Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Uncertainty assessment of the geometry of an aquitard using an EnKF and EnKF-GMM

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfilment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Liselot Arkesteijn

Delft, The Netherlands
September 2014

TNO innovation for life

TU Delft

MSc THESIS APPLIED MATHEMATICS

"Uncertainty assessment of the geometry of an aquitard using an EnKF and EnKF-GMM"

LISELOT ARKESTEIJN

Delft University of Technology

Daily supervisor

Prof. dr.ir. F.C. van Geer

Responsible professor

Prof. dr.ir. A.W. Heemink

Other thesis committee members

Dr.ir. M. Verlaan

Ing. A. Lourens has as a supervisor significantly contributed to the realisation of this master thesis.

September 2014

Delft, The Netherlands

Abstract

Hydraulic head data is frequently used for the calibration of groundwater flow models, usually with the main objective to improve the model performance. The hydraulic parameters in a groundwater flow model depend on the underlying geometry of the subsurface and the permeability of the soil. In the current calibration routines, the geometry is considered known, such that all uncertainties of the hydraulic parameters are ascribed to the permeabilities. After calibration these permeabilities may have physically unreasonable values, indicating that the estimate of the geometry is likely to be wrong.

In this thesis the roles are reversed. A calibration procedure that focuses on the uncertainty of the geometry is performed. This shift introduces a dependency between the hydraulic parameters that is not considered in the current calibration method. An EnKF and EnKF-GMM are applied to a synthetic test case where only the geometry of one aquitard is considered uncertain. The filters are performed with the objectives to 1) localise the extent of the aquitard, and 2) find a probability distribution function for the thickness of the aquitard. In a series of experiments it is shown that the EnKF is only suitable in situations where the exact extent of the aquitard is known. In other cases, the probability distribution function of the thickness is bimodal, resulting in an EnKF that performs poorly and produces physically unrealistic outputs. As a solution, the EnKF-GMM is proposed. This extension of the regular EnKF can be used in more general scenarios as it allows a multimodal distribution as posterior distribution. Roughly speaking, the EnKF-GMM is an application of two EnKFs simultaneously. For two separate populations (modes) an EnKF is performed, and after each iteration the inhabitants of the population have the possibility to cross over to the other population. The cross-over probabilities are based on the likelihood of each population. The results of the experiments show that the EnKF-GMM is able to define probabilities on the extent of the aquitard. However, it fails to find a proper probability distribution function of the thickness. Filter divergence occurs as a result of the insufficient amount of information that hydraulic head data contains.

Acknowledgements

This report is a summary of the results of my master thesis, carried out at TNO Utrecht at the department of Geomodelling. It is the final step to obtain my masters degree in Applied Mathematics at the Delft University of Technology. During this research I have received help and support from many people, to all of whom I would like to express my gratitude. In particular, I would like to thank my supervisor Aris Lourens for being supportive throughout the entire course of this research, providing me the data and synthetic model, and helping me to carry out the calculations at the Deltares linux cluster. I would also like to thank Arnold Heemink and Frans van Geer for offering me this project, and for all their experiences and ideas that they have shared on a regular basis with me throughout the project. In addition, I would like to thank them for proof reading the draft version of this report. Their comments have really improved the quality of this report. Finally, I would like to thank a number of friends for also proof reading the report. In particular Margriet Nieuwenhuis, who read through the entire report, but also a number of other friends that helped to improve the spelling and grammar in parts of the report.

Contents

1 Introduction

The complex structure of the subsurface of The Netherlands is the result of processes taking place over millions of years. Different kinds of soil have been deposited over the past eras and after deposition some of the soil has been displaced by geological hazards such as landslides or faults. Geologists study these processes. One of their activities is to combine historical knowledge and hard data from boreholes to create maps of the subsurface.

These geological maps are useful for many purposes. First of all, it is a tool to convey information on the geology, geological hazards and geological resources of an area. Especially information on resources is valuable for many industries. Secondly, geological maps have a predictive nature. They can, for instance, be used in the planning of construction projects, where it is necessary to know what depth is needed for piles or what the probabilities on subsidence are. Therefore, geological maps play a significant role in the land-use and planning of roads, railways, pipelines, tunnels, dikes and even cities.

Geological maps contain uncertainties as it is impossible to exactly know what soil is present at every location. The maps can be improved if the uncertainty can be adequately assessed. The uncertainty can serve as a guideline to select locations for new boreholes and can furthermore make users aware of the uncertainties. In this thesis, an attempt to assess the uncertainty of geological maps is made, by considering the use of geological maps in groundwater flow models. The variety of strata (layers of sedimentary rock or soil with internally consistent characteristics) and their coherent hydrological properties are partly responsible for the quality and flow of the groundwater. Therefore, extending a geological map to a geohydrological map allows it to be used in groundwater flow models. In this thesis an attempt to assess the structural uncertainty of the subsurface by using observed data of a groundwater flow model is made. Data assimilation techniques are used to integrate the prior geological knowledge with hydraulic head data. This leads to the following problem definition:

1.1 Problem definition

All available geological information about The Netherlands is stored in the digital archive of DINO (Data en Informatie van de Nederlandse Ondergrond), managed and maintained by TNO. The data, especially boreholes, have been interpreted and used to construct various models that describe the subsurface. REGIS II, one of these models, uses the relatively sparse data as an input and combines spatial interpolation and further geological knowledge to estimate the geometry of the subsurface and its geohydrological properties. In geostatistics, many interpolation and simulation algorithms exist, but what all of these algorithms have in common, is that large gaps between data locations will result in unreliable estimates, regardless of the interpolation or simulation algorithm. Adding knowledge from experts' experience might improve the estimates, but errors will always remain.

The geohydrological model REGIS II serves as input data for the physical domain of the groundwater flow model AZURE, that simulates hydraulic heads and flows. At several locations hydraulic heads are measured, and by the application of data assimilation techniques better estimates of the unknown model parameters can become available. The model has been calibrated, but with the main focus on improving the model performance. In this thesis the focus is shifted to calibration with the main purpose to improve the quantification of structural uncertainties of the subsurface.

1.2 Research goals

For AZURE, the geometry itself is not an independent parameter. The thickness of strata is incorporated in the vertical resistance and horizontal transmissivity of strata. A calibration of the geometry can therefore only be indirectly performed by a calibration of the resistances and transmissivities. However, the framework of the current calibration routine does not allow any changes in the extent of strata, and largely ignores the dependencies between the underlying thicknesses. Therefore, a reformulation of the data assimilation framework that allows changes in the extent of strata and incorporates these dependencies is necessary. Furthermore, the main interest here lies in constructing confidence intervals for the thicknesses of strata. This automatically means that first the uncertainties of the resistances and transmissivities of the various strata need to be quantified. The redefined data assimilation framework should allow to translate these uncertainties back to the uncertainties in thicknesses.

A data assimilation method needs to be selected in order to successfully assess the uncertainty of these resistances, transmissivities and the underlying geological structure. After that, this method will be tested on synthetic test cases to evaluate its suitability. Therefore, the following research goals have been formulated to help us achieve the end-goal as sketched in the problem definition:

1. Redefine the data assimilation framework to:
    - allow variations in the extent of strata and honour the dependencies between thicknesses of strata;
    - translate uncertainties in the resistances/transmissivities back to the uncertainty of thicknesses
2. Select a calibration method that is suitable to:
    - improve the estimate of the extent of strata;
    - assess the uncertainties of resistances/transmissivities and thicknesses of strata.
3. Apply the selected method to synthetic test cases in which the thickness of only one aquitard (stratum with low permeability such as clay) is considered uncertain. We aim to obtain posterior probability distribution functions for the extension and thickness of this aquitard and analyse the results.

1.3 Outline

The organisation of the report is as follows. Section 2 provides a short introduction into the geohydrological notions and describes the models that are used during the project. Furthermore, it explains why hydraulic head data may contain information on the geometry. Then, in section 3 a brief literature review of available data assimilation methods is given. In the field of reservoir engineering similar problems are being treated and the possibility to apply techniques used there is studied. An Ensemble Kalman Filter (EnKF) will be selected to gain more insight into the specific behaviour of the model. In section 4 a theoretical formulation of the EnKF is given, along with details on the specific configuration and implementation. After that, section 5 describes the results and a sensitivity analysis of the model. The suitability of the EnKF is thoroughly discussed, and in section 6 a variant of the EnKF, the Ensemble Kalman Filter for Gaussian Mixture Models (EnKF-GMM), is proposed. Section 7 provides the results of the EnKF-GMM applied to the same synthetic test case as in section 5, and finally section 8 concludes.

2 Geohydrology and models

In this thesis hydraulic head data is used to improve the estimates of the geometry of the subsurface. A brief introduction of how these variables quantitatively relate to aquitard thicknesses is provided in this section. It starts with a general introduction into groundwater flow and gives a description of the AZURE model. After that, the geohydrological model REGIS is described and finally the sources of uncertainties in both models and their coherence will be discussed.

2.1 Groundwater flow model

The groundwater in the subsurface is constantly in motion. From areas of recharge the water very slowly moves to areas of discharge, resulting in a journey that can take several months, years, centuries and even millennia. The water flows quickest through aquifers (strata with a high permeability) and slows down significantly when it comes across aquitards (strata with a low permeability). In figure 2.1 (*NERC* (online)) an example of a typical groundwater flow system is given. At the top of the hill lies a recharge area where the water infiltrates. After this process, the water is slowly transported through the subsurface to the lower situated discharge areas.

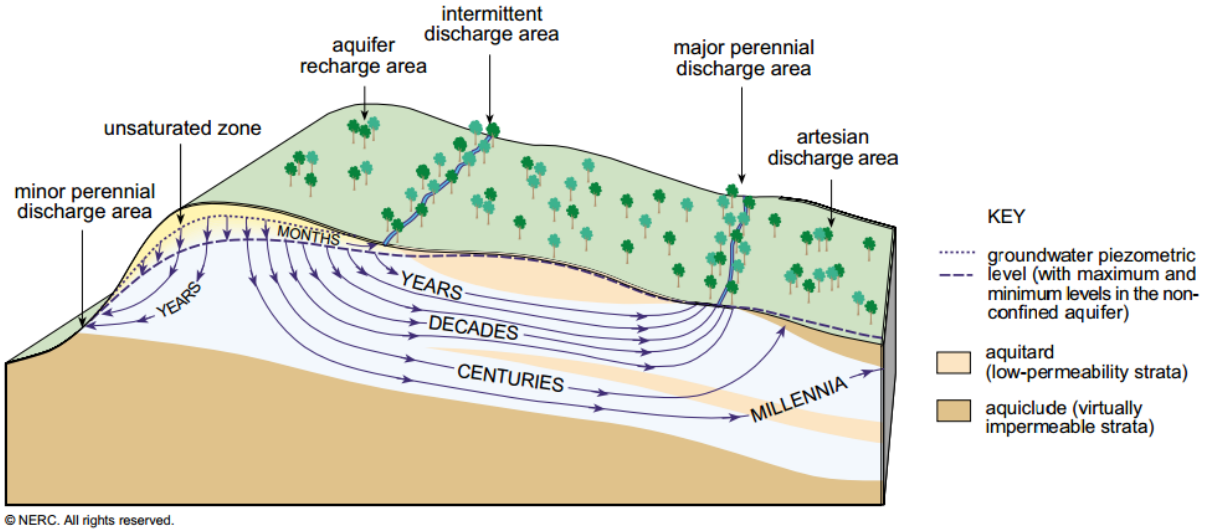


Figure 2.1: Groundwater flow system. From: *NERC* (online)

The driving force behind groundwater flows (and any other type of flow) are the differences in energy potentials. Groundwater flows downhill; from where the potential energy is high to where it is low. The amount of mechanical energy per unit weight that is present at a specific location in a groundwater flow field, can be obtained by summing three potential sources of energy: the pressure energy of the water (pressure head), the potential gravitational energy (elevation head) and the kinetic energy of the flow (velocity head). The latter is negligible compared to the others because of the small velocities of the flow. The summation of the pressure head and elevation head is defined as the hydraulic head.

The hydraulic head is measured at several locations in The Netherlands and has a significant contribution in accurately describing groundwater flows. Figure 2.2 describes a conceptual example of the flow in a 2D polder system to explain the concept and terminology. Two aquifers are present along with a semi-confining aquitard to separate them. At both sides of the domain,

rivers are present, and the water table in the top aquifer is maintained at a constant level. At four locations, the hydraulic heads are measured. The directions of flow can be determined by observation of the measured hydraulic head differences.

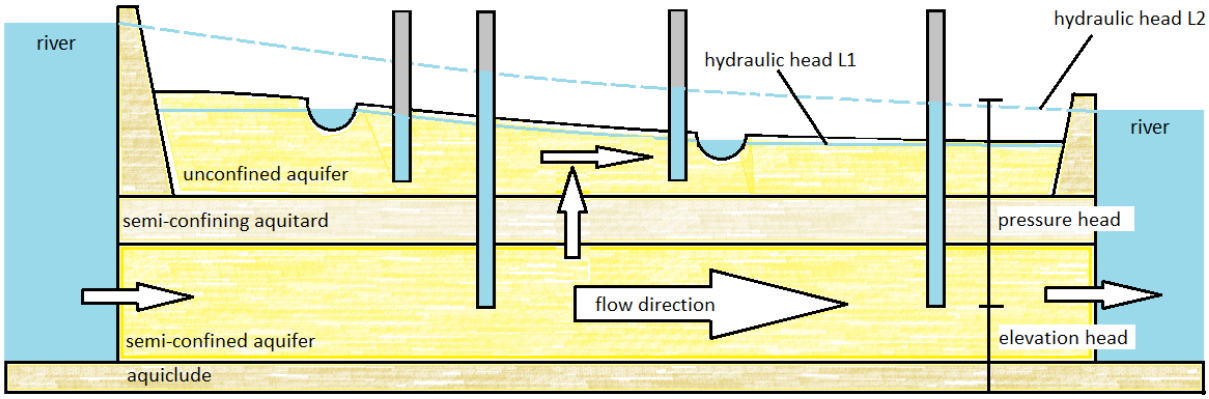


Figure 2.2: Illustrative example of flow in 2D polder system
(based on lecture notes course CIE4420 Geohydrology 2012-2013 M. Bakker TuDelft)

Between the top and the lower aquifer, a semi-confining aquitard is present. This aquitard, often a clay layer, has low permeability and separates the horizontal flows in the aquifers. The hydraulic head in the top aquifer is equal to the water table, while the hydraulic head in the lower aquifer can be higher or lower than the water table, leading to respectively leakage or seepage through the aquitard. Here, the hydraulic head is larger, resulting in an upward vertical flow through the aquitard. The horizontal flow in the aquitard is negligible as a result of the low conductivities.

The relevant issue here is how hydraulic head data can help to improve estimates of the thickness of aquitards. In the case of the simple example above, the hydraulic head difference between the two aquifers is a result of the semi-confining aquitard in between. The aquitard has low vertical permeability and therefore the resistance for water to flow through is much higher than for an aquifer with high permeabilities. This resistance allows a head difference to build up. If the aquitard would have been absent, there would be little resistance and a negligible small head difference would have been measured. This gives the qualitative relationship between the head difference over a layer and the presence of an aquitard.

2.1.1 Mathematical model

The quantitative link between hydraulic heads and groundwater flow was made in 1856 by Henry Darcy. He concluded from a series of experiments that the flow of groundwater is proportional to the gradient of the hydraulic head. His famous law, an empirical formula, is given by:

$$\mathbf{q} = -k\nabla h \tag{1}$$

where $\mathbf{q}$ [LT^{-1}] is the specific discharge vector, k [LT^{-1}] the hydraulic conductivity and h [L] the hydraulic head. In case of anisotropic domains, the constant k is replaced by a tensor K. This tensor K can be represented by a diagonal matrix with main diagonal: $[K_{xx}, K_{yy}, K_{zz}]$ where K_{xx}, K_{yy} and K_{zz} are values of the hydraulic conductivity along the x, y and z coordinate.

Darcy's law can be combined with the mass balance equation to formulate the following frequently used mathematical model to describe the 3D movement of groundwater of constant density through porous media:

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)-W=S_s\frac{\partial h}{\partial t} \tag{2}$$

where
h is the hydraulic head [L]
W represents sources and/or sinks [T^{-1}]
S_s is the specific storage of the porous material [LT^{-1}]
t is the time [T].

This formula can be derived by substituting the fluxes in the mass balance equation by the fluxes from Darcy's law. Note that the specific storage and the hydraulic conductivities may be functions of space, and that the source/sink terms may be functions of space and time. Therefore, equation (2), together with a set of initial and boundary conditions, describes a quite general model for groundwater flow under non-equilibrium conditions in a heterogeneous and anisotropic medium.

2.1.2 Numerical solution

The exact solution of equation (2) only exists for very simple systems. Therefore the solution is usually approximated numerically. Various numerical methods to do so exist. Among those is the finite difference method. In a finite difference method, the continuous system described by equation (2) is replaced by a finite set of discrete points in space and time. The partial derivatives are approximated by terms calculated from the differences in hydraulic head values at these points. This results in a system of equations whose solution yields values of the head at those specific points and times. In a standard finite difference approach in 3D, it is most desirable to partition the domain into cells having both (approximately) uniform hydraulic properties and rectangular cell faces. This means that a cell preferably contains only one type of facies(sand/clay), or in other words, contains material only from either an aquifer or aquitard, and that the boundaries of the strata should coincide with the cell faces. In addition, the grid should preferably not be very fine, as that will greatly increase the required computational effort to solve the problem. However, as the individual strata are continuous bodies that can take almost any shape, it is usually hard to define a coarse rectangular grid with these properties.

The software package MODFLOW was developed to solve the groundwater flow problem in equation (2). Among other routines, there is a routine with a "Quasi 3D" approach in the package that solves all issues described above. The approach uses the principle that flows follow the path of least resistance. This means that the flow in aquitards is almost vertical while the flow in aquifers is by approximation horizontal. Therefore, a single model layer can be used to represent each aquifer whereas the aquitards are removed and are simply represented by a vertical conductance between cells. The 2D layer can be partitioned into rectangles that have approximately uniform hydraulic properties. The finite set of discrete points needed for the finite difference method consists of the centres of all cells. To this node the average hydraulic properties of the entire cell are assigned. The discretisation with respect to time is done using backwards Euler, having as advantage that it is always numerically stable, but with the drawback that it needs to be solved iteratively. The model is first order accurate in time and second order in space (*Harbough* (2005)).

2.1.3 AZURE

AZURE, which stands for "Actueel instrumentarium voor de ZUiderzee REgio", is a geohydrological groundwater flow model maintained by Deltares and is based on the Quasi-Three-Dimensional approach of MODFLOW. In AZURE, all kinds of information are combined to describe the groundwater flow as accurately as possible in the area of the Veluwe until the Utrechtse Heuvelrug, Flevoland and the IJsselmeer (see figure 2.3). The purpose of the model is to provide a tool that can be used in for instance the decision making in water management, climate studies, and in the determination of water levels for maintenance purposes. It was recently developed as a result of a collaboration between Rijkswaterstaat, Vitens, the provinces of Flevoland, Gelderland and Utrecht, the water boards 'Vallei and Veluwe' and 'Zuiderzeeland', Waternet, Deltares, Alterra, Grontmij, Tauw, Royal HaskoningDHV and Acacia Water. (*Hekman et al.* (2014))

Figure 2.3: AZURE model area. From: *Hekman et al.* (2014)

The model can be used to compute hydraulic heads and flows for nine aquifers on a grid with cells of 250x250m. The precision level of input data is 25 meters, making it possible to refine the original grid cells if necessary. Besides MODFLOW, the model MetaSWAP, developed by Alterra, is used for the estimations of the unsaturated zone (soil that lies above the water table). However, in this thesis we will only consider layers deep enough to be certain that they lie in the saturated zone. Therefore we will not discuss this model.

2.2 Geohydrological model

AZURE needs a description of the subsurface as an input for the nine layers of the Quasi-3D domain. This input can be derived from a hydrogeological model, which is defined as "a spatial interpretation of the subsoil in hydrogeological units". Hydrogeological units on its turn are "bodies of rock that posses uniform hydrogeological properties within specified bandwiths, based on rock composition, texture or structure and are separated by the boundaries of layers and facies or the edges of erosion or faults". (*Vernes and van Doorn* (2005))

TNO is in the possession of several of such models: DGM and ReGIS II that describe the top 500 meters of the Dutch subsoil in 2D, or GeoTOP and NL3D which visualise the top 30 respectively top 50 metres in 3D. For the parametrisation of AZURE, the hydrogeological units of REGIS II were used. REGIS II, or REgional Geohydrologic Information System II, was first developed as REGIS in the nineties, and later updated to REGIS II in 1999-2003. The 2D model gives a detailed description of the permeable and impermeable layers in the subsoil. In total, 131 hydrogeological units are distinguished and for each hydrogeological unit the level of top and bottom and the hydraulic properties are recorded in a grid with size 100x100m.

The model is based on various data sources. The main source is a selection set of 14.500 boreholes with a penetration depth between 10-1000m from all over The Netherlands. Other data is obtained from experiments that estimate the hydraulic properties at a specific location, mappings of the surface level, the locations of rivers, faults, etc. With this data a grid filling model is constructed. For each borehole an interpretation was made into the 131 hydrogeological units. After that, using an objective algorithm involving Thiessen Polygon mappings, a first estimate of the extent of each unit is made. This produces a raw model that is subjected to extensive control routines. Many flaws are removed and adjustments to ensure spatial consistency are made. The estimate of the geometry is a single best estimate; no probabilities on the extent of a unit are generated.

After the geometry has been established, the hydraulic parameters (horizontal and vertical permeabilities) need to be assigned to the hydrogeological units. Each hydrogeological unit consists of one or more lithology classes (classes based on the colour, texture, grain size, or composition of the soil). At the borehole sites the permeability has been measured for those individual lithology classes. With formulae 3 and 4 the permeability of the hydrogeological units can be computed when the permeabilities of the underlying lithology classes are known:

$$k_h = \frac{\sum_{i=1}^{N} d_i k_{h,i}}{\sum_{i=1}^{N} d_i} \tag{3}$$

$$k_v = \frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} \frac{d_i}{k_{v,i}}} \tag{4}$$

where k_h is the total horizontal permeability and k_v the total vertical permeability for a hydrogeological unit. Further, d_i and $k_{h,i}$, $k_{v,i}$ are the thickness and horizontal/vertical permeability of lithology class i.

At the borehole sites the hydraulic parameters can be computed. On the other grid points, where no data is available, estimations need to be made. For REGIS II the geostatistical interpolation method Ordinary Kriging (OK) is applied. In section 3.4 an introduction to Kriging is provided. OK is the most commonly used method in geostatistics and generally performs well

when applied to normal distributed variables. For hydraulic parameters it is generally accepted that they follow a log-normal distribution. The OK is therefore applied to the log-transformed permeabilities. In addition to a mean estimate, the OK algorithm also provides confidence intervals for the interpolated variables.

