.

Figure 3.7 illustrates a problem with two objectives (f_1, f_2) to be minimised with respect to two parameters (θ_1, θ_2). Point A is the solution to minimising f_1 and point B is the solution to minimising f_2

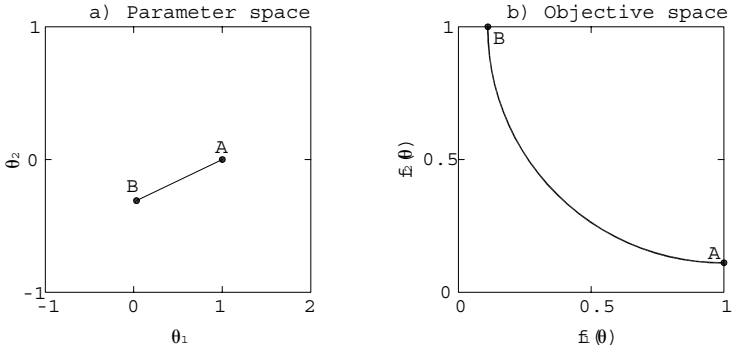

Figure 3.7: Illustration of Pareto optimality. After Yapo et al. (1998)

The solution to the multi-objective problem consists of all parameter combinations on the line from point A to point B. Parameter sets close to point B will result in a small of value of f_1, and as they move towards point A f_1 will decrease and f_2 will increase. Solutions on the line from A to B are called Pareto solutions. (Yapo et al. 1998) All Pareto solutions are acceptable simulators of the system, and the parameter range given by the Pareto solutions reveals the uncertainty due to the choice of objective function. A thorough description of the multi-objective approach and solution methods for minimising multi-objectives can be found in Yapo et al. (1998).

Example 3.3 *Let us consider the flow situation in example 1.1, page 7. Two objective functions are formulated: 1) the least square measure of the head in wells nos. 2 and 3:*

$$J_\psi(\hat{q}, \hat{T}) = \frac{1}{\sigma_{\psi_{h_2^*}}}(\psi_2^* - \psi_2(\hat{q}, \hat{T}))^2 + \frac{1}{\sigma_{\psi_{h_3^*}}}(\psi_3^* - \psi_3(\hat{q}, \hat{T}))^2 \tag{3.12}$$

and 2) the least square measure of the river inflow:

$$J_{Q_r}(\hat{q}, \hat{T}) = \frac{1}{\sigma_{\psi_{Q_r^*}}}(Q_r^* - Q_r(\hat{q}, \hat{T}))^2 \tag{3.13}$$

The minimisation of J_ψ and J_{q_r} will lead to two estimates of q and T that correspond to points A and B on Figure 3.7. If the two objectives are combined

$$\begin{aligned} J_{\psi,Q_r}(\hat{q},\hat{T}) =& \omega(\psi_2^* - \psi_2(\hat{q},\hat{T}))^2 + (\psi_3^* - \psi_3(\hat{q},\hat{T}))^2 \\ &+ (1-\omega)(Q_r^* - Q_r(\hat{q},\hat{T}))^2 \end{aligned} \tag{3.14}$$

and the weight, ω, is varied from 1 to zero, solutions on the Pareto front can be found.

3.8 Parameter statistics

Parameter statistics can be analysed in case of linear models that have non-biased and Gaussian residuals. The covariance of the estimated parameters can be found from Bard (1974):

$$\mathbf{C}(\hat{\boldsymbol{\theta}}) = \frac{J(\hat{\theta})}{N_{obs} - N_{par}} \left[\mathbf{A^k}(\hat{\boldsymbol{\theta}})\right]^{-1} \tag{3.15}$$

where

$\hat{\boldsymbol{\theta}}$ estimated parameter vector
$J(\hat{\boldsymbol{\theta}})$ least square error
$\mathbf{A^k}(\hat{\boldsymbol{\theta}}) = \left[\mathbf{J_D}\left(\boldsymbol{\theta^k}\right)\right]^T \left[\mathbf{J_D}\left(\boldsymbol{\theta^k}\right)\right]$,$(N_{par} \times N_{par})$