The 131 different units can be experienced as complex for users of the model. Therefore, a schematisation is made, translating the model from a hydrogeological model into a geohydrological model that works on a level with only aquitards and aquifers. A top layer, tops and bottoms of aquifers and aquitards, a geohydrological basis (an overall bottom layer) and stowed depositions are distinguished. This is done separately for each of the 12 provinces. As the model area of AZURE is defined in multiple provinces the geohydrological model used for AZURE is slightly different, but the principle of stacking hydrogeological units together in an aquifer or aquitard is the same.

An aggregation process takes place to combine the units of ReGIS II into aquifers and aquitards. For the geometry the aggregation process means that multiple units are taken together and defined as one of the 17 layers (9 aquifers + 8 aquitards). The hydraulic parameters of these layers can again be computed using the permeabilities of the individual units in formulae 3 and 4.

2.3 Uncertainties of REGIS II in AZURE

Both the geometry and the hydraulic parameters in the geohydrological model will contain uncertainties. The uncertainties in the geometry are caused by many different sources. Already at the interpretation of the boreholes in REGIS II, serious errors can occur. All data on boreholes is stored in archives that were built over many years. Depending on the method that was used to make a borehole, the executor and the year, the accuracy of the data will fluctuate. A second source of errors is introduced during the interpolation. Even with the additional knowledge of sophisticated geologists adjusting the estimates of standard interpolation routines with historical knowledge, it will remain impossible to exactly define the boundaries of certain depositions. As the aggregated layers are simply a summation of the REGIS II units, the uncertainties will be inherited directly by the input of AZURE.

Similarly, uncertainties in the hydraulic parameters occur due to measurement errors, wrong meta data and interpolation with the Kriging algorithm. As the individual units of REGIS II are the building blocks for the hydraulic parameters of AZURE, these uncertainties are inherited by the AZURE parameters.

3 Data assimilation and geostatistics

Model predictions of the hydraulic heads from the model AZURE will generally not match the observations, due to parameter uncertainty and model errors. The misfit between those two can be used to obtain better estimates for the model parameters. A so-called objective function that incorporates this misfit in some norm can be defined and minimised. Well-known examples of objective functions are (weighted) least squares, maximum likelihood estimator and Bayesian maximum likelihood estimator (*Valstar* (2001)). The process of integrating data and prior knowledge by minimising the objective function is known as 'data assimilation', 'inverse modelling', 'parameter calibration' or 'history matching'.

Over the last couple of decades many data assimilation methods have been developed. In figure 3.1 a very general overview is given.

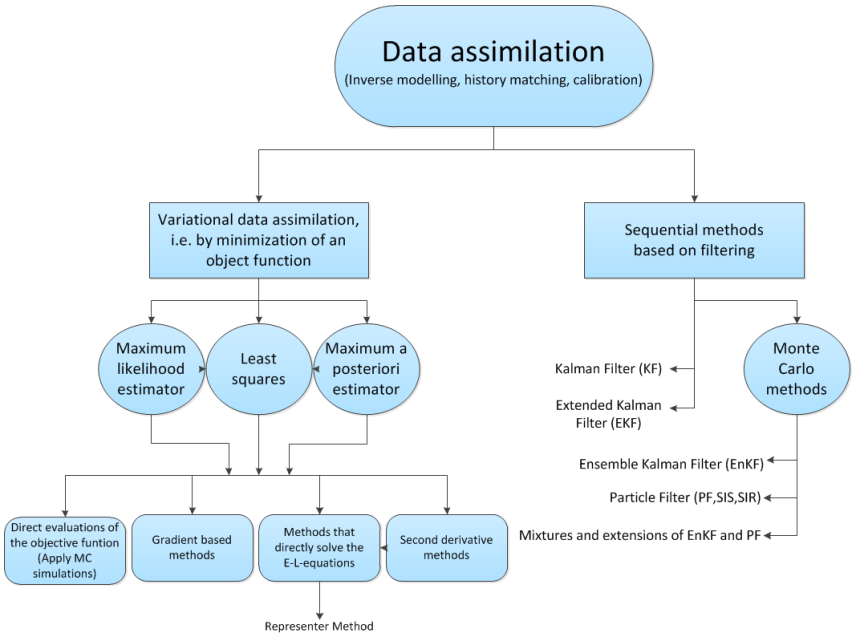


Figure 3.1: General overview of data assimilation methods

There are two important main classes; the variational methods and the sequential methods. In the variational approach, an objective function is directly minimised by for instance direct evaluations of the objective function, gradient based methods, by directly solving the Euler-Lagrange equations or second derivative methods. The other branch is composed of sequential methods. Sequential methods provide an approximation of the least squares solution by simulating the model forward in time and integrating measurements whenever they are available. An advantage of sequential methods is that they are easier to adapt for all models and in case of a continuous supply of data that the data can be assimilated on-line.

In the current calibration procedure of AZURE a variational data assimilation method, the representer method (*Valstar* (2001)), is used for the calibration of the stationary parameters (such as resistances/transmissivities/thicknesses) of the model. In Appendix A a brief description of this method is given. Characteristic for the method is the elegant parametrisation technique that it uses. It allows a reduction of the number of unknowns to the number of measurements using an optimal parametrisation. However, the implementation of the representer method is

quite cumbersome compared to a very straightforward method as an Ensemble Kalman Filter (EnKF). To conceptually test whether hydraulic head data contains enough information to be used to improve thickness estimations, it might therefore not be the most suitable method. The advantage is that there already is a running and working code for calibrating the hydraulic parameters at Deltares. However, adjustments would need to be made to include the uncertainties, and this might be just as time consuming as starting another method from scratch. Therefore we will reconsider what method to choose.

In the field of reservoir engineering, very similar problems are considered. A brief literature review of the application of data assimilation techniques in that field is given along with a discussion of the possibility for successful application here. However, before such a discussion is possible, first a refinement of the data assimilation framework should be given.

3.1 Framework

In the research goals it was formulated that we need to incorporate the uncertainty of the thickness in the data assimilation framework. The thickness of layers is not defined as a parameter for the AZURE model directly, but is included in the hydraulic parameters (horizontal transmissivity and vertical resistances). The relation between the hydraulic and geological parameters is given by:

$$c = \frac{d}{k_v} \tag{5}$$

$$T = dk_h \tag{6}$$

where c is the resistance [T], d the thickness of a layer [L], k_h and k_v the hydraulic conductivities in horizontal and vertical direction [LT^{-1}], and T the transmissivity [L^2T^{-1}]. For the performance of AZURE, it is not useful to break these hydraulic parameters down into a permeability and a thickness for the calibration process. However, for using the calibration method as feedback to the underlying geohydrological model it is.

In the current calibration routine, both c and T are allowed to vary independently of each other. However, if these would be split up into conductivities and thicknesses, an extra constraint rises. The sum of thicknesses of the various layers in a vertical column should remain constant. The 'easy' solution is to keep the thicknesses constant and let the conductivities absorb all changes and hence be responsible for all uncertainties in the hydraulic parameters. However, in some cases this might result in strange conductivities that are not consistent with the type of soil that is described for a layer, indicating that the estimated thicknesses must be wrong. Taking c and T independent in a calibration routine, implicitly uses this assumption.

In this thesis we will swap the roles and fix the k values while making the thickness the uncertain parameter. This introduces the already mentioned dependency between c and T. These dependencies can be introduced either by using penalties such that the constraint of total thickness is "more or less" satisfied, or make the thickness of one layer dependent on all others. However, since we will here consider uncertainties in only one aquitard, another solution to handle the dependencies is required. We will define a maximum allowed thickness for the aquitard of interest and allow the top and bottom to vary within these ranges. Figure 3.2 explains how adjusting the thickness of the layer of interest results in changes in the surrounding layers as well.

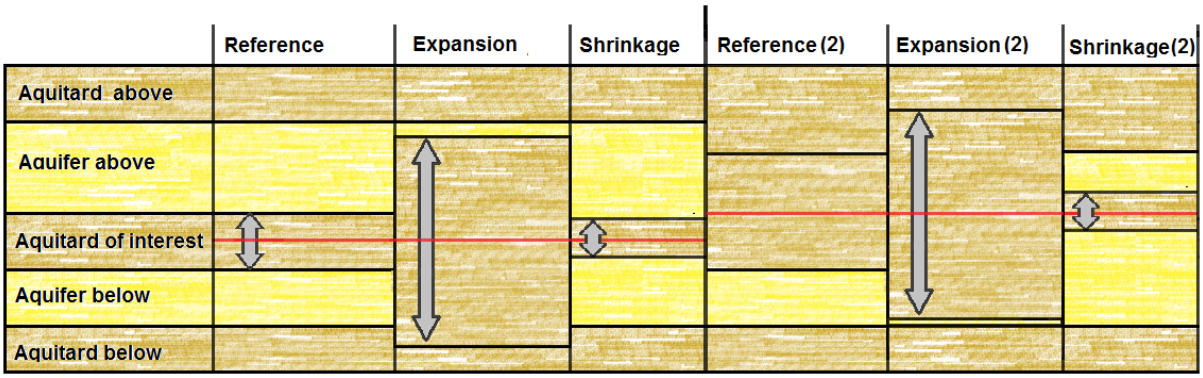


Figure 3.2: Adjustment of surrounding layers for varying thicknesses.
The red line displays the middle of a layer around which the aquitard grows/shrinks. The surrounding layers are adjusted accordingly.

Firstly, note that once the aquitard 'grows' through the aquifer and touches the aquifer on top or below, then the resistances can be summed. If it is assumed that all aquitards have permeabilities of more or less the same order, then the specific configuration of which aquitard has which thickness is negligible for the outcome of the streamflow model. Any further increments may cause the aquitard to even outgrow its maximum allowed thickness. In that case, the thickness will be truncated at the maximum allowed thickness, here defined as the thickness between the top of the aquitard above and bottom of the aquitard below. Secondly, we have to note that if in the prior model an aquifer is not present, then it can emerge if the aquitard shrinks. This may seem odd, but if this framework would be expanded to multiple aquitards and a similar approach would be used for the aquitard on top, then a correction of the layer on top could counteract this effect. Once the thickness is estimated, it will be divided/multiplied by the fixed constant k_v/k_h value (that is allowed to vary spatially) to construct resistance/transmissivity fields for the input of AZURE.

From the viewpoint of performance of AZURE, these dependencies seem like an extra constraint that shrinks the space of feasible solutions. Therefore, the performance of AZURE using the selected optimal parameter set, should remain the same or deteriorate compared to the optimal set that is obtained by calibrating with the representer method. However, in that calibration procedure, penalties are added to prohibit the hydraulic parameters to change more than a factor 10. If the main uncertainties are in the permeabilities (k-value), then this is a reasonable constraint. It is plausible to assume that a permeability is within a factor 10 difference of its original value, otherwise it would not be consistent with the facies type. Also, since the variance of the parameter will grow with magnitude, it makes sense to select a constant multiplication factor to define the boundaries of the parameter. However, if the main uncertainties are in the thickness, then this assumption is less appropriate. The error in the thicknesses depends on the magnitude very little. For very thin or very thick strata, the absolute magnitude of the errors is of the same order. More important, if a layer is not present at a certain location before calibration, then it should be allowed after. However, this is a factor ∞ and it therefore cannot be estimated in the current calibration routine. Therefore a calibration with respect to geometry might also improve the overall model performance of AZURE.

3.2 Literature review

Many different data assimilation methods exist and it is an impossible task to compile a complete overview. Therefore, in this literature review, we will limit ourselves to data assimilation methods that have been applied in reservoir engineering problems. In reservoir engineering, lots of research is done towards the estimation of permeability fields and porosity fields that are used in reservoir modelling simulations. The permeability and porosity in a cell in the discretised problem, are largely dependent on the facies type of the cell. In many applications, only two facies are distinguished; permeable sand and impermeable shale. Only small fluctuations in the permeability and porosity occur within these facies types, and they are often even ignored. The comparison to groundwater engineering is straightforward; the aquifer and aquitard play a similar role as the sand and shale. A large difference between both problems is the amount of uncertainty in the prior fields. Much more data is available in the shallow subsurface, automatically decreasing the amount of uncertainty. This is relevant to keep in mind, but is not a reason to not apply a method from the reservoir engineering methods here.

In many reservoir engineering studies, a re-parametrisation is done to reduce the number of parameters that is estimated independently. The reasons for this are mainly two-folded. First of all, not more parameters than independent data can be uniquely optimised, and secondly, lowering the amount of parameters is often computationally attractive. If a parametrisation is carefully chosen, and the number of parameters is not reduced a too large amount, then the model estimate using the parametrised problem compared to the estimate that would have been made with a full parametrisation (defining each quantity of interest in each grid block) is very similar. When searching for a suitable data assimilation method, it is therefore usually not only a method that needs to be selected, but a method in combination with a parametrisation technique.

Oliver and Chen (2011) provide a review on the (in 2011) recent progress of data assimilation techniques for reservoir engineering purposes. Many of the recently applied techniques are the combination of new parametrisations with an EnKF. In the mid-1980's and early 1990s there was a revolution in the way people thought about geologic models of the reservoir and their uncertainty. Major developments in the field of geostatistics took place, and instead of making single best estimates, geologists began to create possible realisations that had much more heterogeneity than previously thought necessary or desirable. This formed a natural demand for ensemble-based methods to be used for the data assimilation.

A fairly recent trend is to define the permeability and porosity fields by their rock type (facies fields) and assign a constant permeability and porosity to each type (*Liu and Oliver* (2005), *Armstrong et al.* (2011), *Sebacher et al.* (2013)). By using a plurigaussian truncation simulation technique, the boundaries of the geological facies can be estimated. An example of plurigaussian truncation is shown in figure 3.3. The main idea is to generate a Gaussian random field and assign, depending on the location of a Gaussian random field point in a so-called rock-type-rule mapping, the facies. In the example, the lower values (red) in the Gaussian random field are mapped to the first/green facies (F1), while the high values (blue) are mapped to the third/yellow facies (F3). As the green and yellow facies (F1 and F3) do not touch in the rock-type-rule mapping, they also do not touch in the simulated facies field. The second/orange facies (F2) is always in between. The vertical lines in the rock-type-rule mapping are the thresholds. The horizontal position of these lines determines the proportions of the simulated facies. In this example, the surface area of F1 in the rock-type-rule mapping is larger than the surfaces

area of F2 and F3. Therefore the green facies (F1) most often occurs in the facies map. In this example, only one Gaussian Random Field was used. In a plurigaussian approach, this can be extended to arbitrarily many fields. The specific formulation of the rock-type-rule mapping can be adjusted to fit the problem.

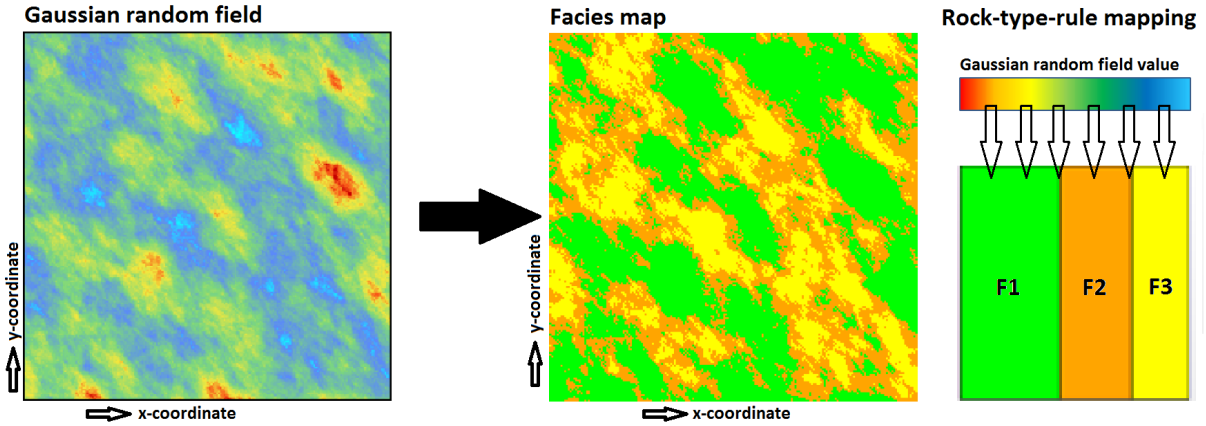


Figure 3.3: Plurigaussian truncation simulation.
The images were created using PluriDemo (*Armstrong et al.* (2011))

Plurigaussian truncation methods can be combined with automatic data assimilation methods. Namely, it allows the facies to be represented by a continuous variable, eliminating the problem of working with a categorical variable that is not differentiable (necessary for variational assimilation methods) or Gaussian (preferable for EnKF). *Liu and Oliver* (2005) combine the plurigaussian approach with the EnKF. A clever choice for the definition of the rock-type-role mapping makes this possible. The results they obtain are satisfactory, but they do assume that the proportions of occurrence of each facies are known or estimated in advance. These proportions are needed to determine the position of the thresholds in the rock-type-rule mapping. A calibration of this position (and hence the proportions) is not automatic but requires a trial-and-error procedure.

Sebacher et al. (2013) define a different rock-type-rule mapping that allows the thresholds themselves to be calibrated. Moreover, they introduce the concept of probability fields in order to give an interpretation to the Gaussian random fields. By defining a mapping from the Gaussian random fields towards a probability field, the probability on a specific lithotype at a specific location is estimated. The mean thresholds and a prior confidence interval are estimated and by integrating the parameters into the EnKF they are all estimated during the calibration. For test cases in the reservoir engineering, the results are promising. Therefore the possibilities to apply this approach to the AZURE model were investigated further. A mathematical summary of the method can be found in Appendix B.

Although the application seems very similar, the transition to our problem is not straightforward. The problem treated by *Sebacher et al.* (2013), is in 2D and defined for categorical variables. The permeability is considered constant for cells within the same facies type. This last assumption is (by approximation) valid for permeabilities in our problem setting. However, for transmissivities and resistances, it is also strongly dependent on the thickness. We are looking for an estimate in 2D of a continuous variable or in 3D of a categorical variable. This means

that we will either need to introduce a large amount of different classes (the former facies types) or start working on a full 3D grid. Both approaches will introduce much more parameters than unknown variables and are therefore unattractive from a computational point of view. For a full 3D groundwater model, the 3D categorical field might provide a reasonable solution, but because AZURE is only Quasi-3D, it would be a waste to pursuit here. The input in AZURE is not dependent on the exact configuration of thicknesses in a vertical column, but only on its proportions. As the application applies an EnKF after the parametrisation, severe memory problems would occur. It would be inefficient to make different realisations of something that is not used at all.

It would be much more natural to interpret the probability fields as proportion fields directly (since both properties can have values between 0 and 1). No thresholds would have to be defined, but only mappings from one or multiple Gaussian Random Fields to the proportion fields. However, for the generation of the Gaussian Random Fields the data is required to have a Gaussian distribution. This already fixes our choice for a mapping and leaves no room for parameters defining it. Furthermore, we should note that for *Liu and Oliver* (2005) one of the reasons to use the plurigaussian truncation simulation in combination with data assimilation, was the transformation of a categorical variable into a continuous one to make it easier to apply data assimilation methods. Here, however, we are already trying to estimate a continuous variable, releasing this benefit.

Another branch of popular parametrisation methods in combination with an EnKF uses grid deformations. *Seiler et al.* (2009) propose an elastic 3D approach for the modelling of structural uncertainty. A standard rectangular grid is adjusted to match the top and bottom of the reservoir. The main benefits of this parametrisation for the reservoir engineering, is that it allows to simulate on a fixed number of grid blocks, while the domain can shrink and expand. They apply the elastic deformation in combination with calibrating parameters on the internal grid blocks. We, however, have no desire at all to model the third dimension as all parameters (also the ones not calibrated here) are taken constant in the vertical direction. Our only interest is in the location of the top and bottom layer and grid deformations are only useful if they are applied in a scenario with varying parameters in the vertical. The only case where that would happen in our defined framework, is if multiple layers are considered at once. This, however, requires that we would have to start directly with a more complicated scenario then intentionally aimed at. We will therefore leave an application of this method for future research.

3.3 Motivation for EnKF and EnKF-GMM

From the literature review we can conclude that none of the researched parametrisation techniques from the reservoir engineering were found to be applicable to the problem here. The main cause is the specific model that we are using for the calibration. The techniques from the reservoir engineering are specified to either 2 or 3 dimensions. Applying a method in 2D is not sufficient, while using a full third dimension does not allow the methods to reach their full potential. If a full 3D groundwater model and a full 3D geohydrological model would have been used, then a comparison between the fields would be better on its place. Here a method that just focuses on the thickness itself will suit better.

Therefore a general EnKF, without any additional parametrisation techniques, will be used. With the main purpose of this thesis in mind; to test whether we can perform a calibration with respect to geometry, it is best to select a method as simple as possible. The main restriction of

this method, is that it requires Gaussianity of the posterior distribution. With the assumption that thicknesses are log-normal distributed, a straightforward transformation can be performed to satisfy this assumption. Around the edges of the aquitard, the situation will be slightly more complicated. The uncertainty whether the aquitard is present or absent, may locally introduce a bimodal distribution for the thickness. Whether the application of the EnKF will be successful in this case, will depend on the information content of the hydraulic head data. If it is large enough to directly penalise those members having a wrong extent of the aquitard, then it might be reasonable to assume that the posterior distribution is log-normal. If not, we will have to pursuit other methods that can deal with the bimodality.

After the application of the EnKF, we can conclude that the bimodality causes the EnKF to perform weakly. Another method should be selected. Typically mixtures with particle filters are applied to relax the strict assumptions of a Gaussian distribution. Similarly to the EnKF, the Particle Filter(PF) uses Monte Carlo simulations to represent the prior and posterior probability density functions by a set of random samples. Instead of updating the ensemble members forward in time, the PF updates the weights of the random samples and computes the estimates based on these samples and weights. The main issue with PFs is that often a much larger ensemble is needed to reduce the negative effects of filter degeneration. Filter degeneration means that all but one particle will have negligible weight. Increasing the ensemble size may require a lot of computational power.

Many combinations of both ensemble filters have been reported in literature. However, not all of the mixtures are intended to relax the assumption of Gaussianity of the prior and posterior distribution. An example of a method that does so, is the Particle Ensemble Kalman Filter (PKF, *Hoteit et al.* (2012)). In this algorithm, the distributions are considered to be Gaussian Mixture Probability Distribution Functions (PDFs). For a mixture consisting of N Gaussian PDFs, N parallel EnKFs are applied, with after each update a computation of the posterior weights. A large drawback of this method is the required computational effort that is needed since N can become quite large. In the situation here, it seems that $N = 2$, as we have only 2 facies types. However, each cell separately (with some correlation structure) is sampled from a bimodal distribution; making it necessary to use many more Gaussian distributions, up to $N = 2 *$ number of cells. Each separate Gaussian distribution needs an ensemble to be approximated, requiring too much effort.

A computationally less requiring method is proposed by *Dovera and Della Rossa* (2011). Their algorithm also involves multiple EnKFs to be applied parallel, but instead of resampling the distributions it resamples the members. Using an inverse Cholesky and regular Cholesky transformation the members are transferred from one Gaussian distribution to another. This allows us to define from which distribution the thickness in the cells of each individual member are originating, without having to extensively sample all possible combinations of facies types. This method was applied after the regular EnKF. Details on the theoretical formulation and the results are given in sections 6 and 7.

3.4 Geostatistics

The thickness of an aquitard can be considered a random function that models the joint uncertainty of interdependent thicknesses distributed in space. As input to the EnKF and EnKF-GMM we need an ensemble of realisations of this random function, conditioned on the data (boreholes) that is available. Luckily there is an abundance of geostatistical software available to generate such realisations. For instance SGeMS (Stanford Geostatistical Modeling Software) which is an open-source computer package for solving problems involving spatially related variables (*Remy et al.* (2009)). In this thesis SGeMS will be used, however, to be able to apply one of the simulation algorithms included in the package, it is useful to have a very general introduction into the different methods and their underlying concepts.

Traditionally, there are estimation and simulation algorithms. Estimations are used to obtain probability distributions whereas simulation algorithms are meant to construct equi-probable realisations from these probability distribution functions. As estimations describe averages, they are usually much more smooth than simulations. Both estimation and simulation algorithms will be discussed here.

3.4.1 Estimation algorithms

Estimation problems can generally be defined as follows:

> Given n measurements $y_1, y_2, .., y_n$ at locations with spatial coordinates $\mathbf{x}_1, \mathbf{x}_2, ..\mathbf{x}_n$, estimate the value of y at location $\mathbf{x}_0$.

Note that $y_1, y_2, ..y_n$ is a random process. A semi-variogram is a structural model that indicates how the constitutive random variables of the random processs relate to each other and to the data. It describes the degree of spatial dependence between any two variables based on the distance between them and is therefore called a 2-point statistic. Usually semi-variograms are estimated from the data.

The semi-variogram, denoted by $\gamma(.)$, is defined as:

$$\gamma(h) = \frac{1}{2}E[(y(\mathbf{x}) - y(\mathbf{x}'))^2], \qquad h = \|\mathbf{x} - \mathbf{x}'\|$$

A solution to the estimation problem can be constructed by assuming that y is a linear combination of $y_1, y_2, .., y_n$. In that case, $\hat{y}$, the estimation of y, can be written as:

$$\hat{y}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i y_i(\mathbf{x}_i) \tag{7}$$

where λ_i determines the weight of each observation $y_i(\mathbf{x}_i)$.

Many linear interpolation algorithms exist, but preferably we select an estimator that meets the following specifications:

Minimisation of variance:

$$\min_{\lambda_1,..,\lambda_n} \text{Var}\left[\hat{y}(\mathbf{x}_0) - y(\mathbf{x}_0)\right] \tag{8}$$

Unbiasedness:

$$E\left[\hat{y}(\mathbf{x}_0) - y(\mathbf{x}_0)\right] = 0 \tag{9}$$

If we introduce the assumption that $E[y(\mathbf{x}_0) - y_i(\mathbf{x}_i)] = 0$ for all locations $\mathbf{x}_i$, then a sufficient condition to live up to the unbiasedness condition is $\sum_{i=1}^{n} \lambda_i = 1$. Namely, from $E[y(\mathbf{x}_0) - y_i(\mathbf{x}_i)] = 0 \; \forall \mathbf{x}_i$, it follows that all points y_i have the same mean m. If it holds that $E\left[\hat{y}(\mathbf{x}_0)\right]$ is also equal to m, then the condition of unbiasedness is satisfied. To show this, we note that:

$$E\left[\hat{y}(\mathbf{x}_0)\right] = E\left[\sum_{i=1}^{n} \lambda_i y_i(\mathbf{x}_i)\right] = \sum_{i=1}^{n} \lambda_i E\left[y_i(\mathbf{x}_i)\right] = \sum_{i=1}^{n} \lambda_i m = m$$

where the last equality holds if $\sum_{i=1}^{n} \lambda_i = 1$.

Kriging provides a Best Linear Unbiased Estimator to the problem. The system can be minimised using Lagrange multipliers. The derivation is not included here, but the minimisation yields the following Kriging system to be solved (*Kitandis* (1997)):

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{10}$$

where

$$\mathbf{A} = \begin{pmatrix} -\gamma(\|\mathbf{x}_1 - \mathbf{x}_1\|) & -\gamma(\|\mathbf{x}_1 - \mathbf{x}_2\|) & \dots & -\gamma(\|\mathbf{x}_1 - \mathbf{x}_n\|) & 1 \\ -\gamma(\|\mathbf{x}_2 - \mathbf{x}_1\|) & -\gamma(\|\mathbf{x}_2 - \mathbf{x}_2\|) & \dots & -\gamma(\|\mathbf{x}_2 - \mathbf{x}_n\|) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\gamma(\|\mathbf{x}_n - \mathbf{x}_1\|) & -\gamma(\|\mathbf{x}_n - \mathbf{x}_2\|) & \dots & -\gamma(\|\mathbf{x}_n - \mathbf{x}_n\|) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix},$$

$$\mathbf{x} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ v \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} -\gamma(\|\mathbf{x}_1 - \mathbf{x}_0\|) \\ -\gamma(\|\mathbf{x}_2 - \mathbf{x}_0\|) \\ \vdots \\ -\gamma(\|\mathbf{x}_n - \mathbf{x}_0\|) \\ 1 \end{bmatrix}$$

The here described Kriging algorithm is Ordinary Kriging (OK). Other forms are Simple Kriging (SK), Kriging with a Trend(KT) and Indicator Kriging(IK). SK also assumes the mean to have a constant value, but in addition it assumes that this value is known. It is therefore less general and not used here. KT allows the mean to vary with a trend over the domain and can in specific applications yield better results. IK is Kriging for indicator variables.

3.4.2 Simulation algorithms

Kriging provides best estimates and variances, but for the EnKF we require realisations of the random function. Sequential Gaussian Simulation (SGS) and Sequential Indicator Simulation (SIS) are Kriging-based simulation algorithms. Both algorithms make use of the same basic algorithm, but have a different domain of application. SGS is suited for continuous variables (like thickness) whereas SIS can be used for the simulation of categorical variables (present or absent). The general outline of both algorithms is as follows:

1. Define a random path through all the grid nodes.
2. Sequentially visit all nodes on this path:
   - Estimate the local conditional probability distribution function conditioned on the data and previously simulated nodes by applying Kriging.
   - Draw a realisation from the resulting distribution.

- Add the realisation to the data set

3. Repeat step 2 until all nodes are visited.

The realisations of SIS are categorical fields displaying which category is present in each grid cell. The fields that are generated with SGS are Gaussian random fields conditioned to the data. The only difference between the algorithms of SGS and SIS is in the estimation of the local conditional distribution function; SGS uses SK whereas SIS applies IK. However, this has additional consequences for the distribution of the data. For SGS the data must follow a Gaussian distribution and if this is not the case a normal score transformation must be performed. In case of bimodal PDFs this can create unsatisfactory results. For SIS, the data must be categorical and otherwise transformed. If the variable is not Gaussian or categorical, an alternative is to apply Direct Sequential Simulation (DSS) instead. DSS does not use a transformation, but has its own issues. The variants included in the SGeMS packages either fail to correctly reproduce the histogram or the covariance structure (*Soares* (2001)).

4 Method: EnKF

For the first series of experiments, an EnKF was used. In this section a general introduction and specifics regarding the implementation and configuration of the filter is given.

4.1 Introduction

The Ensemble Kalman Filter (EnKF) was initially introduced by Evensen in 1994 (*Evensen* (1994)) and improved by Burgers et al. in 1998 (*Burgers et al.* (1998)) and has since then been applied widely. Over the years, many researchers have given their own touch to the EnKF, improving the EnKF for a specific problem setting. Many examples can be found, some of the nature of better estimations and others for better computational efficiency (*Evensen* (2003)). However, all methods are still characterised by their simple concept that directly relates back to the original Kalman Filter(KF). Here an introduction to the EnKF is given that generally follows the description of *Aanonsen et al.* (2009). It starts with a description of the regular KF and Extended Kalman Filter (EKF, a KF for nonlinear models), and then provides the differences with the EnKF.

4.2 Theoretical formulation

All Kalman type filters use a probabilistic framework to give an estimate of the whole system state sequentially by propagating information only forward in time. It is assumed that the system state $\mathbf{x}$ at time $t+1$ only depends on the state at time t, but not on $(0, 1, .., t-1)$, and that the observations $\mathbf{y}$ at time t depend only on the system state $\mathbf{x}$ at time t (Markov assumption). This means that the true system state can be expressed by equation 11:

$$\mathbf{x}_t = M_t(\mathbf{x}_{t-1}) + \mathbf{w}_{t-1} \tag{11}$$

where M_t is the model describing the dynamics of the system (AZURE) and the subscript t the time index. The model M_t is time dependent sine we assume here that the time-dependent boundary conditions are part of the model description. The vector $\mathbf{w}_{t-1}$ represents model uncertainties and is a realisation of a random variable with mean 0 and covariance matrix $E[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T] = Q_t$.

At the measurement locations, parts of the truth are indirectly observed, e.g. in general not all components of the state vector are observed and all observations will contain errors. The set of observations of the model that is observed at various locations at time t can be written as:

$$\mathbf{y}_t = H_t\mathbf{x}_t + \mathbf{v}_t \tag{12}$$

where $\mathbf{x}_t$ is the true system state at time t and H_t a linear observation operator that relates the observations to the system state. We observe $\mathbf{y}_t$, a summation of the true state and some noise $\mathbf{v}_t$ to express the uncertainty in the observations. Here $\mathbf{v}_t$ is a realisation from a random variable with known probability distribution function, such that $E[\mathbf{v}_t] = 0$ and $E[\mathbf{v}_t\mathbf{v}_t^T] = R_t$.

In the Kalman Filter the true state at time t is approximated. In two steps, the state of the system at the next time step is estimated by propagating the model forward in time and updating it by applying a statistical analysis of the observations.

The forecast step is given by:

$$\mathbf{x}_t^f = M_t(\mathbf{x}_{t-1}^a) \tag{13}$$

where the superscripts f and a denote the outcomes of the forward and analysis steps.

In 1960 Kalman suggested to update the forecasted state vector with observations by using the following formula:

$$\mathbf{x}_t^a = \mathbf{x}_t^f + K_t \left(\mathbf{y}_t - H_t \mathbf{x}_t^f \right) \tag{14}$$

He derived the gain K_t in such a way that it minimises the trace of the covariance matrix of the estimation error, e.g., it minimises the trace of $P_t^a = E[\tilde{\mathbf{x}}_t^a (\tilde{\mathbf{x}}_t^a)^T]$, where $\tilde{\mathbf{x}}_t^a = \mathbf{x}_t^a - \mathbf{x}_t$. Several different derivations of the Kalman gain K_t have been published. A derivation that respects Kalman's original concept using orthogonal projections on the vector space of random variables is given in Appendix C. The Kalman gain can be expressed as:

$$K_t = P_t^f H_t^T (H_t P_t^f H_t^T + R_t)^{-1} \tag{15}$$

where P_t^f is the forecast error covariance matrix. The state forecast is assumed to be unbiased, such that the forecast error covariance matrix is given by: $P_t^f = E[\tilde{\mathbf{x}}_t^f (\tilde{\mathbf{x}}_t^f)^T]$, where $\tilde{\mathbf{x}}_t^f = \mathbf{x}_t^f - \mathbf{x}_t$.

In the original Kalman Filter the forecast covariance matrix is updated at each time step using former estimates. The following equations can be used for the updates (see Appendix C for the derivation):

$$P_t^f = M_t P_{t-1}^a M_t^T + Q_{t-1} \tag{16}$$

$$P_t^a = (I - K_t H_t) P_t^f (I - K_t H_t)^T + K_t R_t K_t^T \tag{17}$$

For the computation of the covariance matrices, linearity of the model M_t is required. In the regular Kalman Filter this is an assumption. For nonlinear models it is an option to linearise the model. In that case, the Extended Kalman Filter (EKF) is obtained. The EKF has been frequently used but has lost popularity since the EnKF was introduced. In case of highly non-linear models the shortcomings of the EKF become severe, and moreover, for large-scale systems there are large problems with storage and the updating of the covariance matrices P_t^f and P_t^a. In the EnKF these problems are omitted by approximating the error covariance matrix by a sample error covariance matrix.

The EnKF is a Monte-Carlo variant of the KF/EKF. A large amount of ensemble members is used to approximate the true statistical moments. In the forward step, the model equations are applied to each of the n_e ensemble members $(\mathbf{x}_t^a)_i$, $i = 1, .., n_e$ and an approximation of the true state is obtained by taking the mean $\bar{\mathbf{x}}_t^a$:

$$\bar{\mathbf{x}}_t^a = \frac{1}{n_e} \sum_{i=1}^{n_e} (\mathbf{x}_t^a)_i \tag{18}$$

A sample covariance matrix can then be computed to approximate the true covariance matrix:

$$P_t^f \approx \frac{1}{n_e - 1} \sum_{i=1}^{n_e} \left((\mathbf{x}_t^f)_i - \bar{\mathbf{x}}_t^f \right) \left((\mathbf{x}_t^f)_i - \bar{\mathbf{x}}_t^f \right)^T \tag{19}$$

where $\bar{\mathbf{x}}_t^f$ is defined as

$$\bar{\mathbf{x}}_t^f = \frac{1}{n_e} \sum_{i=1}^{n_e} (\mathbf{x}_t^f)_i \tag{20}$$

Next, the ensemble members can be updated by applying the analysis equation to each ensemble member. An updated estimate for the system state is found by taking the mean of the analysed ensemble member state vectors.

4.3 Configuration

The EnKF as described above is in its most general form. The updates of the EnKF are mostly steered by the sample covariances that are computed in the filter, but other factors can also play a role. In this section a few issues will be discussed to control and improve the performance of the EnKF.

4.3.1 Errors in model and measurements

In the EnKF two errors are present. A model error $\mathbf{w}_t$ and a measurement error $\mathbf{v}_t$. In real applications, the standard deviation of these errors should be determined according to the estimated accuracy of the model and measuring equipment. The application here, however, is purely synthetic and a twin experiment. In a twin experiment, a true scenario will be constructed along with the other members. The outcomes of the forward model of this true scenario can be used to generate synthetic measurements. As a result of using a twin experiment, the model will be an exact representation of the truth and the measured hydraulic heads in the true scenario are also free of errors.

However, the errors also influence the EnKF algorithm. The measurement error fulfils a role in the analysis step of the EnKF. Without perturbed observations, the variance of the analysed ensemble is too low(*Burgers et al.* (1998)). Therefore, the standard deviation of the measurement error should be taken nonzero. A large measurement error represents large uncertainties in the observed data and will decrease the magnitude of filter updates. The value of the standard deviation of the error can therefore be used in this synthetic test case to steer the magnitude of updates, but both errors should be kept within reasonable ranges. In the field, an error in the other of centimetres is standard, and therefore we will use a measurement error with a standard deviation between 2 and 5cm.

The model error also influences the magnitude of the updates but has the reverse effect. Larger model errors imply a lower model accuracy and will increase the magnitude of updates. In none of the experiments a model error will be considered.

4.3.2 Errors in initial conditions

The same initial conditions are prescribed to all members at the start of the EnKF. These conditions are uncertain and will contain errors. Therefore, for all members a warming-up period of one year is included to remove the effect of initial errors.

4.3.3 Confirming EnKF

In the EnKF, both stationary thicknesses and dynamic hydraulic heads are updated. If data is available at time t, then the model continues the forward simulation to time $t+1$ with updated variables after assimilating the data at time t. However, if the initial guess is poor, large errors in the updated dynamic states can occur. The EnKF updates are linear whereas the underlying flow equations are nonlinear. The updated dynamic variables may be inconsistent with the static ones. Moreover, as time continues, these errors will continue to build up as there is no mechanism to correct the inconsistencies. The hydraulic heads may attain physically unreasonable variables, causing the EnKF to diverge.

In 2005, *Wen and Chen* (2005) described a 'confirming' EnKF to avoid errors in the dynamic state variables. In the regular EnKF all updated state variables are used for the next forward simulation. In case of the confirming EnKF, only the static variables are used. The new forward simulation is restarted at time $t-1$ instead of at time t. The forward simulation from $t-1$ to t is called the confirming step. In this step the dynamic components are 'confirmed' such that the forward simulation from time t to $t+1$ can be performed using updated static variables and confirmed dynamic variables. Compared to the standard EnKF, the performance of the confirmed EnKF was shown to be much better (*Wen and Chen* (2005)). Moreover, the static and dynamic variables are consistent in the confirming filter. A drawback is that the computational effort of the confirming EnKF can be up to two times the amount of effort of the standard EnKF, since all forward model simulations are done twice.

In the applications here, we will use the confirming approach. As the majority of the computational effort is spent on the initialisation of the AZURE model, the specific number of days that is simulated forward is of rather limited influence for the total computational effort. Therefore the confirming step is chosen maximal; for each update only the static updates are used and the dynamic variables are simulated again starting from time $t = t_0$.

4.3.4 Logarithmic transformation

The EnKF will be applied to the logarithmic transformed thicknesses. The reason for this is two-folded. First of all, the thickness is assumed to be lognormal. By taking the logarithm we therefore transform it to a normal distribution which is aligned with the assumptions of the EnKF. Secondly, it ensures that no negative thicknesses can occur after updates.

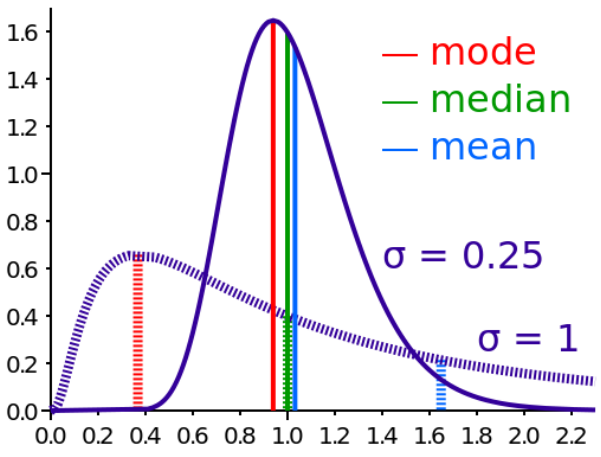


Figure 4.1: Two log-normal distributions with different skewness.
From: http://en.wikipedia.org/wiki/Log-normal_distribution

For the interpretation of the results it is necessary to go back to a normal scale again. For a Gaussian distribution the mode, mean and median are all the same. In case of a lognormal distribution they differ (see figure 4.1). The following relations will be used to transform the logarithmic distribution back to the regular scale:

For $X \sim lnN(\mu, \sigma)$ we have:

$$\mathrm{E}[X] = e^{\mu+\frac{1}{2}\sigma^2} \tag{21}$$

$$\mathrm{Var}[X] = \left(e^{\sigma^2}-1\right)(E[X])^2 \tag{22}$$

$$\mathrm{Mode}[X] = e^{\mu-\sigma^2} \tag{23}$$

We will regard to the mode as the most important estimator because it is the most likely estimate.

4.4 Implementation

To describe the physical state of the subsurface, we formulate a state vector in which the log thickness of each cell and the hydraulic head data at measurement locations is gathered. This yields the following state vector:

$$\mathbf{x}_t = \begin{bmatrix} \log(\text{thickness})_1 \\ \vdots \\ \log(\text{thickness})_{n_x * n_y} \\ \text{head observation}_1 \text{ at time } t \\ \vdots \\ \text{head observation}_{n_m} \text{ at time } t \end{bmatrix}$$

of size $(n_x * n_y + n_m)$x1 where n_x is the number of cells in x-direction, n_y the number of cells in y-direction and n_m the number of measurements.

As we restart the EnKF it suffices here to only include the hydraulic heads at the measurement locations. We note that since AZURE counts 9 aquifers, this is roughly a factor 10 difference in the state vector length compared to an EnKF without confirming step. Hence using the confirming EnKF significantly reduces the required memory.

The measurement operator H_t can be written as a composition of two block matrices:

$$H_t = \begin{bmatrix} 0_{n_x * n_y, n_m} \\ I_{n_m} \end{bmatrix} \tag{24}$$

where $0_{n_x * n_y, n_m}$ is a matrix of size $(n_x * n_y)$xn_m containing only zeros and I_{n_m} the identity matrix of size n_mxn_m.

The model M_t is the AZURE model along with prescribed initial and boundary conditions. Since we restart the model from the first time step after each update, the input of the model does not require any hydraulic heads. We will provide these hydraulic heads as input to avoid having to introduce another variable, but we emphasise that these heads are not used as input to the model.

Figure 4.2 presents a flow chart of the EnKF with confirmation step.

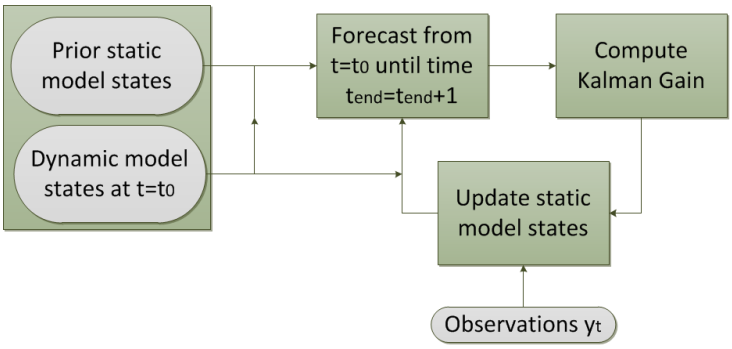


Figure 4.2: Flowchart of EnKF

The implementation will be most efficient if the equations are in matrix form. The Ensemble Kalman Filter algorithm can then be summarised in a few lines: (n_e is the number of ensemble members)

1. Construct an initial ensemble $X_0^a = [(\mathbf{x}_0^a)_1, \ldots, (\mathbf{x}_0^a)_{n_e})] \in \mathbb{R}^{(n_x n_y + n_m) \mathrm{x} n_e}$ that represents the prior statistics.