$\mathbf{J_D}$ Jacobian matrix of state variable, ψ, with respect to parameters, θ, $(M \times L)$

$$\mathbf{J_D} = \begin{pmatrix} \frac{\partial\psi_1}{\partial\theta_1} & \frac{\partial\psi_1}{\partial\theta_2} & \cdots & \frac{\partial\psi_1}{\partial\theta_{N_{par}}} \\ \frac{\partial\psi_2}{\partial\theta_1} & \frac{\partial\psi_2}{\partial\theta_2} & \cdots & \frac{\partial\psi_2}{\partial\theta_{N_{par}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\psi_{N_{obs}}}{\partial\theta_1} & \frac{\partial\psi_{N_{obs}}}{\partial\theta_2} & \cdots & \frac{\partial\psi_{N_{obs}}}{\partial\theta_{N_{par}}} \end{pmatrix}$$

N_{obs} number of observations
N_{par} number of parameters

The Jacobian is a local linear approximation of the response surface in the region around the optimal solution, $\hat{\boldsymbol{\theta}}$.

The size of the components of $\mathbf{A^k}$ is proportional to the number of observations and consequently the estimated parameter variation becomes (through Eq.

3.15) inversely proportional to the number of observations. As the number of observations increases the parameter error will diminish.

The linear confidence intervals for the parameter θ_i can be found (Hill 1992, p. 58: Seber and Wild 1989, p. 191-194):

$$\theta_i \pm t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right) s_{\theta_i} \tag{3.16}$$

where $t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right)$ is the Student-t statistic for $N_{obs} - N_{par}$ degrees of freedom and a significance level of α and $s_{\theta_i} = \sqrt{C(\theta_i, \theta_i)}$ is the standard deviation of θ_i.

Example 3.4 *We now want to consider the statistics for the parameter estimate described in example 3.2, page 45. From Eq. 3.15 we have the covariance of the estimated parameters:*

$$\mathbf{C} = \begin{pmatrix} \sigma_{\hat{q}}^2 & \sigma_{\hat{q}}\sigma_{\hat{T}} \\ \sigma_{\hat{q}}\sigma_{\hat{T}} & \sigma_{\hat{T}}^2 \end{pmatrix} = \frac{J(\hat{q}, \hat{T})}{3-2} \left[\mathbf{A^k}(\hat{\boldsymbol{\theta}})\right]^{-1} \tag{3.17}$$

where $\mathbf{A^k}(\hat{\boldsymbol{\theta}}) = \mathbf{J_D}^{\mathbf{T}}\mathbf{J_D}$ *and the Jacobian matrix,* $\mathbf{J_D}$*, is given as*

$$\mathbf{J_D} = \begin{pmatrix} \frac{\partial \psi_2}{\partial \hat{q}} & \frac{\partial \psi_2}{\partial \hat{T}} \\ \frac{\partial \psi_3}{\partial \hat{q}} & \frac{\partial \psi_3}{\partial \hat{T}} \\ \frac{\partial Q_r}{\partial \hat{q}} & \frac{\partial Q_r}{\partial \hat{T}} \end{pmatrix} \tag{3.18}$$

The 95% confidence intervals of $\hat{q}$ *and* $\hat{T}$ *are*

$$\begin{aligned} \hat{q} \quad &\pm t\,(3-2, 0.975)\,\sigma_{\hat{q}} \\ \hat{T} \quad &\pm t\,(3-2, 0.975)\,\sigma_{\hat{T}} \end{aligned} \tag{3.19}$$

3.9 State variable statistics

Linear confidence intervals for any system state variables ψ_l are given as (Hill 1992 ,p. 58).

$$\psi_l \pm t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right) s_{\psi_l} \tag{3.20}$$

where s_{ψ_l} is the standard deviation of ψ_l calculated from

$$s_{\psi_l} = \left[\sum_{i=1}^{N_{par}} \sum_{j=1}^{N_{par}} \frac{\partial \psi_l}{\partial \theta_j} C_{\theta_i, \theta_j} \frac{\partial \psi_l}{\partial \theta_i}\right]^{\frac{1}{2}} \tag{3.21}$$

where C_{θ_i,θ_j} are components in the covariance matrix for $\boldsymbol{\theta}$