2. Iterate for $t = 1 : n_t$;

   (a) Do a forward simulation from time t_0 (confirming step) until time t for each ensemble member, i.e compute $(\mathbf{x}_t^f)_i = M_t((\mathbf{x}_{t-1}^a)_i)$ for $i = 1, .., n_e$

   (b) Collect the ensemble members in a matrix, $X_t^f = [(\mathbf{x}_t^f)_1, \ldots, (\mathbf{x}_t^f)_{n_e}] \in \mathbb{R}^{(n_x n_y + n_m) \mathrm{x} n_e}$

   (c) Create a matrix Y_t, holding n_e differently perturbed copies of the observation vector: $Y_t = [\mathbf{y}_t + (\mathbf{v}_t)_1, \ldots, \mathbf{y}_t + (\mathbf{v}_t)_{n_e}] \in \mathbb{R}^{n_m \mathrm{x} n_e}$

   (d) Define $\hat{Y}_t$ as the matrix holding the ensemble of predicted measurements: $\hat{Y}_t = [H_t(\mathbf{x}_t^f)_1, \ldots, H_t(\mathbf{x}_t^f)_{n_e}] \in \mathbb{R}^{n_m \mathrm{x} n_e}$

   (e) Compute an auxiliary covariance matrix A_t^f, such that $P_t^f = A_t^f (A_t^f)^T$: $A_t^f = \sqrt{\frac{1}{n_e - 1}} \left(X_t^f - \frac{1}{n_e} X_t^f \mathbf{1}_{(n_e, n_e)} \right)$

   (f) Compute the Kalman Gain: $K_t = A_t^f ((A_t^f)^T H_t^T) \left((H_t A_t^f)((A_t^f)^T H_t^T) + R_t \right)^{-1}$ The inverse was implemented in Matlab using the backslash function.

   (g) Update all ensemble members: $X_t^a = X_t^f + K_t(Y_t - \hat{Y}_t)$

5 Application of EnKF to a synthetic model

The EnKF was applied to a synthetic model, derived from the real AZURE and REGIS II model. Using a twin experiment, the performance of the EnKF was tested in two scenarios. In the next subsection the synthetic model and these two scenarios are described. After that, a sensitivity analysis is performed and the measurement locations are discussed. Finally, the results are presented and the section concludes with a general evaluation of the application of the EnKF to the scenarios here.

5.1 Synthetic model

The synthetic model is derived from the true model to ensure that it is geologically plausible. A region of 20x30 km^2 around the intersection of the provinces of Utrecht, Flevoland and Gelderland was selected from the domain of the AZURE model. There, at a depth of approximately 30-40 metres, a thick aquitard is situated in parts of the domain. The aquitard is situated between the fourth and fifth aquifer of AZURE, when counting from the top. In figure 5.1 the model thickness of the aquitard is shown. The thicknesses in this figure can be considered as a single best estimate of the truth, that is based on the prior geological data. Here we will not consider this model as the truth, but rather use it as a starting point for the construction of an initial ensemble and the truth. Figure 5.1 is defined on a grid with 200x300 cells with a cell size of 100x100m. For the simulations in the AZURE model an upscaling is done to 80x120 cells of 250x250m.

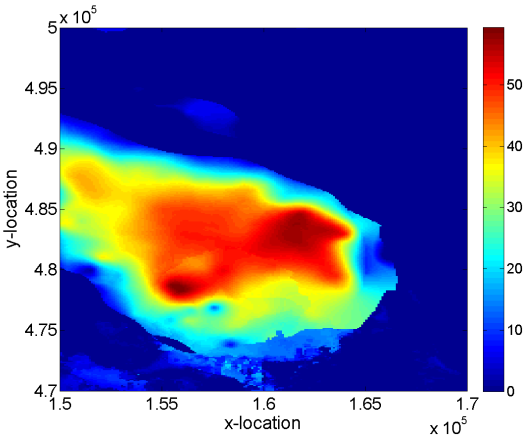


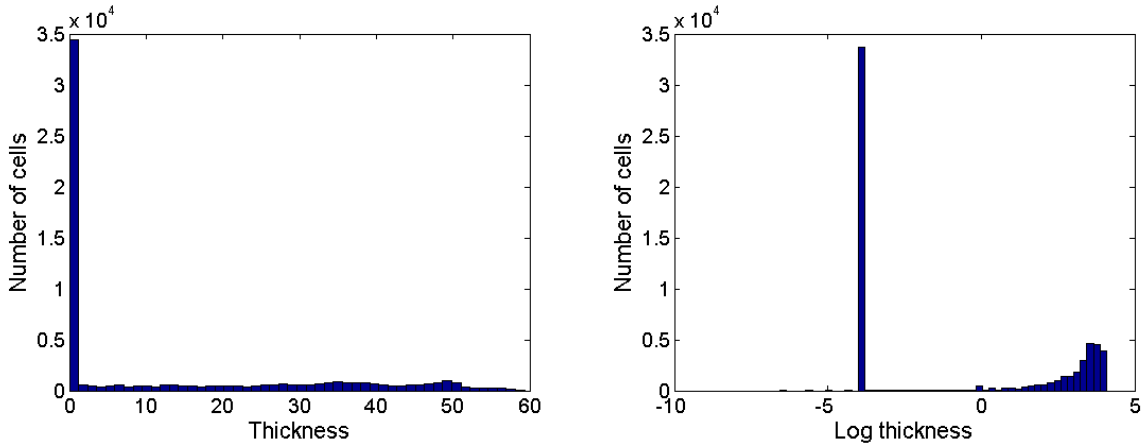
Figure 5.1: Thickness of model layer [m] according to geological model.

5.1.1 Construction of a prior ensemble

From the model in figure 5.1, a synthetic set of boreholes can be constructed by (randomly) selecting the thickness at several locations. For the geological model REGIS II, on average around 200 boreholes of varying quality are available in a region with surface area equal to the domain in figure 5.1. Here, however, we can vary the amount of boreholes and their locations to construct ensembles with a varying prior uncertainty. The obtained sets can be used as input for geostatistical methods to generate multiple equi-probable realisations.

As discussed in the literature review, the traditional simulation methods are SGS, SIS and DSS. There is a general consensus that the transmissivities, resistances, and permeability parameters follow a lognormal distribution. Since the product of two lognormal distributed random variables is again lognormal, it is reasonable to assume here that the thickness also is a lognormal distributed random variable. However, the assumption only holds in cells where the aquitard is present. This leads to a bimodal distribution of thicknesses over the entire domain. Figure 5.2a shows a histogram of the model thicknesses in the cell centres of figure 5.1.

For the application in the EnKF, we need the logarithmic transformation of the model thicknesses. In cells where the aquitard is absent, this transformation of the thickness leads to a



(a) Thickness [m] (b) Log-thickness

Figure 5.2: Histogram of thickness and log-thickness of model layer according to geological model.

value of $-\infty$. The introduction of a truncation constant c is necessary for practical purposes. Preferably this constant is chosen in such a way that the effects of the introduced truncation on the groundwater flow model are negligible. This means that the resulting hydraulic head differences should be less than 1cm. Meanwhile, the truncation constant should not be chosen too small, in order to prevent that the groundwater flow model will be insensitive to any changes in the thickness. In the AZURE model, a minimum resistance of 1 day is assigned to each cell. Therefore, it would be best to choose the truncation value as k_v for each cell. In this case no additional errors will be introduced and the constant is on the threshold of the sensitive region. However, for reasons provided in the discussion at the end of this section, the constant c was chosen at -4 on log-scale ($\approx$ 1cm on the regular scale). In figure 5.2b a histogram of the log-transformed model thicknesses is shown. The application of the truncation constant results in a spike at -4.

The truncation constant allows us to define a logarithmic transformed probability distribution function, but it will remain to be a bimodal distribution. In the application of SGS to a bimodal distribution it is necessary to perform a normal score transformation in advance. The transformation will join variables that are far apart on the normal scale and merge the two modes into a Gaussian distribution. Afterwards, the obtained realisations are transformed back to the regular scale to adjust the Gaussian histogram to the original (bimodal) histogram again. As a result, large jumps in the thicknesses of neighbouring cells can occur.

A better approach is to first simulate the edges of the aquitard with SIS, and after that condition the thicknesses of cells where the aquitard is present on the boreholes with SGS. In that case, the data points that are used for the SGS are log-normal distributed. Figure 5.3 shows a realisation that was generated with this approach. An ensemble of such realisations will yield enough variety in the thicknesses, but the rough extent of the aquitard will be the same for all members. Near the edges, the aquitard of each member will contain small holes in the aquitard, giving rise to short circuits in the flow patterns. The groundwater flow model may be insensitive to the exact locations of these holes. These speckled patterns are caused by an implicit maximum entropy assumption on the multi-point-statistics in the SIS and SGS algorithms (*Remy et al.* (2009)).

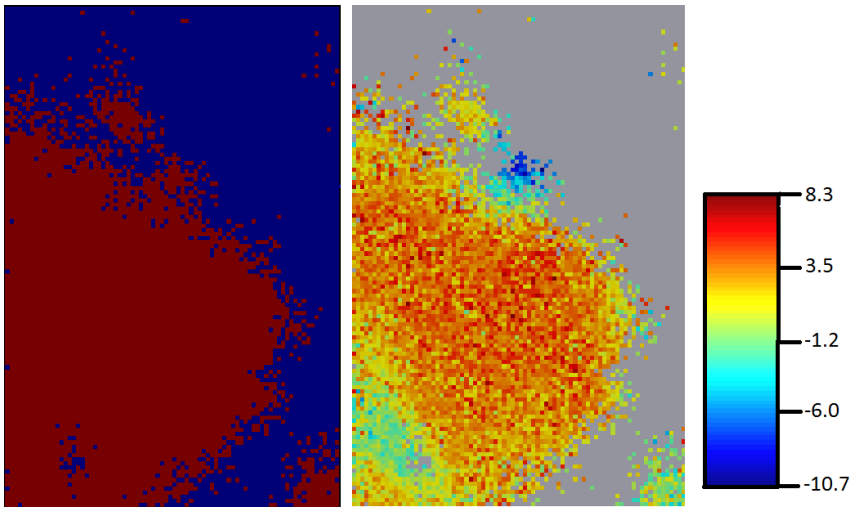


Figure 5.3: Realisation of SIS and SGS
Left: Indicator red=present, blue=absent. Right:Thickness on logarithmic scale

Indicator Kriging (IK) and Ordinary Kriging (OK) are estimation algorithms and will yield much smoother estimates. Ideally, we would use these estimators and find a way to bring some diversity into the mean estimates. Therefore, the following approach was used to generate the ensemble members:

1. For each member randomly pick n data points from the model and apply IK.

2. Determine a percentage of aquitard presence and select an area of this size with highest IK values.

3. Apply OK within the selected region.

The main drawback of this approach is that the 'true' set of boreholes is not used (here the set that is used to generate the reference scenario, in practice it is the set that is taken from the field). A separate technique must be applied to condition on the true measurements. Here, this is not done; in case of true applications, a better look at how to generate an ensemble is required anyway. For synthetic test purposes the ensemble members are geologically plausible and therefore the generated ensembles will suffice for the applications in this thesis.

5.1.2 Scenarios

Two different scenarios were constructed by varying the amount of boreholes and allowing a wider or narrower range of percentages of aquitard presence:

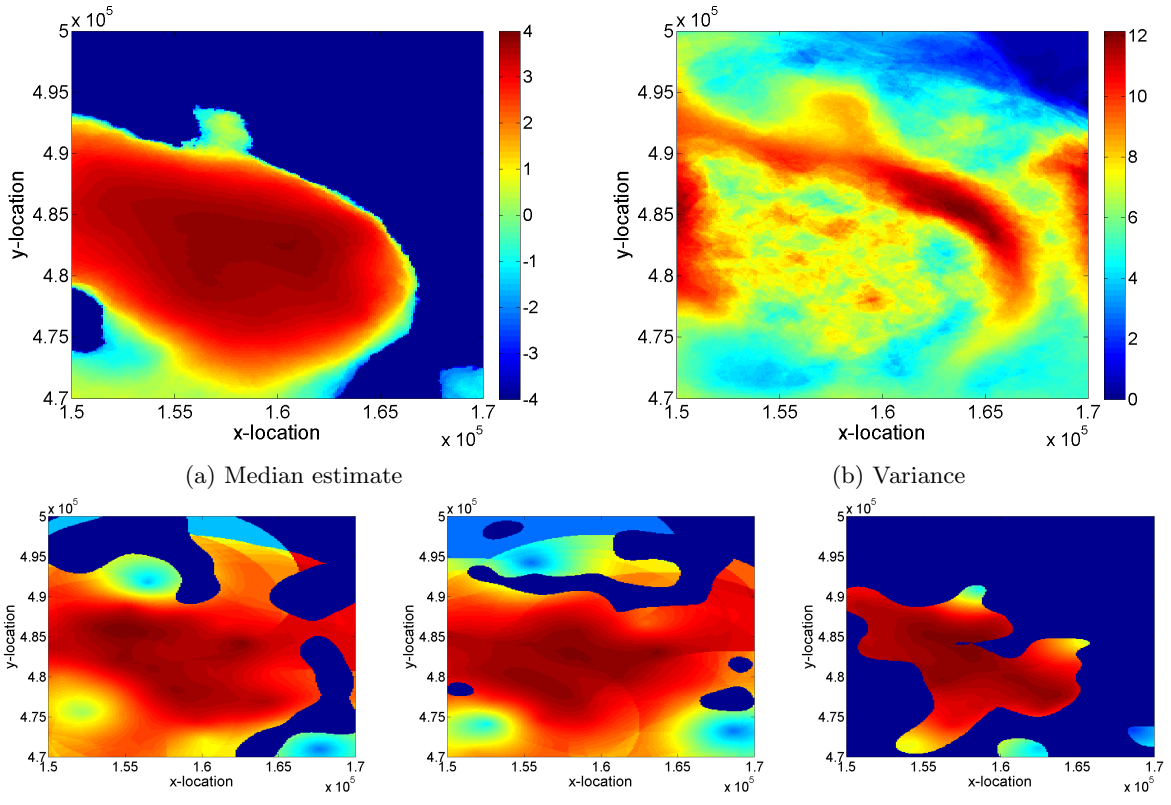
- Scenario A: A rather uncertain situation. The number of sampled synthetic boreholes was set to a total of 50, whereas the percentage of aquitard randomly fluctuates between 10 and 90 percent among the members.

- Scenario B: The edges of the aquitard are known exactly; only the thicknesses need to be estimated. All ensemble members and the true scenario have the same contour. For the OK 120 fictive borehole locations are used.

The variogram was fitted as an omnidirectional variogram on randomly drawn points of the model data. For the IK a spherical variogram with sill 0.33 and range 17000m was taken and for the OK a spherical variogram with sill 17 and range 17000m.

The quality of the prior distribution and the initial ensemble, e.g the representativeness of the true state by the initial ensemble, is a requirement for the successful application of the EnKF. The true state should be in the span of the initial ensemble. Otherwise it cannot be approximated by the filter. To make sure that this is the case, the percentage of aquitard presence for the reference case will be fixed at the mean percentage of aquitard presence among all ensemble members. This leads to 50% for the reference in scenario A and 47.5% in scenario B. The number of boreholes is chosen in accordance with the number of boreholes used in both scenarios. We note that this means that there are two different reference situations.

Figure 5.4 shows (all on a logarithmic scale) the prior ensemble median and variance and the first three members of ensemble A. We have used the ensemble median rather than the mean, since it reduces the effect that the truncation constant c has. Figure 5.5 provides a similar overview of ensemble B. Although it looks as if the variance of ensemble B is really low except for a few places in the domain, we should note that it is all on a logarithmic scale. The colour scale of the individual members is the same as for the ensemble median.



(a) Median estimate

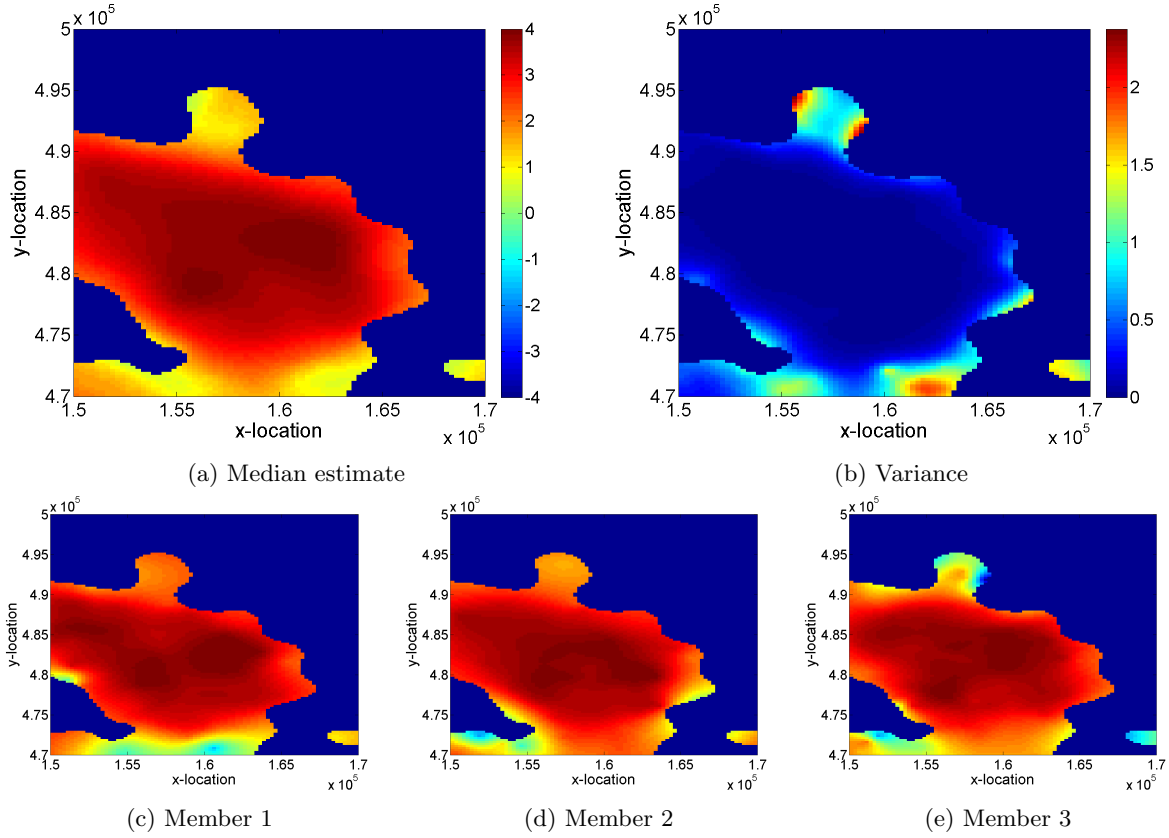
(b) Variance

(c) Member 1

(d) Member 2

(e) Member 3

Figure 5.4: Prior ensemble A on logarithmic scale.



(a) Median estimate

(b) Variance

(c) Member 1

(d) Member 2

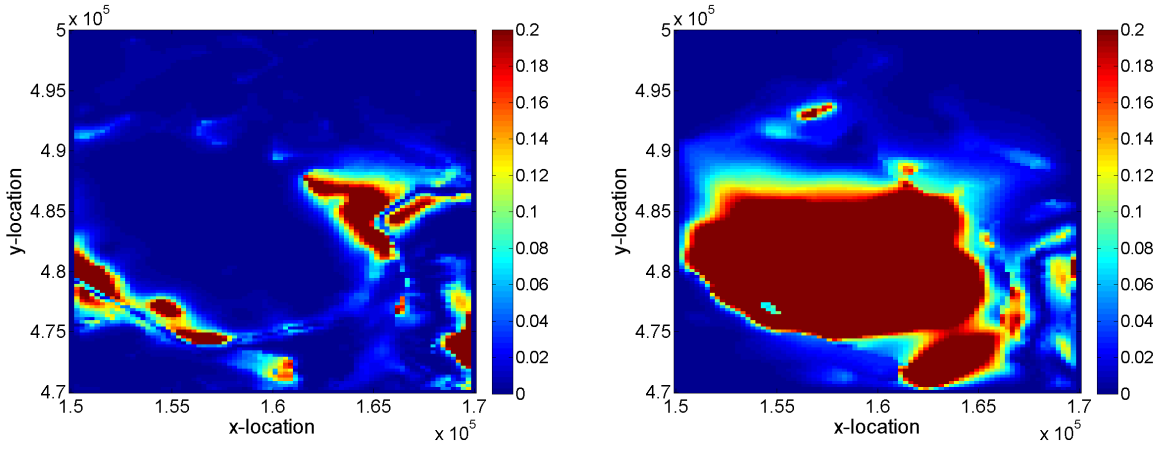
(e) Member 3

Figure 5.5: Prior ensemble B on logarithmic scale.

5.1.3 Sensitivity analysis

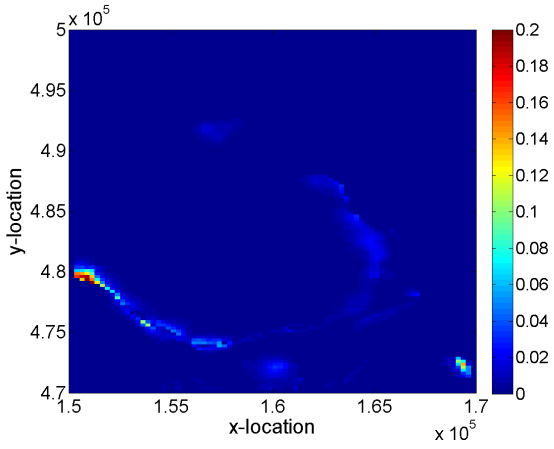
A sensitivity analysis was performed to get an indication how the variances of thicknesses influence the variances of hydraulic heads. First of all, both ensemble A and B are simulated forward for one year and the resulting variances in hydraulic heads are compared. After that, the influence of holes in the aquitard is studied.

Figure 5.6 shows the results of the hydraulic head variances of the ensembles. All subfigures are cut-off at a value of 0.2m^2. As expected, ensemble A shows much more variances in the hydraulic heads than ensemble B. To avoid updating only noise with the EnKF, we need at least a variance of approximately 0.01m^2. At the colour scale as printed, this level is the smallest that can be distinguished. In the fourth layer of ensemble B, a small variance only occurs near the aquitard's edges. There the thicknesses, and hence the resistances, are smallest and therefore variations in the thickness have a larger influence than elsewhere. A similar situation holds for the variations in transmissivities and therefore large hydraulic head variances can also be expected at locations where the surrounding aquifers are thin. Figure 5.7 shows the model thicknesses of the surrounding aquifers. Especially the fifth aquifer is locally rather thin.

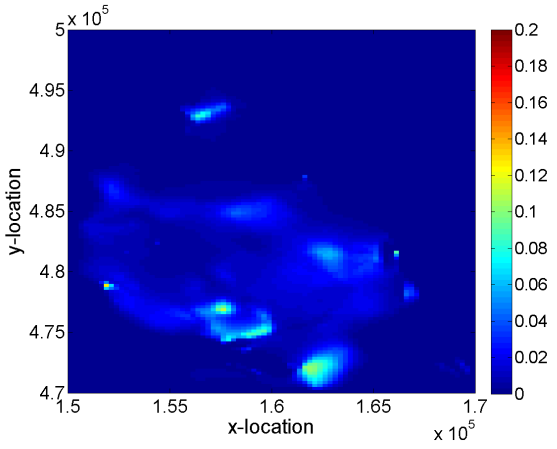


(a) Ensemble A - layer 4

(b) Ensemble A - layer 5

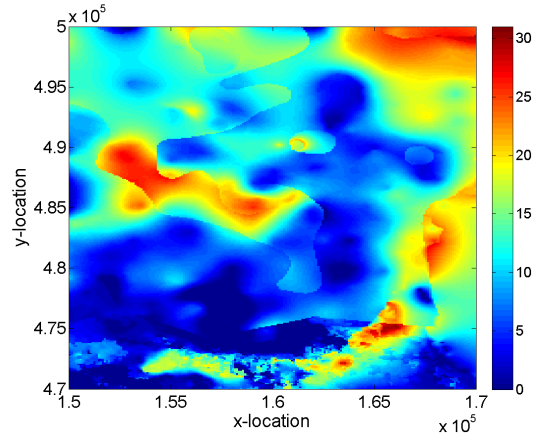


(c) Ensemble B - layer 4

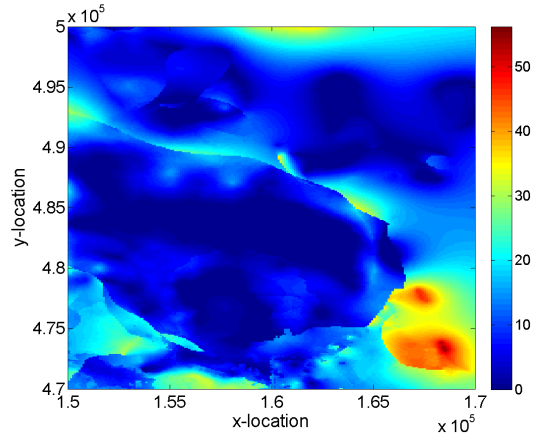


(d) Ensemble B - layer 5

Figure 5.6: Hydraulic head variances of prior ensembles A and B [m^2]



(a) Aquifer 4 (above)



(b) Aquifer 5 (below)

Figure 5.7: Thickness of surrounding aquifers [m]

Figure 5.8 shows the difference of the hydraulic head variances after one year between ensemble B and ensemble B without adjusting the thickness of the surrounding aquifers. A difference larger than 0 implies that the variance with the required dependencies is larger. Although not very significant, we can conclude that the hydraulic head variances have increased by including the dependencies, especially in regions where the aquifer is thin.

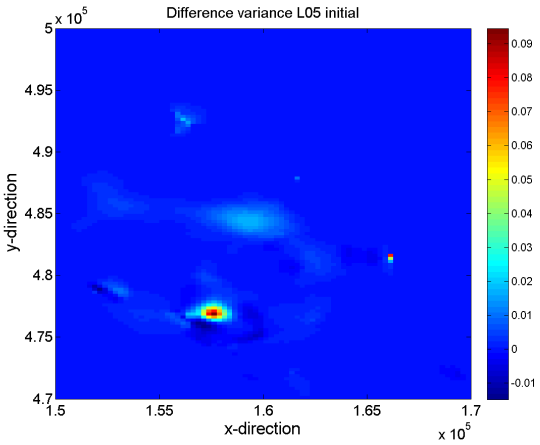
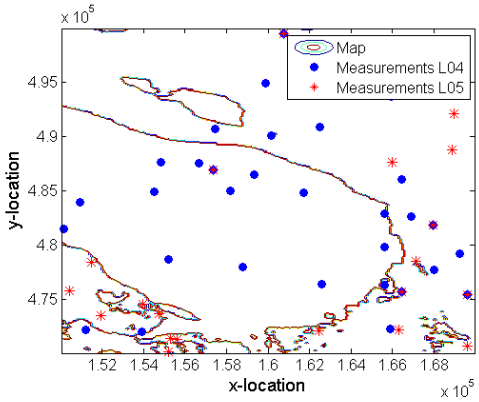


Figure 5.8: Differences in hydraulic head variance (with dependencies-without dependencies) for prior ensemble B [m^2]

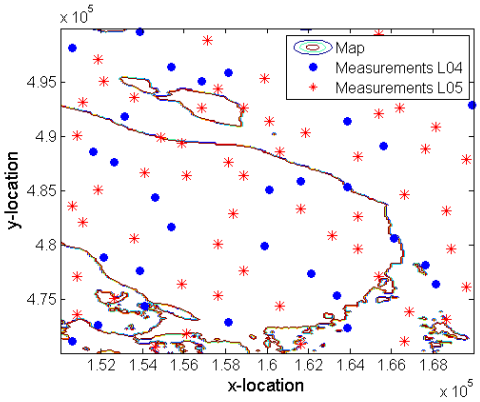
Finally, we will discuss the influence of holes in the aquitard on the solution. Earlier we claimed that the combination of SIS and SGS to generate the prior ensemble would not be suitable to apply. It was assumed that the realisations would give little variation in the output of the groundwater flow model, making it unfit to serve as input to the EnKF. For Ensemble A this is definitely true, since the state vector is scaled up before application of the model. Therefore the effective variance of the input is much smaller than what is observed from the prior ensemble on the refined grid. On the coarse grid in ensemble B the hypothesis can be verified/rejected by means of simulations. In Appendix D the results of these simulations are presented. Overall, it can be concluded that the hypothesis is incorrect. Large variances occur in multiple cells. However, the holes may destroy the correlation structure. The EnKF will only perform well if the hydraulic heads in neighbouring cells are correlated. Therefore the research goals will be easier to achieve if we assume that no holes can occur in the aquitard.

5.1.4 Measurements

Hydraulic head data from multiple observation locations is available in the model area. Here, however, we will create our own data by taking the outcome of simulations in the reference scenario. In all nine layers hydraulic heads are simulated and therefore in each of those layers measurements could be made. Of course we have the freedom to select any location we desire, but for an application in a real scenario it is useful to see how the model would have reacted in case the original measurement locations were used. Figure 5.9a shows the true measurement locations. Only measurements in the aquifers that are in direct contact with the aquitard were selected. The contour of the aquitard is plotted on the background as reference. For ensemble B measurement set 2 was used. This set is shown in figure 5.9b. It contains more locations than the original set. The measurement locations were chosen rather random. We note that the most valuable measurements are located there where the initial hydraulic head variances are large. For both experiments the hydraulic heads are assimilated each 30 days over a period of approximately 2 years.



(a) Location set 1: True measurement locations in layer 4 and 5



(b) Location set 2: Randomly selected

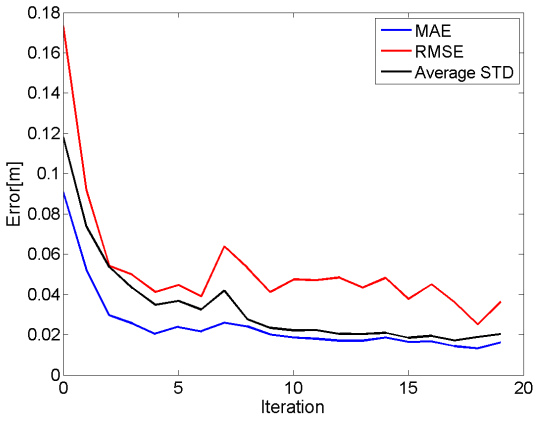
Figure 5.9: Measurement sets 1 and 2

5.2 Results

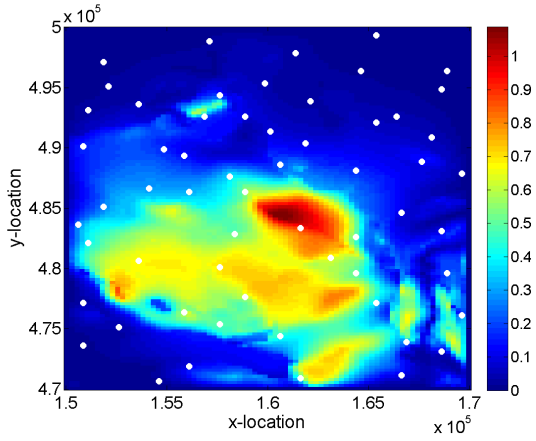
The application of the EnKF was tested in scenarios A and B. In this subsection the results of the experiments are presented. For both scenarios we will discuss the EnKF's influence on the model performance (better predictive properties for hydraulic heads) and the EnKF's ability to correctly assess the uncertainty of the thickness. Although a better predictive nature of the model is not a research goal on itself, it is an important indicator for the performance of the EnKF.

5.2.1 Experiment A

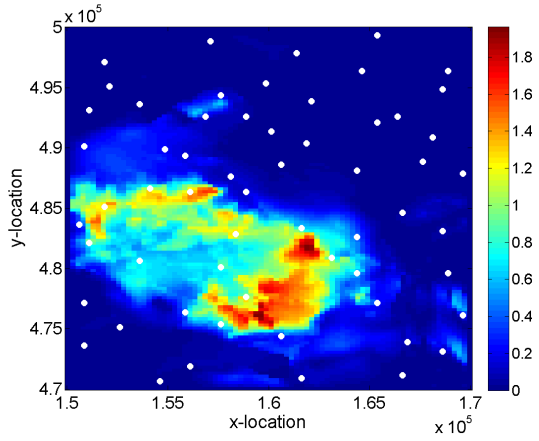
Experiment A was performed using 160 members and a measurement error with a standard deviation of 0.05m. Figure 5.10 shows results related to the model performance.



(a) Average head errors at measurement locations [m]



(b) MAE of head errors initially at 2002-01-01 [m]



(c) MAE of head errors after updates at 2002-01-01 [m]

Figure 5.10: Hydraulic head errors experiment A

In figure 5.10a, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the average hydraulic head at the measurement locations are shown. That is, at each measurement location the average hydraulic head was computed and of these 56 errors, the MAE and RMSE were computed. In Appendix E a brief overview of these error measures is provided. On the x-axis the time steps are shown. These time steps are the iterations of the EKF and with the

configuration here, $t = 0$ is 2002-01-01 and each time step resembles 30 days. Hence at $t = 20$ we have assimilated $1.5 - 2$ years of data.

Figure 5.10a illustrates that the hydraulic head error at the measurement locations decreases over time. This indicates that the performance of the groundwater flow model at the measurement locations improves as a result of updating the aquitard thickness. As a reference, also the average Standard Deviation (STD) at the measurement locations of the members is shown. Preferably, we would see that the STD at each individual measurement location is larger than the error in both MAE and RMSE. In that case the truth lies within the range of one standard deviation from our most likely estimate. The distance between the RMSE and MAE indicates that there is some variance among the errors at measurement locations. Figures 5.10b and 5.10c show the MAE of the hydraulic head in each cell. In both figures the hydraulic head errors at 2002-01-01 are shown. In figure 5.10b the prior thicknesses for the aquitard are used in the simulation, while for figure 5.10c the updated thicknesses are used. The measurement locations are indicated with white dots. From the updated figures it follows that while the hydraulic head errors at the measurement locations significantly decrease, large errors remain at the other locations.

The convergence of the hydraulic heads at the measurement locations indicates that the EnKF is functioning properly. The remaining question is now whether the research goals can be achieved. Figure 5.11 shows the updated ensemble median and figure 5.12 the reference case on a logarithmic scale. Some improvement of the global structure can be observed although it is also quite obvious that large errors remain.

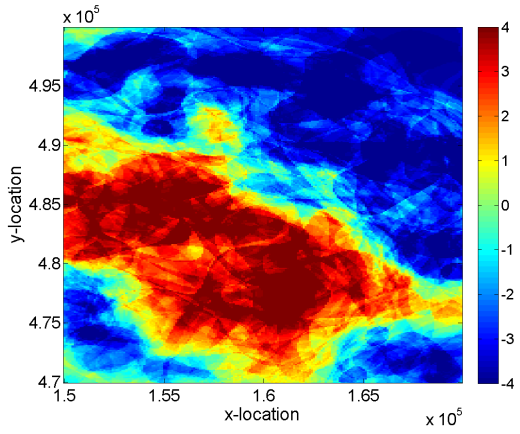


Figure 5.11: Updated ensemble median experiment A (logarithmic scale)

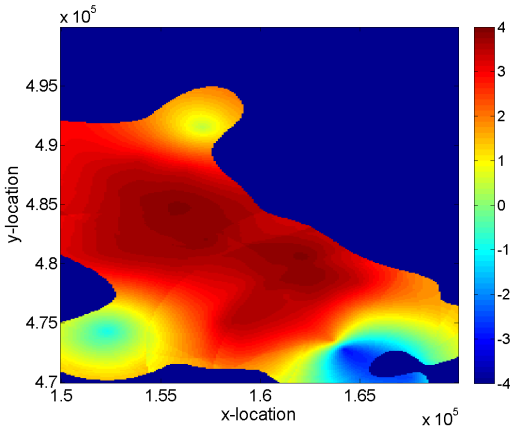
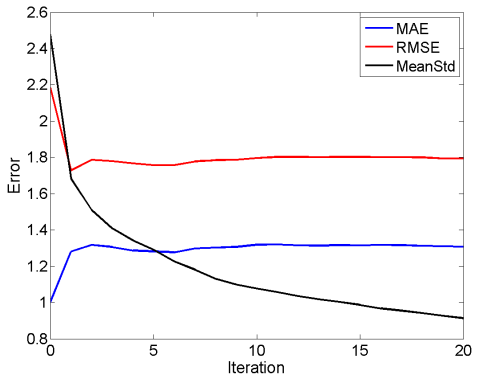


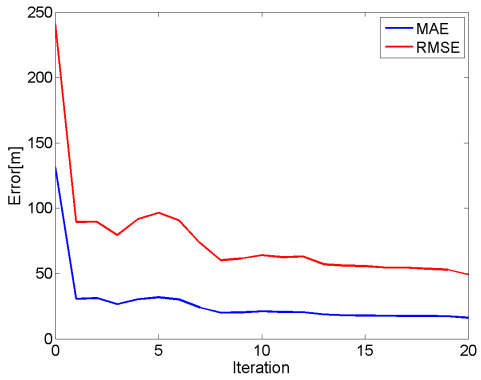
Figure 5.12: Reference scenario experiment A (logarithmic scale)

Figure 5.13 shows the average thickness error at each iteration. Figure 5.13a displays the average STD and MAE and RMSE of the estimated thickness on a logarithmic scale. The graph illustrates that the variances decrease, while the error-levels remain constant after the first update. Eventually the distribution on the logaritmic scale has to be transformed back to a distribution on normal scale. Figures 5.13b-5.13d show the average error computed using respectively the mean, median and most likely estimate of the thickness. Initially the variance is extremely large, resulting in bad estimations for both the mean and most likely estimate. The median, which is not depending on the variance shows the smallest initial error. During the iterations, the

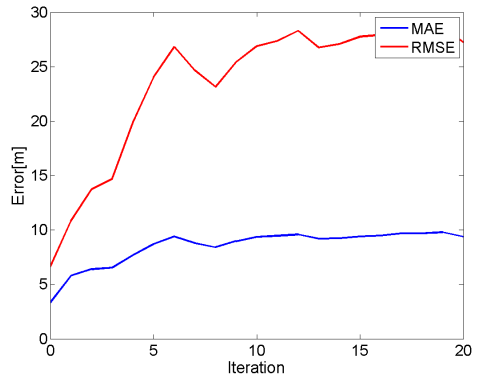
variance shrinks and hence the differences become smaller. This leads to a comparable MAE of the median and mode after all updates. Unfortunately the variance remains too large to obtain a complete (physically reasonable) posterior distribution. As a result, the mean estimate is an overestimation in many cells, resulting in an error in the order of tens of metres. The large average errors indicate that there are cells where the error estimate of the mean is over hundreds of meters. This is only possible if one, or in this case multiple members, adopt physically unrealistic values in some cells. It can be concluded that a valuable quantitative estimate of the thickness and its uncertainty is out of reach here.



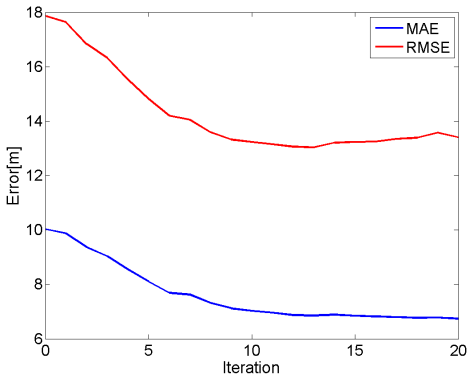
(a) Average thickness error on logarithmic scale.



(b) Average thickness error; mean estimate[m].



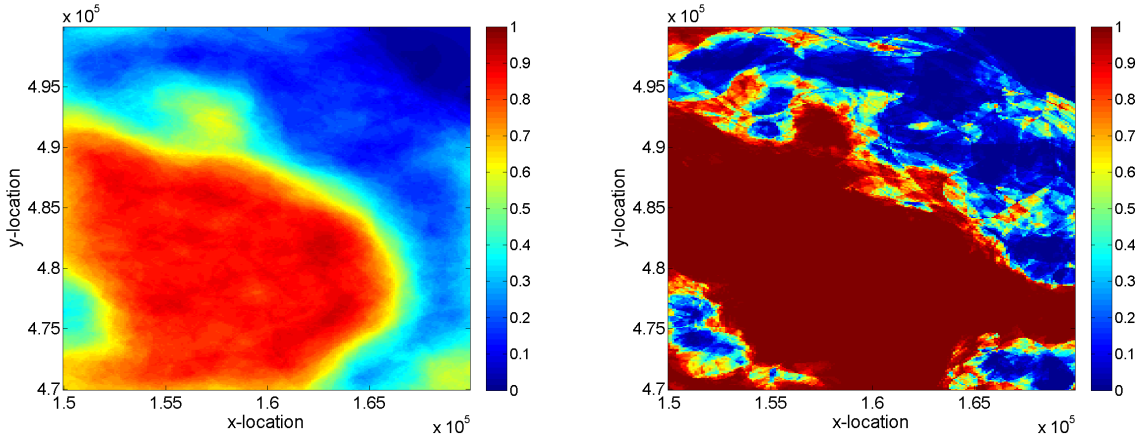
(c) Average thickness error; median estimate[m].



(d) Average thickness error; most likely estimate [m].

Figure 5.13: Thickness errors experiment A

The other research goal is to find probabilities on the extent of the aquitard. A problem that occurs immediately is how to define absence of the aquitard. In the prior ensemble we truncated all thicknesses smaller than $\exp(-4)$. Therefore, the most straightforward definition would be to again use this truncation constant to define absence of the aquitard if thickness in the updated ensemble is smaller. However, during the updates of the EnKF, the thickness distribution per cell was merged into a normal shaped distribution. In the case that the aquitard is absent, the mean log-estimate will be around -4, leading to an aquitard presence of 0.5 if the same definition would be used for the truncation constant. Therefore a different truncation constant should be used. By trial-and-error the truncation constant after updates is defined as $-2.3(\approx 5\text{cm})$. Figure 5.14 shows the percentage of members that shows presence of the aquitard initially and after all updates. Compared to the contour of the reference situation in figure 5.12 the aquitard presence is overestimated.



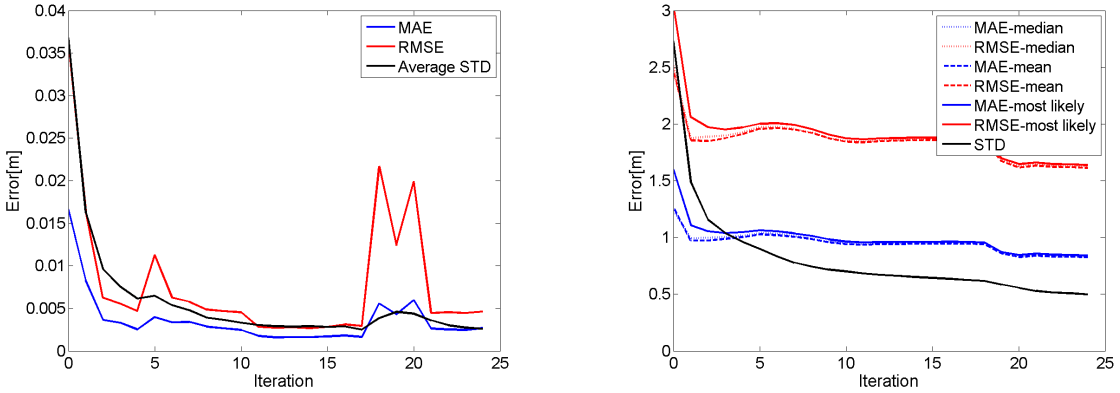
(a) Prior probability (b) Posterior probability

Figure 5.14: Probabilities on the extent of the aquitard for experiment A

5.2.2 Experiment B

The second experiment that was performed describes a situation in which the boundaries of the aquitard are known exactly; only the thicknesses need to be calibrated. Furthermore, the grid cells were scaled up to 250x250m to allow a comparison later on with the ensembles in section 7. A risk of having an ensemble with fixed contour is that it will result in small variances for the hydraulic heads; making it impossible for the EnKF to make significant updates. Therefore we decrease the measurement error to having a standard deviation of 0.02m, to allow slightly larger updates. Again 160 members and no model errors are considered.

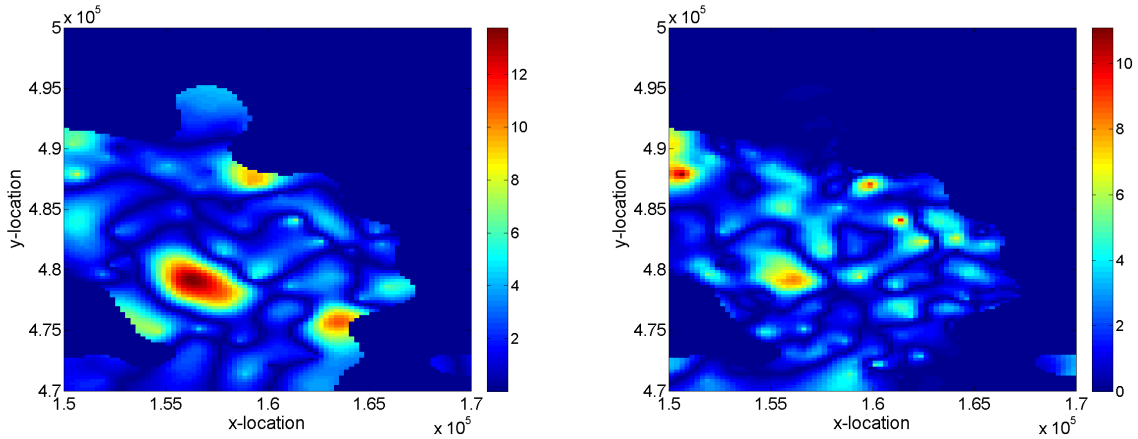
Figure 5.15 shows the results. Figure 5.15a shows the hydraulic head errors at the measurement locations of set 2. While in experiment A we had to be aware that only small updates could occur, it goes naturally here. Decreasing the standard deviation of the measurement error does not result in too large updates. Figure 5.15b shows the thickness error. The mean, median and most likely estimate are very close to each other. All error estimates are larger than the average STD, indicating that the EnKF converges to the wrong solution.