Approximate linear prediction intervals are calculated as (Hill 1994 ,p. 32)

$$\psi_l \pm t\left(N_{obs} - N_{par}, 1.0 - \frac{\alpha}{2}\right)\left(s^2_{\psi_l} + \frac{s_r^2}{\omega_l}\right)^{1/2} \tag{3.22}$$

where s_r is the calculated standard error of the regression and ω_l is a weight that equals σ_r^2/σ_l^2, where σ_r^2 is the estimated common error variance of the regression, and σ_l^2 is the measurement error variance associated with ψ_l. The error variance of ψ_l is not usually known. Christensen and Cooley (1999) assumed that σ_l^2 was spatially distributed proportionally to the variance of the observed head measurements. Error in predictions becomes equally distributed according to the assumed errors deriving from observation error.

Example 3.5 *We now want to consider the statistics for the estimate of ψ_1, see example 1.1, page 7, and 3.2, page 45. From Eq. 3.21 we have the standard deviation on ψ_1*

$$\begin{aligned} s_{\psi_1} =& \frac{\partial\psi_1}{\partial q} C_{q,T} \frac{\partial\psi_1}{\partial T} + \frac{\partial\psi_1}{\partial q} C_{q,q} \frac{\partial\psi_1}{\partial q} \\ &+ \frac{\partial\psi_1}{\partial T} C_{T,q} \frac{\partial\psi_1}{\partial q} + \frac{\partial\psi_1}{\partial T} C_{T,T} \frac{\partial\psi_1}{\partial T} \end{aligned} \tag{3.23}$$

From Eq. 3.20 the 95 % confidence intervals can be found

$$\psi_1 \pm t\,(3-2, 0.975)\, s_{\psi_1} \tag{3.24}$$

3.10 Solving the regression problem

3.10.1 Gauss-Newton minimisation

The Gauss-Newton algorithm has often been used as a minimisation algorithm in problems concerning the estimation of groundwater parameters. The algorithm starts with an initial parameter vector $\boldsymbol{\theta}^0$ and converges iteratively to a local minimum. The local optimal parameter set is found from minimisation of (e.g. see Yeh 1986)

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \rho^k \mathbf{d}^k \tag{3.25}$$

with

$$\mathbf{A}^k \mathbf{d}^k = \mathbf{g}^k \tag{3.26}$$

where

$\mathbf{A^k} = \left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]^T \left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]$,$(L \times L)$;

$\mathbf{g^k} = \left[\mathbf{J_D}\left(\theta^{\mathbf{k}}\right)\right]^T \left[\boldsymbol{\psi}(\boldsymbol{\theta^k}) - \boldsymbol{\psi}^*\right]$,$(L \times 1)$;

$\mathbf{J_D}$ Jacobian matrix of state variable, ψ, with respect to parameters, θ, $(M \times L)$

ρ^k scalar step size

$\mathbf{d^k}$ Gauss-Newton direction vector

N_{obs} number of observations

N_{par} number of parameters

A description of the modified Gauss-Newton Optimisation Method, including weighted residuals, can be found in Hill 1992, p. 76-82.

3.10.2 Other methods

An alternative to the Gauss-Newton minimisation is offered by global random search methods, where the parameter space is sampled randomly and global parameter set/sets are estimated on the basis of one or more objective functions.

Other algorithms combine local search methods with random search methods. The Multi-Objective Complex Evolution (MOCOM-UA) method (Yapo et al. 1998) and the shuffled complex evolution (SCE-UA) method (Duan et al. 1992) are examples of such methods. Madsen and Kristensen (2002) applied UCODE (inverse programme for Gauss-Newton minimisation) and the SCE-UA method on a MIKE SHE application and found that "The UCODE solutions were trapped in local optima far from the Pareto front. Even when the initial parameter set was close to the Pareto front, UCODE failed to converge into a Pareto optimal solution."