(a) Hydraulic head error [m] at measurement locations (b) Thickness error [m]

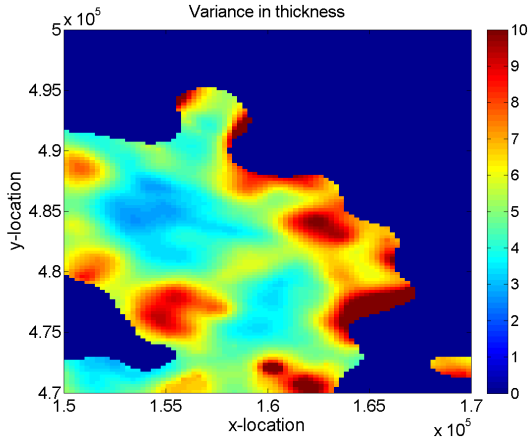
Figure 5.15: Results experiment B

To analyse this, we observe the distribution of errors and standard deviations over the entire domain initially and after all updates in figure 5.16. After the updates, the largest errors are located near the boundaries of the domain. There, the influence of the thicknesses is rather limited as a result of the fixed hydraulic heads at the boundaries. In case of a true application, the boundaries would be placed further away, damping the effect. This leads to the first conclusion that the estimate has improved more than figure 5.15b suggests. Secondly, we note that it is clear that at several places in the domain, the true state lies not within a range of one standard deviation from the most likely estimate in the posterior probability distribution; multiple spots where MAE>>STD occur. Most of these locations are there where no measurement locations are near. It can be concluded that filter divergence occurs.

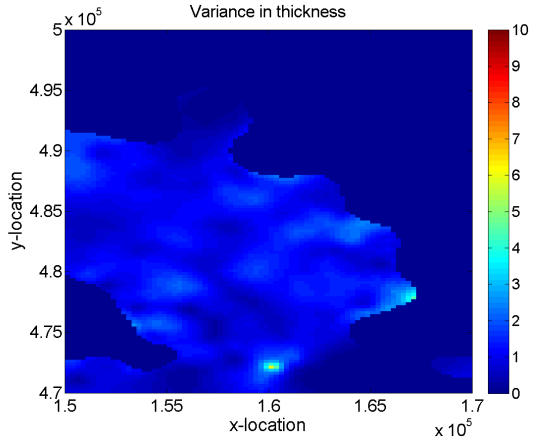


(a) MAE prior to EnKF[m].

(b) MAE after all updates[m].



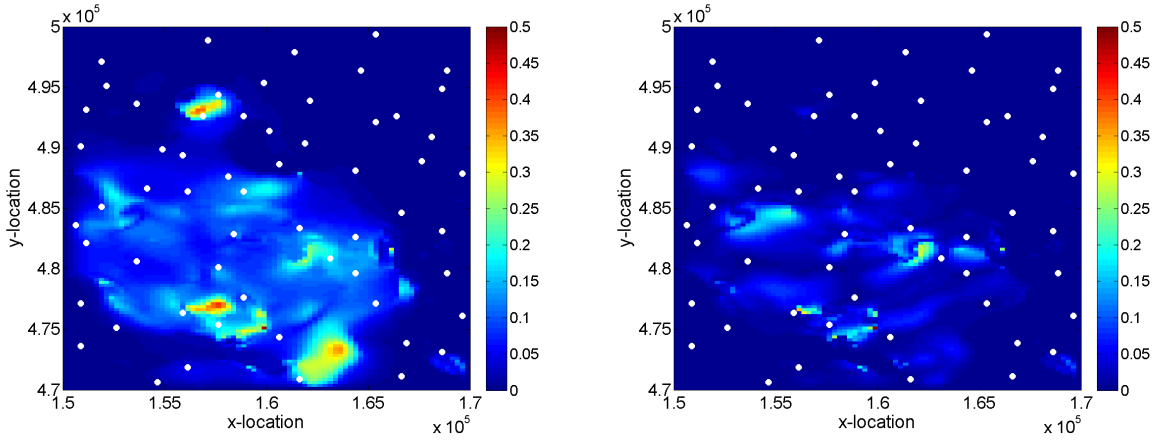
(c) STD prior to EnKF[m].



(d) STD after all updates[m].

Figure 5.16: MAE of most likely thickness and STD for experiment B

Finally, we consider the hydraulic head error in the entire domain instead of just at the measurement locations. Earlier in experiment A, little improvement was found at locations other than the measurement locations. Here, however, we have used a different measurement set, with more and better located observation points. Figure 5.17 shows the decrease in hydraulic head errors over the entire domain at 2002-01-01. A better performance throughout the entire domain is observed.



(a) Prior to EnKF

(b) After all updates

Figure 5.17: MAE of hydraulic heads at 2002-01-01 in aquifer 5 in experiment B.

5.3 Analysis and discussion

Two experiments were performed, using different prior distributions and measurement sets. At the start of both experiments, a significant error of the hydraulic heads at the measurement locations was present. During the EnKF updates, the hydraulic head errors diminished, until a level in the order of centimetres was reached. This indicates that the EnKF is able to improve the performance of the AZURE model. However, the goal was to obtain a posterior probability distribution for the thickness of the aquitard. This goal has not been achieved due to the following reasons:

1. The magnitude of the first update in the first experiment is extremely large; introducing geologically unrealistic values in a significant amount of the members. The EnKF is not able to correct this at later time steps.

2. Covariance is a measure to estimate the linear relationship between variables. The relation between the log-thickness of the aquitard and hydraulic heads is likely to be very nonlinear.

3. The transformation back from a distribution on log-normal scale to the regular thickness scale is only valid in case the posterior distribution is normal. Hence it cannot be applied to the prior distribution, making it difficult to assess the quality of the prior ensemble and any improvements.

4. The role of the truncation constant is indistinct. There is no straightforward way to define a truncation constant for after the updates. Also, the influence of the chosen constant plays a much larger role than desirable.

5. Filter divergence occurs.

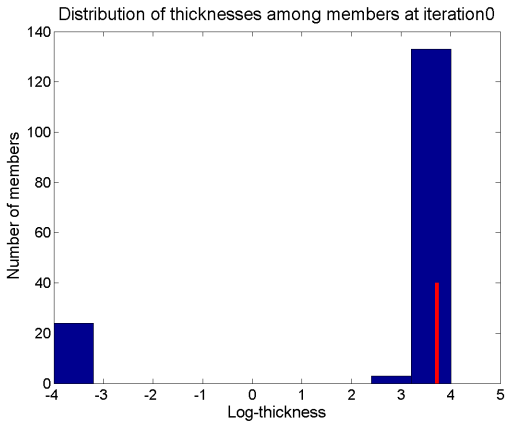
In this subsection we will discuss the causes of these problems and evaluate whether the results would contribute to our research goals if the here defined true scenario would be reality.

5.3.1 Magnitude of the update

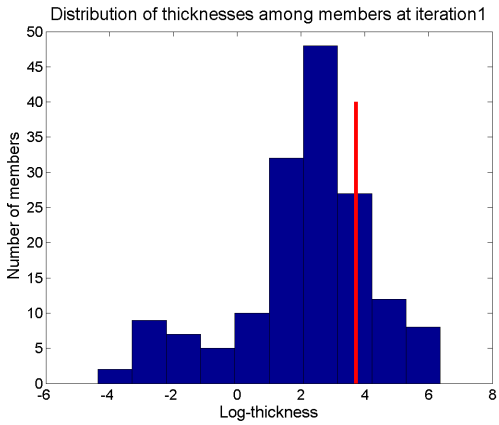
The first issue is related to the bimodality of the prior distribution in cells near the edges of the aquitard. Before the application of the EnKF, the assumption was made that in the majority of

the cells the probability distribution function could be approximated by a Gaussian distribution, if a logarithmic transformation was made. In the second experiment this assumption is valid, because we fixed the boundaries of the aquitard. In the first experiment, however, it is not, causing serious problems.

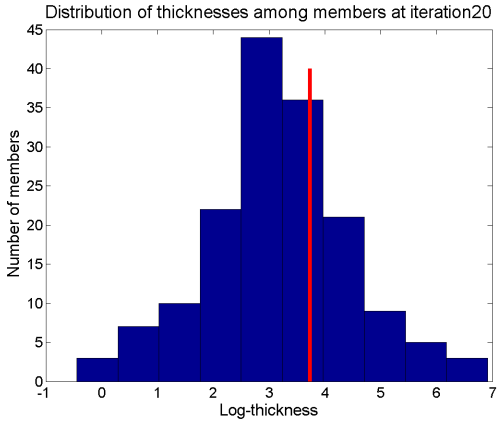
We can illustrate this with a histogram of the thickness in a specific cell among different members. Figure 5.18 provides such a histogram of a cell at location (154750,482650) in scenario A. The EnKF tries to fit the prior distribution with two modes into the shape of a normal distribution. This results in a very large first update, where some of the members adopt values that are highly unlikely. In the following updates the variance shrinks, but it is not enough to let the thickness in all cells return to a physically reasonable value. Part of the problem is that the members where the aquitard is thicker than the 'maximum allowed thickness' as defined in section 3.1, are not penalised enough due to the truncation. However, the values should not become larger than this maximum allowed thickness in the first place. An explanation for the large first update involves the second issue ans is therefore deferred to the next paragraph.



(a) Prior ensemble



(b) After 1 update



(c) After all updates

Figure 5.18: Histogram of log-thicknesses at location (154750,482650) in experiment A. The red line indicates the value in the reference scenario

5.3.2 Covariance measures linear relationships

The driving force behind the magnitude of updates is the covariance between the thickness of a cell and the observed data. The logarithmic transformation negatively influences these covariances. As the covariance is defined to describe the linear relationship between two variables, it may give wrong estimates if the underlying problem is highly nonlinear. We do not suggest that the relationship between hydraulic heads and thicknesses on a normal scale is perfectly linear, but it is definitely a finite value and of similar order for both small and larger thicknesses. On log-scale this is different. The larger thicknesses are after transformation packed on a rather small interval, while the interval between 0-10 cm is infinitely large. Since the covariance is not normalised, it is influenced by the magnitudes of the variables it is describing.

The problem clearly follows from the definition of covariance; the covariance of the log-thickness d in a cell and the measured head h at a measurement location is given by:

$$\mathrm{cov}(d, h) = E\left[d - E[d]\right] E\left[h - E[h]\right]$$

In the situation of figure 5.18, the magnitude of the covariance (and hence the updates) is extremely boosted by the first expectation. However, updating the values in the right mode of the distribution with these large covariances yields serious errors.

5.3.3 Prior distribution is not log-normal

The prior distribution is bimodal and definitely not log-normal. The large variance in the prior thickness, caused by the two modes, makes it impossible to do a transformation back to the regular thickness scale (third issue). It causes the most likely estimate to be underestimated, while the mean is severely overestimated. This means that we cannot adequately asses the quality of the prior distribution, making it difficult to judge how much the updated distributions have improved.

5.3.4 Role of the truncation constant

Prior to starting the EnKF, we suggested that the artificial thickness should have negligible influence on the groundwater flow and hydraulic heads. Ideally, the artificial thickness should have a value at the lower bound of the interval where the sensitivity of the groundwater flow model seizes. For other artificial thicknesses, the model would either be insensitive to the updates of the thickness in the analysis stage of the EnKF, or it would introduce large errors. This means that the artificial thicknesses should be defined in such a way that the resulting resistance is exactly one day. In our test area this equates to a variable truncation constant with a minimum of -9. This yields two problems. First of all, the covariance of the thickness in a cell and the hydraulic heads, is strongly influenced by the chosen truncation constant. As a result, the EnKF updates are almost doubled using $c = -9$ compared to $c = -4$. Secondly, taking a variable truncation constant requires smoothness of the permeabilities (k-values) for adjacent cells to ensure the persistence of the correlation structure. If k is not smooth, then the magnitude of updates of neighbouring cells can greatly differ as a result of the large differences in covariances for the two cells. In the test area, the k-values are not always smooth, causing the correlation structure near the edges of the aquitard to fade over time. It was experimentally verified that it is not possible to find a value for the truncation constant c that allows us to have updates without extremely high thicknesses, and at the same time enables us to distinguish between the smaller thicknesses. Here the compromise of $c = -4$ was chosen, as it allows us

to still distinguish thicknesses in the order of centimetres, while the transformed values of the updated ensemble median remain within a physically reasonable range.

As the EnKF will reshape the bimodal distribution to a normal shaped distribution, it is also not straightforward how to define the truncation constant after the updates. In the situation in figure 5.18, there is a clear indication that the aquitard is present. However, it is not always this obvious, especially if the true thickness is small. For instance, if a cell has a logarithmic thickness of -2 ($\approx$ 13cm), then it becomes difficult where to define a truncation value. By a method of trial and error, a value of -2.3 was selected, however, this value was chosen rather random and therefore no serious value can be attached to the estimations of the extent of the aquitard.

5.3.5 Filter divergence

From the discussion of the first four issues, we can conclude that they are all related to the bimodality of the distribution in some cells, in interplay with the logarithmic transformation and/or the truncation constant. This means that these are all issues which can be solved by selecting a different calibration method. Therefore the EnKF-GMM will be applied in the coming sections. The last issue, however, is not a consequence of our choice for a calibration method. In both experiments A and B, we have seen that after all updates the overall hydraulic head errors are negligible. The EnKF finds a wrong solution that satisfies the hydraulic head data. As a result, the width of the confidence interval shrinks, while large errors remain in the estimation of the aquitard thickness. This shows that with the limited amount of available hydraulic head observations, there is still too much freedom for the estimation of the aquitard thicknesses. Filter divergence occurs as a result of an amount of information contained by the hydraulic head observations that is too little.

A major concern is that no indication of the filter divergence is given, except for a non-decreasing thickness error. In practice, we are not able to compute this error since the true state is unknown. As the hydraulic head variance and error decrease both at a similar rate, it is not possible to conclude that something is wrong based on only the hydraulic head errors. A cross validation with other hydraulic head data fails to discover the problem and the only possibility to verify whether the obtained posterior distribution covers the measured thickness is by a cross validation of more boreholes. However, this means that data must have been intentionally left out during the construction of the geohydrological model to allow such verifications later on.

5.3.6 Contribution to the research goals

Summarising, we can conclude that the combination of a covariance based method with a log-transformation only works in case the exact extent of the aquitard is known. To assess the uncertainty of the extent of the aquitard, an approach where no covariances of absent and present values are combined will have to be used, as it is not possible to objectively determine a truncation constant for after the updates. The EnKF is therefore not applicable in situations where the exact extent of the aquitard is unknown. If a good estimation of the extent is available, than the EnKF can be applied within these boundaries. A most likely estimate that better satisfies the data can be found. However, the information content of the hydraulic heads is not large enough to make valid statements about the uncertainties.

The amount of information in the hydraulic heads partially depends on the chosen observation

locations. Two different sets were used for the experiments. Measurements can best be made at locations where the hydraulic head variances are large. For the second experiment, synthetic locations were used, whereas experiment A used the true locations. For experiment A it holds that while the error of hydraulic heads at observation locations are negligible, serious errors occurred in rest of the domain. The amount of measurement locations is modest, but not necessarily the limiting factor. It is the absence of measurements in the aquitard or near its edges that restricts the amount of information that the observations contain. As the synthetic model here shows strong resemblance to the true AZURE model, this is worrisome. Although the EnKF is clearly not functioning due to many reasons, we are apart from that able to conclude from this test case that the hydraulic head data does not contain enough information to achieve the research goals regardless of the calibration method. The hydraulic heads are simply measured at the wrong locations. The second set is already better and is able to decrease the hydraulic head errors in the largest part of the domain. The fact that filter divergence also occurs in the second experiment, indicates that the hydraulic heads in general do not contain enough information and that extending the measurement set is unlikely to give positive results.

Another possibility is to sample more/less frequent and for a longer period of time. In both experiments, approximately 2 years of data was assimilated monthly. Sampling more frequently is not useful as changes in hydraulic heads are gradually. Sampling over a longer period of time might allow the uncertainty ranges of the thickness to shrink a little further; but as we have already incorporated data from each season twice it is unlikely that we will obtain very different results for the most likely, mean and median estimate. In general, the simulated hydraulic heads will follow a pattern that is repeated each year, but every once in a while the hydraulic head variance will increase if something happens with the input data. For instance, extreme weather conditions or the installation of a new infiltration or extraction wells can cause the hydraulic heads to behave differently then in the past years. An example of this is in figure 5.15a where a jump in the hydraulic head error and standard deviation leads to a significant jump in lowering the thickness error. However, as these events are not likely to happen very often, the errors will remain rather large. Hence this cannot solve the filter divergence.

6 Method: EnKF-GMM

One of the main issues that occurred during the application of the regular EnKF was the large influence of the artificial thickness. As the EnKF updates are covariance steered, the value of the truncation constant determines the magnitude of the update. The truncation constant should not play any role at all and therefore a different solution must be considered, where the updates are independent of the value of the truncation constant c.

Dovera and Della Rossa (2011) describe a multimodal Ensemble Kalman Filter using Gaussian Mixture Models (EnKF-GMM). The method was succesfully applied to a 2D reservoir facies model, to estimate the permeabilities of two different facies. As discussed before, this can be seen as a rather similar application and therefore this method in a slightly adapted format was used here. Here, if we consider the thickness in case of absence of the aquitard to follow a Gaussian distribution somewhere far out of the reach where the sensitivity of the groundwater flow model seizes, then we have a situation with a multimodal distribution, constructed from two Gaussian distributions.

6.1 Introduction

In the EnKF-GMM, the EnKF is extended in such a way that the assumption of a normal prior and posterior distribution is loosened. Instead, it is assumed that the prior and posterior distribution can be approximated by a Gaussian Mixture Model. Roughly speaking, the EnKF-GMM combines multiple EnKFs in one single algorithm. In the algorithm, the members are divided into subpopulations on each of which an EnKF is applied. After each iteration, the members have the possibility to switch population with a probability determined by the likelihood of the population. The EnKF-GMM is consistent with the EnKF since in case of a unimodal distribution, it will reduce to the regular EnKF.

Prior to starting the algorithm, the number of components/modes must be known. Here two components are distinguished; 'present' and 'absent'. All members are assigned to belong to a specific component type, based on the prior probability of both components. Then, before each update is performed, the performance of the members in both components is evaluated and new probabilities are determined. The component-type of all members is resampled according to those new probabilities and after that, the subpopulations are individually updated by an EnKF.

A drawback of the method in its current form, is that each member should belong to a component type, whereas we require flexibility for each individual cell of each member. In the experiment of *Dovera and Della Rossa* (2011), where production data is assimilated for the estimation of permeability in a reservoir facies model, this assumption was loosened by combining the multimodal filter with a localisation technique. In that case, it is possible to consider a smaller area for the updates. Localised updates are made in the neighbourhoods around the measurement locations using the EnKF. After that, SIS and SGS are applied to neighbouring cells not updated with the EnKF-GMM, to secure spatial consistency.

The most significant difference between history matching or data assimilation in the deep and shallow subsurface is the amount of measurements that is available. While it is plausible in the reservoir engineering to assume that the local neighbourhoods around the measurement locations do not overlap, it is a very weak assumption in the context here. The level of uncertainty in the shallow subsurface is much lower, requiring a cell-by-cell possibility to assign a component

type. Therefore, the algorithm of *Dovera and Della Rossa* (2011) was slightly adjusted to allow for more flexibility. The main structure, however, remains the same.

6.2 Theoretical formulation

Suppose that the Probability Distribution Function (PDF) of each element j of the forecast state vector $\mathbf{x}_t^f$ at time t is a Gaussian Mixture of N_c components, e.g.,

$$f_j(\mathbf{x}_t^f) = \sum_{k=1}^{N_c} \pi_k^f(j)(f_{j,k}(\mathbf{x}_t^f))(j), \qquad f_{j,k}(\mathbf{x}_t^f) = N(\mathbf{x}_t^f; \mu_j^k, P_j^{f,k})$$

where $\sum_{k=1}^{N_c} \pi_k^f = \mathbf{1}$ and $\pi_k^f \geq \mathbf{0}$. The vector π_k^f has length $(n_x * n_y)$x1 and provides the prior mixture weights for each cell. It can be interpreted as a vector that holds the probability on component k for each cell. Further, μ_j^k is the conditional component mean, which is the mean of the state of all members having component k in cell j and similarly $P_j^{f,k}$ is the conditional component forecast covariance matrix. The dimensions of μ_j^k and $P_j^{f,k}$ are $(n_x * n_y + n_m)$x1 respectively $(n_x * n_y + n_m)$x$(n_x * n_y + n_m)$. We note that this covariance matrix can be defined in multiple ways. The definition here will be discussed in section 6.3.1. Further, we should note that most of the time indices have been left out to simplify notation. However, all covariance matrices, component mean vectors, mixture weight vectors and probability distribution functions are dependent on time. The notation $(.)(j)$ is used to indicate the j^{th} element of a vector.

At time t, observations $\mathbf{y}_t$ become available. The posterior PDF can be written in a form similar to the prior PDF by The Law of Total Probability. It is given by:

$$f_j(\mathbf{x}_t^a) = f_j(\mathbf{x}_t^f|\mathbf{y}_t) = \sum_{k=1}^{N_c} P(\text{type in cell } j = k|\mathbf{y}_t)(f_j(\mathbf{x}_t^f|\mathbf{y}_t, k))(j) = \sum_{k=1}^{N_c} \pi_k^a(j)(f_{j,k}(\mathbf{x}_t^f|\mathbf{y}_t))(j)$$

The posterior weights directly follow from Bayes' theorem:

$$\begin{aligned}
\pi_k^a(j) &= P(\text{type in cell } j = k|\mathbf{y}_t) \\
&= \frac{P(\mathbf{y}_t|\text{type in cell } j = k)P(\text{type in cell } j = k)}{\sum_{l=1}^{N_c} P(\mathbf{y}_t|\text{type in cell } j = l)P(\text{type in cell } j = l)} \\
&= \frac{N\left(\mathbf{y}_t; H_t\mu_j^k, H_t P_j^{f,k} H_t^T + R_t\right)\pi_k^f(j)}{\sum_{l=1}^{N_c} N\left(\mathbf{y}_t; H_t\mu_j^l, H_t P_j^{f,l} H_t^T + R_t\right)\pi_l^f(j)}
\end{aligned}$$

Once the posterior weights are computed, the ensemble members need to be adjusted to fit the posterior probabilities. For this we need an algorithm that 1) will redistribute the component types among the members such that the probability of a component type in each cell satisfies the posterior probabilities 2) allows a correlation structure for the indicator after redistribution, and 3) preferably minimises the number of cells that need a transfer. Here we use the following algorithm:

Algorithm 1: Redistribution of component types
At the first assimilation step:

1. Select all areas with a probability of at least 0.9 on aquitard presence;
2. Locate the boundaries of these areas and regularly select a subset of points on each contour. (here each fifth node was selected);
3. Compute a Thiessen Polygon mapping around the selected nodes to construct influence areas for each of the selected nodes.
4. Randomly pick a starting node and assign a random drawn value from $U(0,1)$ to this node. Use this node as start and end point to simulate a Brownian bridge along the selected nodes on the contour.
5. Store the realisations of the Brownian bridge and the Thiessen Polygon mapping.

At all assimilation steps:

1. Assign indicators to all cells. For each selected node consider its influence area. Assign indicator 1 ('present') to all cells within the influence area for which the probability in a cell is larger than the simulated Brownian bridge value at selected node and assign 0 ('absent') if smaller.

By means of simulations it can be shown that the above algorithm produces an ensemble with the correct probabilities of occurrences for large sample sizes. The verification of a correlation structure is harder, but by observation it follows that in most cases, moving away from the boundary of the aquitard will give a decreasing posterior probability. Therefore the cells remain connected and some sort of indicator correlation structure is maintained. Preferably we also minimise the number of transfers. A minimisation is easily computed, however, priority should be given to satisfying the posterior probabilities and maintaining a correlation structure. By fixing the realisations of the Brownian Bridge and influence areas in the proposed algorithm, the number of transfers after the first assimilation step is relatively low, as the posterior probability depends on the prior probability.

Once the new component types are known, the transfers can be established. In *Dovera and Della Rossa* (2011) the component covariance matrices are decomposed using a Cholesky decomposition, and after that an inverse Cholesky transformation and a normal Cholesky transformation are performed subsequently. This transforms the error vector with a covariance structure given by $P^{f,k}$ (k the old component type) first to an uncorrelated random vector, and after that to a vector with covariance matrix $P^{f,l}$ (l the new component type). Note that $P^{f,k}$ and $P^{f,l}$ are not depending on j in the research of *Dovera and Della Rossa* (2011). In Appendix F the background of Cholesky transformations is given.

This is a robust method for members that belong entirely to one component type. Here, however, the situation is rather complicated. We require a method that assigns values to the transferred cells that are both correlated with the other transferred cells and with the cells that already had the new component type. Application of the Cholesky transformation on all cells that get transferred using the component covariance matrices per cell, will only secure the correlation structure among cells that are transferred all together. As the filter continues, the group will each time have different compositions, causing the correlation structure around the edges of the

aquitard to fade over time.

It is probably best to apply a geostatistical method here. Either SGS or OK can be performed, conditioning the cells that need a transfer on the cells that already contained the new component type before. For this, either the prior variogram or a new variogram estimated after the updates until then, could be used. However, as we are here only testing whether the ENKF-GMM has potential at all, we will choose a method that is easy to implement and computationally efficient as long as the assumptions made are not too restrictive. Although computationally not very efficient, the Cholesky decompositions are easy to implement. A drawback, however, is that for small sample sizes the covariance matrices can contain large errors, and furthermore, that the matrices can fail to be positive definite. A discussion on this is included in Appendix E. Here we will just state that it is impractical to use Cholesky decompositions and therefore a different approach will be used.

In the specific problem here, one component has an artificially assigned distribution. This provides us with the flexibility to also assign any desired correlation structure to it, without consequences. If the same correlation structure is assigned to the areas where the aquitard is present and absent, then adding/subtracting the difference in means is enough to satisfy the correlation structure. That is, without any additional updates of the correlation structure by the EnKF. Since the covariance structures of the components will be updated independently, errors will be introduced.

The posterior PDFs of the conditional components are Gaussian PDFs, since any conditional distribution of a Gaussian distribution is Gaussian again. Each cell of each member can be updated with the normal EnKF equations, where the mean and covariance matrix are the component mean and component covariance matrix. This forms the basis of the EnKF-GMM and can be used to construct the following algorithm:

Algorithm 2: EnKF-GMM

1. Determine likelihood of component k per cell.

   For each cell $j = 1 : (n_x * n_y)$:

   (a) compute membership probability $\pi_k^f(j)$, component mean μ_j^k and forecast error covariance matrix $P_j^{f,k}$ for component $k = 1, ...N_c$.

   (b) compute $\pi_k^a(j)$ for $k = 1, ...N_c$ as

$$\pi_k^a(j) = \frac{N\left(\mathbf{y}_t; H_t\mu_j^k, H_tP_j^{f,k}H_t^T + R_t\right)\pi_k^f(j)}{\sum_{l=1}^{N_c} N\left(\mathbf{y}_t; H_t\mu_j^l, H_tP_j^{f,l}H^T + R_t\right)\pi_l^f(j)}$$

2. Assign new component types to each cell. For each member $m = 1 : n_e$:

   (a) Apply algorithm 1 to define the new component type per cell.

3. EnKF update: For each cell $j = 1 : (n_x * n_y)$ loop over all ensemble members $m = 1 : n_e$

   (a) set k as the old component of cell j of the member m

   (b) set l as the new component of cell j of the member m

(c) compute the auxiliary state $(\mathbf{x}_t^{f'})_m$ with respect to element j, for member m according to

$$(\mathbf{x}_t^{f'})_m = \mu_j^l + (\mathbf{x}_t^f)_m - \mu_j^k$$

(d) update statevector entry j using the updating equation for the new component l by:

$$(\mathbf{x}_t^a)_m(j) = (\mathbf{x}_t^{f'})_m(j) + K_j^l(\mathbf{y}_t - H_t(\mathbf{x}_t^{f'})_m)$$

where $K_j^l = P_j^{f,l} H_t^T (H_t P_j^{f,l} H_t^T + R_t)^{-1}$.

6.3 Configuration

For the EnKF-GMM we will again use a confirming EnKF. Also, the assumptions that were made on the model and measurement error in the regular EnKF, remain. In addition, we will discuss how to deal with small ensemble sizes. Moreover, the settings of algorithm 2 and the likelihood computation are discussed.

6.3.1 Component covariance matrices

The component covariance matrices should reflect the correlation structure of the component type. Earlier it was suggested to divide the members into two groups based on their component type in a specific cell and after that compute sample covariance matrices of both types. However, multiple component types can occur in other cells. Therefore a member should only be allowed in the covariance computation if both cells have the same component type. The other members will be ignored, resulting in adjusted covariance matrices, whose elements are given by:

$$P_j^k(a,b) \approx \frac{1}{n_e(j,a,b)-1} \sum_{i=1}^{n_e(j,a,b)} \left(\alpha_i^k(j)\alpha_i^k(a)((\mathbf{x}_t^f)_i - \bar{\mathbf{x}})\right)\left(\alpha_i^k(j)\alpha_i^k(b)((\mathbf{x}_t^f)_i - \bar{\mathbf{x}})\right)^T$$

where

$$n_e(j,a,b) := \{\#\text{of ensemble members with component } k \text{ in cell } j, a \text{ and } b\}$$

and α_i a coefficient vector of length $n_x * n_y$ whose elements are defined as:

$$\alpha_i^k(j) = \begin{cases} 1 & \text{if cell } j \text{ of member } i \text{ has component type } k \\ 0 & \text{otherwise} \end{cases}$$

Remark that the correlation between element a and b is computed solely using members with component type k in cell j while other members with component type l in cell j might have component k in cells a and b.

6.3.2 Ensemble size

The EnKF uses a finite sample approximation of the covariance matrices. For small sample sizes, large errors can be introduced, and this, on its turn, will cause errors in the estimation of the Kalman Gain matrix. When this happens, spurious correlations occur, which lead to unrealistic updates of the model parameters and state variables. A regularisation process applied on either the covariance or the Kalman gain matrix is then necessary. The most common method to resolve this issue, is distance-dependent covariance localisation. However, with an ensemble size of 160 in the regular EnKF, the spurious simulations seem limited and hence it was not necessary to apply a localisation technique.

The EnKF-GMM can be seen as the application of two EnKFs simultaneously for two separate populations, where inhabitants have the possibility to switch population. Therefore, we require a sufficiently large sample size for both separate subpopulations. If two components are used, each cell will have a percentage of members labelled as component 1, and the remainder as component 2. In the most extreme case, only one member will belong to component 1 and no distribution can be estimated. However, then the probability on component 2 is close to $\frac{n_e-1}{n_e}$, which for large n_e will approach 1. In this case, and for all other cells with an ensemble sizes smaller or equal to 3, the probability λ_j^k in step 1b is manually put to 1. This means that an ensemble with size 5 is also considered by the filter. One can be certain that rigorous errors are made in estimating the covariance matrices. However, as the iterations continue, the likelihood estimator will either decide that more members should be added and the ensemble size will increase, or the members will cross over to the other ensemble. A third possibility, is that the members are insensitive for the likelihood estimate. In that case, one population can make strange updates, but it will remain to have a small probability, making it not a major concern. Therefore, taking the above under consideration, it was deemed not necessary to apply any localisation techniques. However, if the ENKF-GMM shows encouraging results, it might be used for a fine-tuning later on.

6.3.3 Likelihood computation

The computation of the posterior probabilites of the component types uses a likelihood computation. Theoretically, the likelihood of all components is equal in case there is no correlation between the observation of an hydraulic head and the thickness/component type in a specific cell. Therefore, at large distances from the observation locations, the posterior probabilities should equal the prior probabilities. In practice, this often appears to be an invalid assumption. As a result of the finite sample sizes, spurious correlations that entirely eliminate component types in specific cells can occur. This makes it inevitable to apply a localisation technique. We introduce a correlation length of 2km, and measurements at a larger distance are ignored. If no measurement are present within these 2km, the measurement nearest by was selected. This correlation length was chosen by trial-and-error, where a value was selected that allowed the probabilities to gradually vary and not directly make all probabilities 1 or 0 after the first update.

6.3.4 Prior distribution

The prior distribution for the EnKF-GMM should not be different from the prior distribution of the EnKF. Special attention is only required for the artificial distribution of the component type 'absent'. The artificial distribution should be a Gaussian distribution, somewhere far outside of AZURE's sensitivity range. Furthermore, it should have the same correlation structure as cells the distribution of component type 'present'. A prior ensemble satisfying these specifications was constructed by applying OK to the entire domain, and after that adding a sufficiently large negative value (-12) to the values in the 'absent' region.

6.3.5 Transfer of head data

In step 3 of algorithm 2 the auxiliary state vectors are computed. In this paragraph we will focus on the consistency of the elements of the state vector. An example is given to illustrate the problem that occurs:

Example:
In this example we consider the update of the EnKF-GMM of one ensemble member. Denote $\mathbf{h}_t^f = \mathbf{x}_t^f(n_x n_y + 1 : n_x n_y + n_m)$ as the simulated hydraulic heads at the measurement locations of this member. In figure 6.1, four cells of this member and their component types are shown before (figure 6.1) and after the EnKF-GMM update (figure 6.1b-d). Prior to performing an update with the EnKF-GMM, the simulated hydraulic heads at measurement locations are consistent with the thicknesses and component types. In the redistribution step of the EnKF-GMM, the component type of two cells is shifted (see figure 6.1b-d). With this shift of component type, the thickness in the cells is adjusted and therefore the updated static state variables are not consistent with the hydraulic heads anymore.

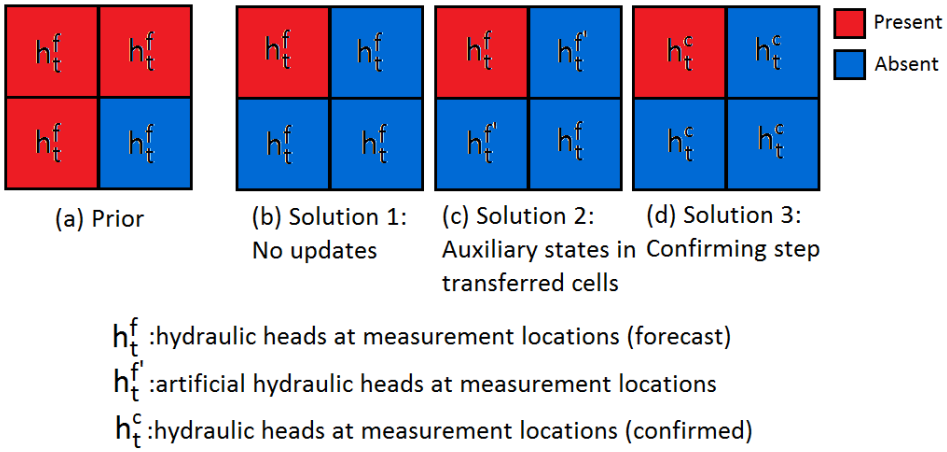


Figure 6.1: Transfer of head data

A first solution (figure 6.1b) is to ignore these inconsistencies and assume that the perturbations of the hydraulic heads are small. In that case it suffices to not make any updates. A second solution (figure 6.1c) is to compute auxiliary hydraulic heads for the shifted cells. This may compensate for some of the errors, but meanwhile it is a cause for confusion as there may be more than one hydraulic head at a measurement location for the same member. This means that not all cells are updated using the same simulated dynamic variables. A third solution is to introduce another confirming step between steps 3c and 3d of algorithm 2. Once the static state variables have been updated in step 3c, the model is restarted and simulated forward until time t. Hydraulic heads at the measurement locations that are consistent with the static variables and the same for all cells can be obtained. A drawback, however, is that this will almost double the amount of required computational effort. Therefore, in this thesis solution 2 will be used.

6.3.6 Updating the artificial distribution

The thickness of cells with type 'absent' is an artificial random variable. Therefore, once the realisation appears to be of type 'absent', its value is actually deterministic. As a result, the variance of the artificial distribution is theoretically 0. Therefore the covariances between thickness in a cell and hydraulic head observations are also theoretically zero. As discussed earlier, we have assigned a correlation structure to the artificial thicknesses, that is similar to the structure of the component type 'present'. The issue is now whether we should update the artificial distribution, or not. On the one hand, it makes no sense at all to make artificial updates since the entire distribution was defined under the assumption that the specific realisations would not

influence the groundwater flow model. Therefore all correlations are by assumption spurious. On the other hand, although they are spurious, they will keep the updated thicknesses close to each other. If we consider two cells near the edge of the aquitard; then their correlation with respect to the hydraulic head measurements is approximately the same, even if their component types differ. This is result of the (by approximation) identical correlation structure of the thicknesses and the property that covariance is centred around the mean of a variable. In the EnKF-GMM this yields updates that are also similar, and hence the thicknesses will remain correlated. Therefore, allowing updates for the artificial distribution may also allow the correlation structure to better be maintained. In this thesis the artificial distribution is updated in all experiments.

6.4 Implementation

The EnKF-GMM can be summarised in the flow chart that is shown in figure 6.2.

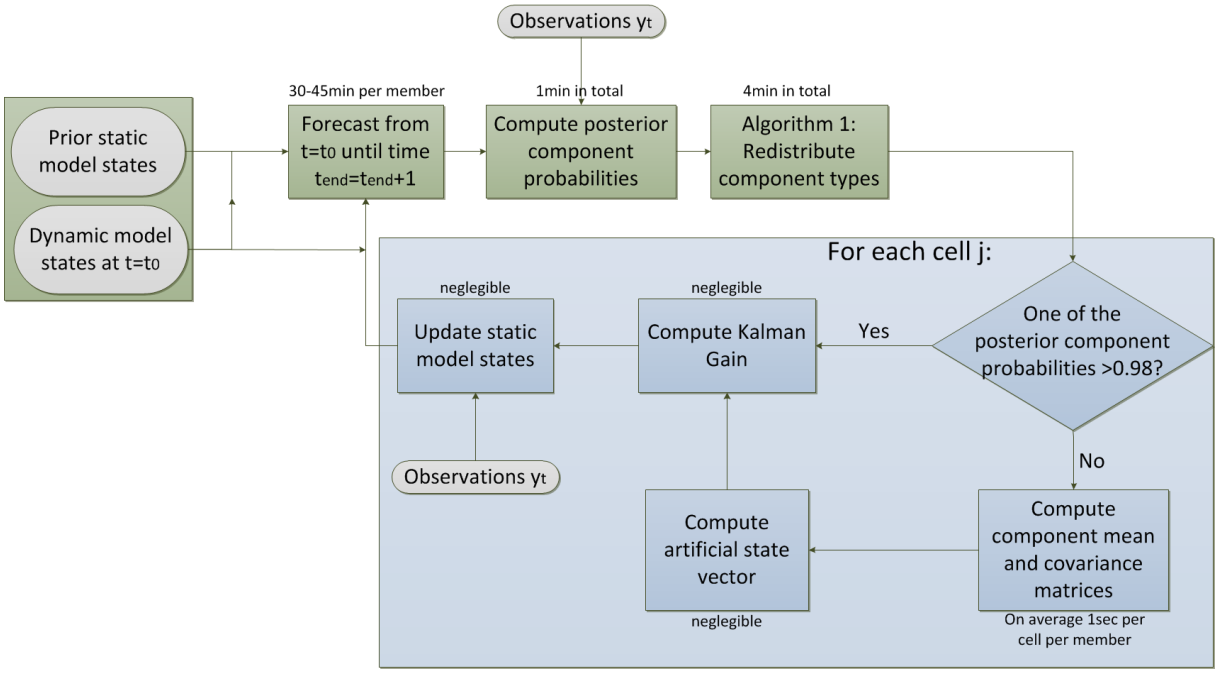


Figure 6.2: Flowchart of EnKF-GMM

Compared to the flow scheme of the standard EnKF, a number of processes have been added. The blue box starts with a decision point, where it is decided whether a regular EnKF or an EnKF-GMM is performed. If only one component type is present, then the probability is 1 and hence the branch "Yes" is followed. This leads to a regular EnKF update.

Since processes have only been added, the required computational time will increase. In the flow chart, approximate computation times are indicated for the experiments performed in section 7. The computation of the posterior component probabilities and algorithm 1 have a negligible contribution to the total required computation time. The blue box, to perform the EnKF per component type, takes a lot more time if multiple component types are present. In the regular EnKF, only one covariance matrix is computed at each iteration, whereas up to $N_c * n_x * n_y$ different covariance matrices (with lower dimensions) are needed here. The computation of these matrices requires quite some time (and memory). Some of these matrices are the same

and therefore the re-use of already computed covariance matrices and approximations should be considered. For instance, in cells with a unimodal distribution the EnKF-GMM reduces to the EnKF. We should take advantage of that. The diamond shaped box therefore divides the cells into three different groups, depending on their component type. The first two groups will contain cells with one type only, whereas the last group has a Gaussian mixture distribution. Only this last group needs the update steps of step 3 of the EnKF-GMM, the others can be updated all together using the regular EnKF. The amount of time needed for the updates of the first two groups (path indicated by "yes") is again negligible compared to the third group (path indicated by "no").

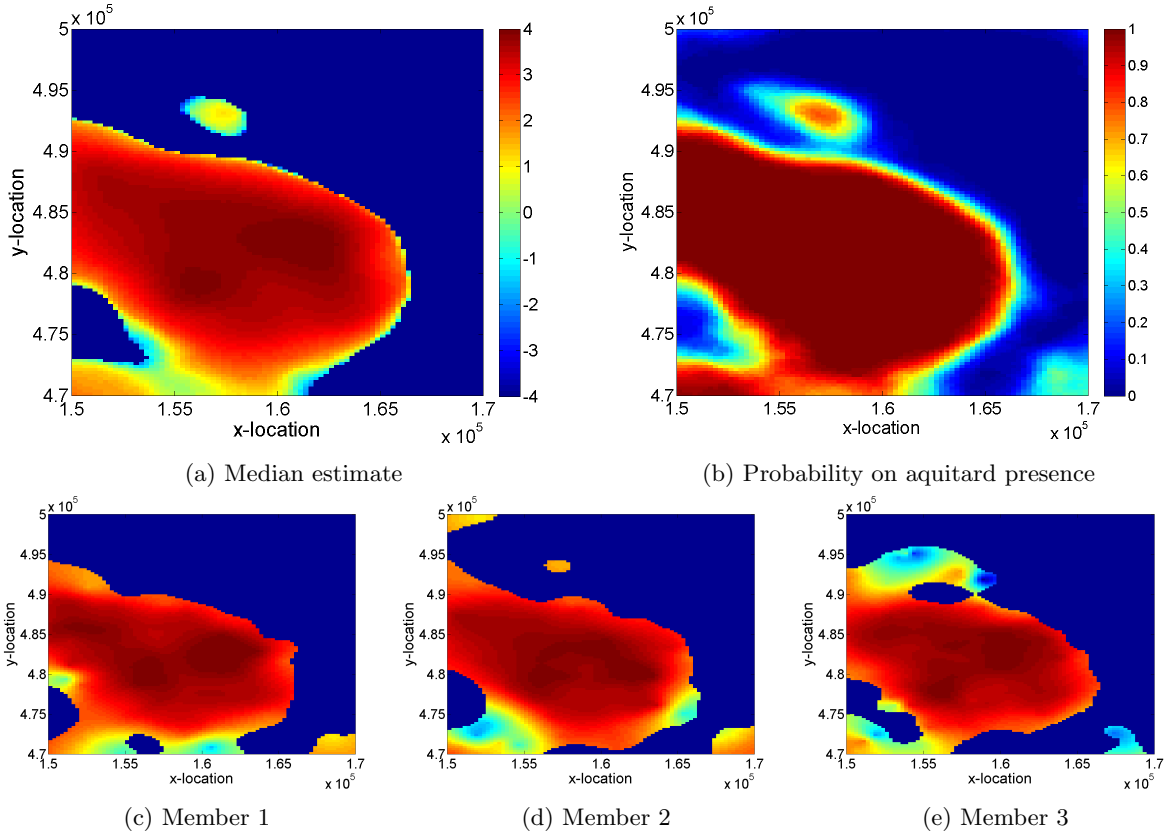
Even with these modifications, the computational effort is rather large. The EnKF-GMM introduces a for-loop over all cells in the update step and the computational time of this step is therefore linearly dependent on the number of cells. As the AZURE model already scales up the grid before simulating, it is a logical step to decrease the computational effort further by also using this coarser grid in the state vector.

7 Application of EnKF-GMM to a synthetic model

The EnKF-GMM was applied to a synthetic test model, similar to the model in section 5. In the coming subsections, the prior ensemble, measurement sets and results are discussed.

7.1 Initial ensemble

The EnKF-GMM was applied to ensemble C, shown in figure 7.1. Ensemble C is related to ensemble B in the sense that both ensembles are based on the same boreholes. As a result, the members have the same thicknesses at each cell where the component type is equal. The reference situation of both scenarios is the same to allow for a comparison.



(a) Median estimate

(b) Probability on aquitard presence

(c) Member 1

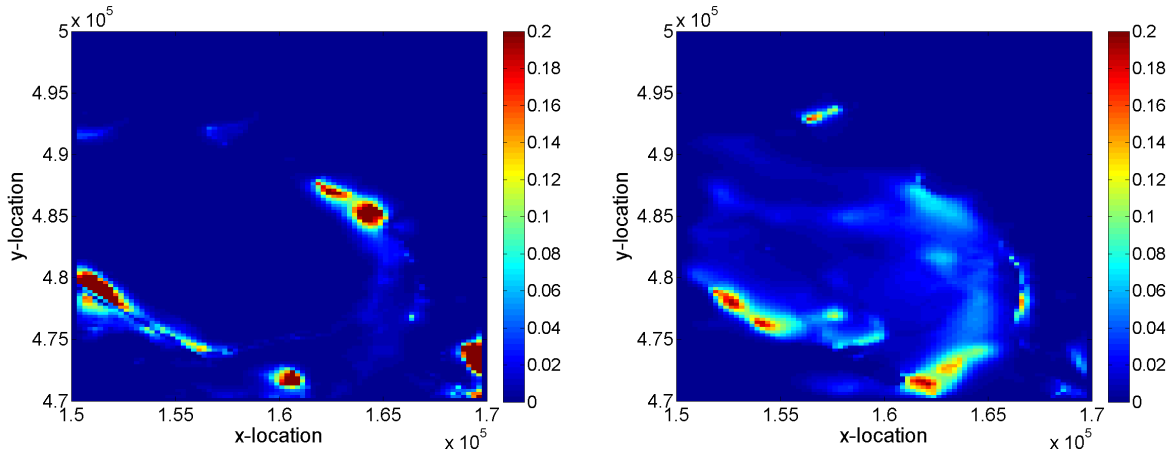
(d) Member 2

(e) Member 3

Figure 7.1: Prior ensemble C on logarithmic scale.

7.2 Measurement sets

The prior hydraulic head variances of ensemble C in aquifers 4 and 5 are shown in figure 7.2. Compared to ensemble B (figure 5.6), the variance of hydraulic heads is much larger near the edges of the aquitard. A third measurement set with 90 measurement locations near the edges of the aquitard is therefore defined. This set is shown in figure 7.3.



(a) Aquifer 4 (above aquitard)

(b) Aquifer 5 (below aquitard)

Figure 7.2: Prior hydraulic head variances [m^2] ensemble C

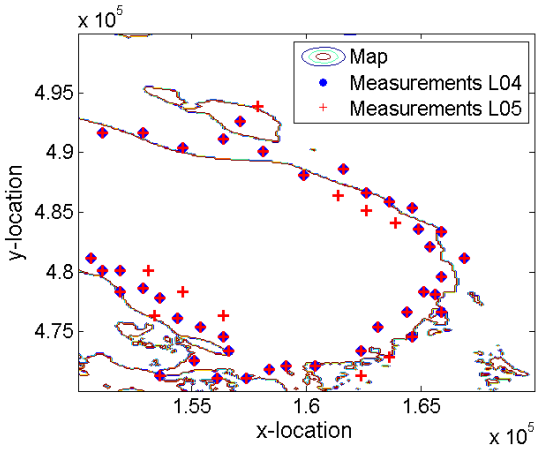


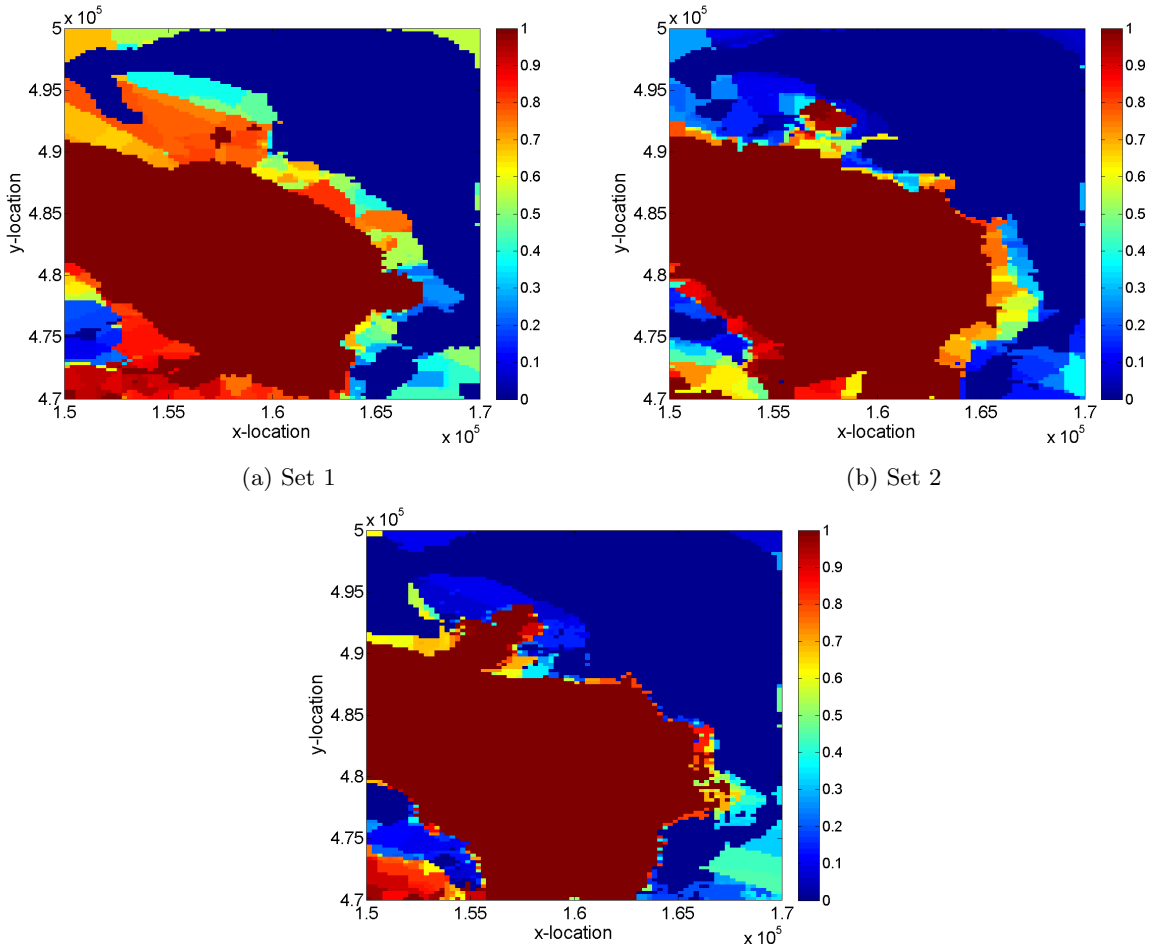
Figure 7.3: Measurement set 3

7.3 Results

The EnKF-GMM was applied to ensemble C in combination with all three measurement sets. Figures regarding the hydraulic head data are not shown here, as they all show a significant decrease of the hydraulic head errors at the measurement locations. As the EnKF's performance of improving the model performance (at measurement locations) was already good, this was to be expected. Therefore we will only discuss the estimation of the extent of the aquitard and the posterior PDF of the thickness. In addition, we will discuss an experiment based on only likelihood updates, and the influence of holes on the aquitard.

7.3.1 Estimating the extent of the aquitard

The first research goal involves a better estimation of the extent of the aquitard. For all three experiments a probability plot of occurrence of the aquitard is shown in figure 7.4.



(a) Set 1

(b) Set 2

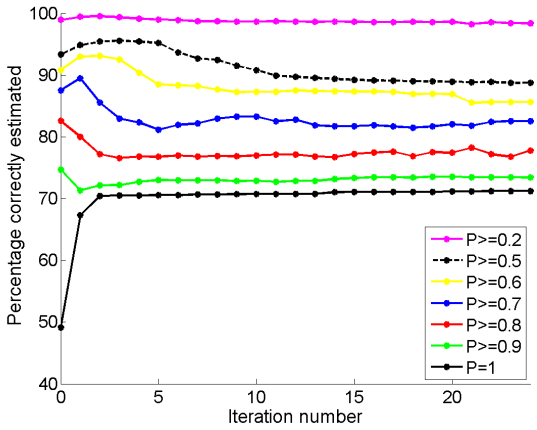
(c) Set 3

Figure 7.4: Updated probability of occurrence for measurement sets 1,2 and 3

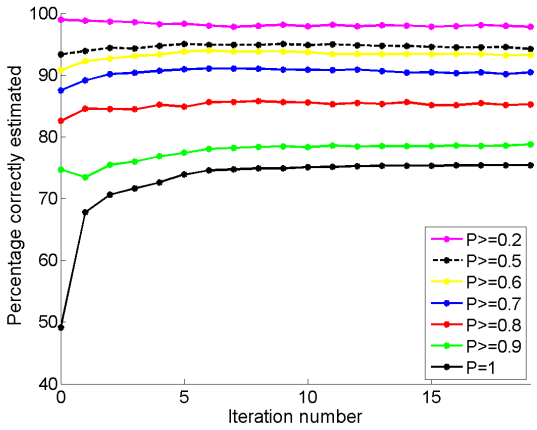
Figure 7.4a shows much more uncertainty than figures 7.4b and 7.4c. Especially in the upper left corner where the aquitard is absent in the true scenario, a large area with a probability between 0.3 and 0.7 on aquitard presence occurs. This can be explained by the absence of measurement locations in this corner in the first measurement set.

For the first and third set, there are cells where no measurement locations are present within the estimated correlation length of 2km. Therefore the measurement location nearest by was selected for the computation of the posterior probability distributions. The result of this selection can be observed in figures 7.4a and 7.4c; straight lines divide the areas of where cells are closest to a particular observation location. Possible solutions to this error are discussed at the end of this section. In the second experiment this phenomena is absent. The measurement locations are distributed more uniformly over the domain, having no areas where cells have no measurement location within the prescribed correlation radius.

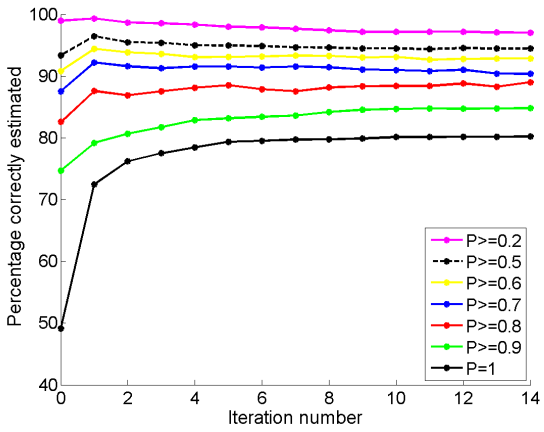
An estimate with large uncertainties where the truth lies within these uncertainty ranges, is much more valuable than an estimate that claims presence with probability 1 while the aquitard is absent. Figure 7.5 shows the percentage of cells for which the probability of occurrence on the correct component type was at least $0.2, 0.5, 0.6, 0.7, 0.8, 0.9$ or 1.



(a) Set 1



(b) Set 2



(c) Set 3

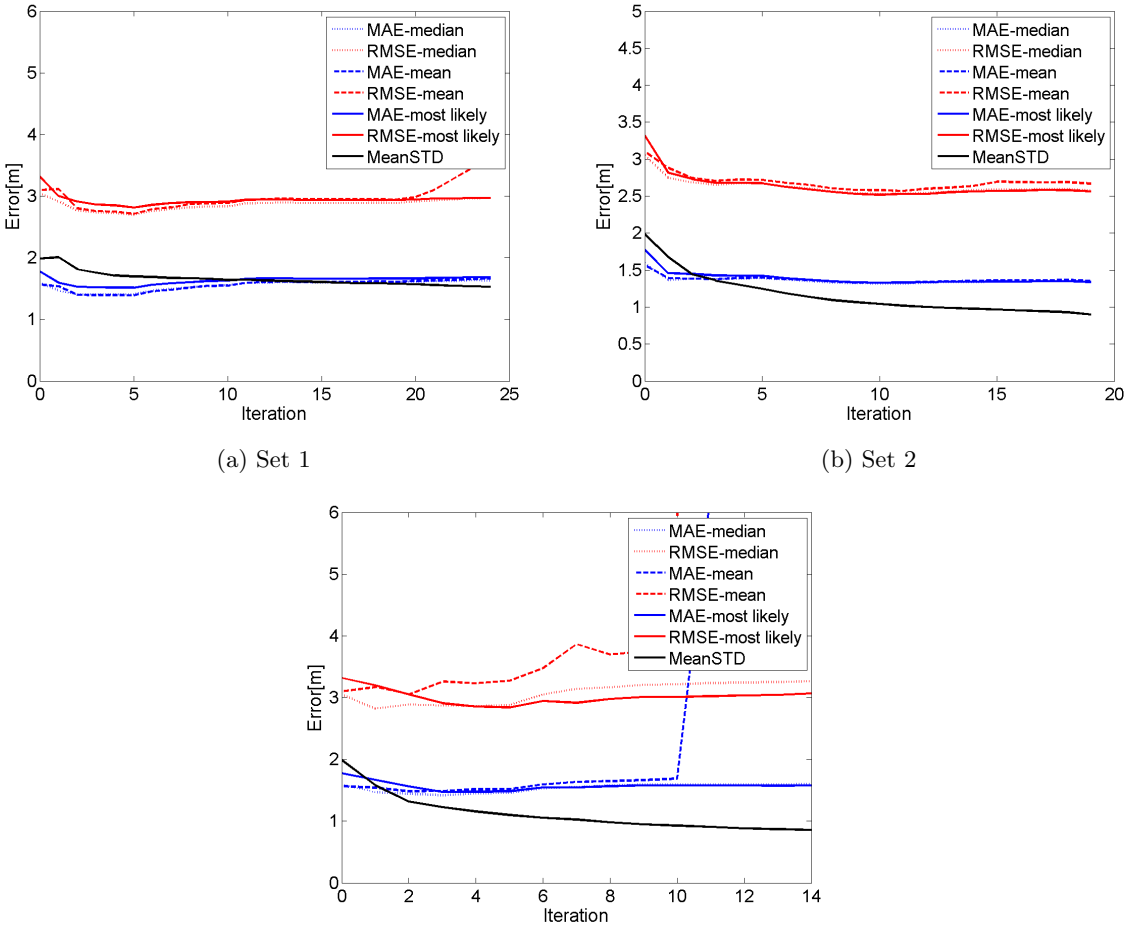
Figure 7.5: Percentage plot for measurementsets 1,2 and 3

Typically, the bundle of graphs will grow towards each other. Namely, once the component type of a cell is estimated with probability 1, it cannot change anymore. This means that the black graph that belongs to P=1 will always be a monotonically increasing graph. However, the number of severe errors (here defined as a wrongly estimated cell with $p = 0.8$) will also increase as the filter continues and hence the graph belonging to 0.2 will decrease.

From a comparison of figures 7.5a-c it follows that the estimation made with the third measurement set is rather accurate. The percentage of cells that is correctly estimated with probability 0.7, 0.8, 0.9 or 1 is significantly better than for measurement sets 1 and 2. In practice, it is most likely that a component type is assigned if its probability is larger than 0.5. Therefore, the most important graph is $P >= 0.5$. For this percentile, set 2 and 3 perform comparable, while set 1 is by far the worst performer. The amount of correct estimations even decreases with respect to the prior estimate. With respect to the severe errors set 1 performs best. However, also set 2 and 3 perform quite well.

7.3.2 PDF of the thickness

The posterior distribution of the thicknesses is unimodal in all cells where the probability of occurrence is 0 or 1, and bimodal in the remainder of the cells. To assess the quality of the distribution after each iteration, we will first select the mode with the largest probability, and after that compute the mode, median, mean and variance of this mode. For the artificial distribution it holds that these are all 0 after transformation. Figure 7.6 shows the average thickness errors per measurement set.



(a) Set 1

(b) Set 2

(c) Set 3

Figure 7.6: Thickness errors [m] for measurement sets 1,2 and 3

In all three experiments an improve in the thickness error is observed during the first few iterations. After that, the errors remain more or less constant or even slightly increase. Meanwhile, the STD remains decreasing. Again we can conclude that filter divergence occurs, just as in the experiments with the regular EnKF. Also, we should note that the state variables (log-thicknesses) were manually bounded by -9 and 5 before transformation back to the regular scale. This was done in order to remove the outliers. In the first two experiments this is sufficient to obtain physically reasonable distributions. In the third experiment, however, a too large variance remains in a number of cells. This results in a mean estimate with large errors at later time steps.

The performance, when measured this way, is strongly influenced by the level of the black graph $P = 0.5$ in figure 7.5. The errors will be larger if the presence/absence of the aquitard is wrongly estimated. However, if with a probability of 0.51 the wrong type is estimated, a large error is accounted for, while in practice this means we are uncertain whether the aquitard is present or absent. In the first experiment, this will most frequently occur as there the level of the $P = 0.5$ graph is lowest and the spread of the various graphs in figure 7.5 largest. Figure 7.7 shows the thickness errors in a situation where the mode, mean and median of the 'correct' type were selected if the probability on the 'correct' type was at least 0.2. We observe that the graphs in figure 7.6 show an increasing trend, while the errors in figure 7.7 remain constant or even slightly decrease.

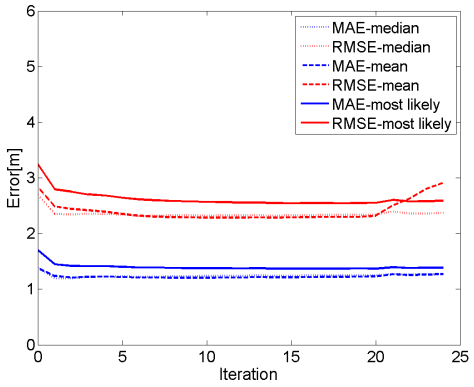
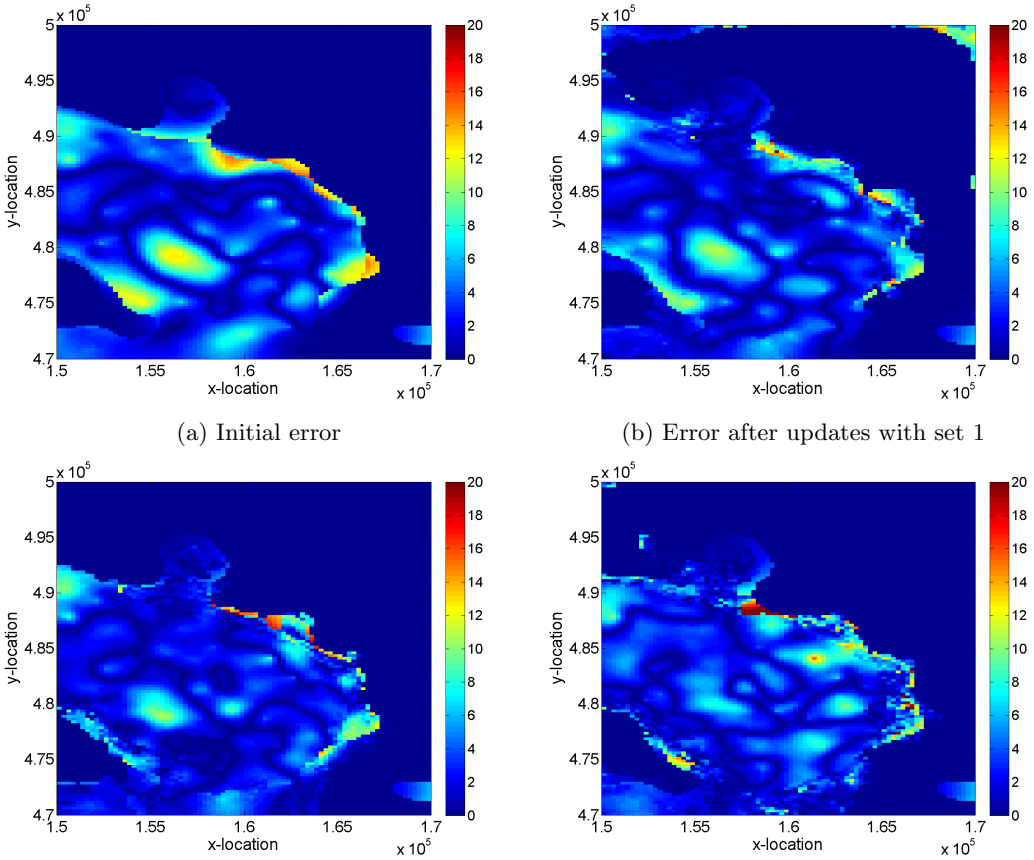


Figure 7.7: Adjusted thickness [m] for measurement set 1

In figure 7.8 the thickness errors of the most likely estimates in the entire domain are shown. Figure 7.8a shows the intial error, and figures 7.8b-d the errors for measurement sets 1-3. These figures confirm that overall, the thickness has improved. Only in a few cells the estimates are really wrong. The colour scale was cut-off at 20 metres. In figures 7.8b and 7.8d this is not a cut-off since all errors are below 20m. In figure 7.8c there are several pixels where the error is larger than 20 meters, with as maximum a pixel at $(x, y) = (16590, 477400)$ where the most likely thickness error is 63m. The error grows larger as the updates continue. This explains the large increase in the mean error in figure 7.6c.



(a) Initial error

(b) Error after updates with set 1

(c) Error after updates with set 2

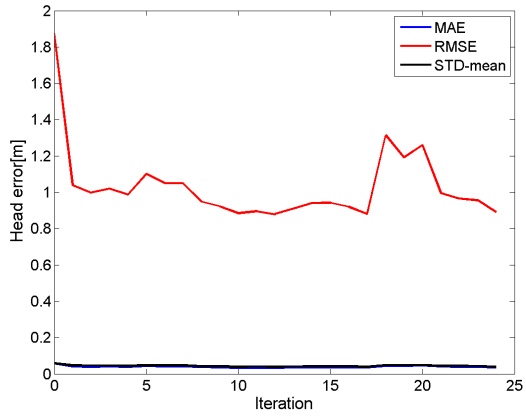
(d) Error after updates with set 3

Figure 7.8: Thickness errors (MAE) [m] initially and for measurement sets 1,2 and 3.

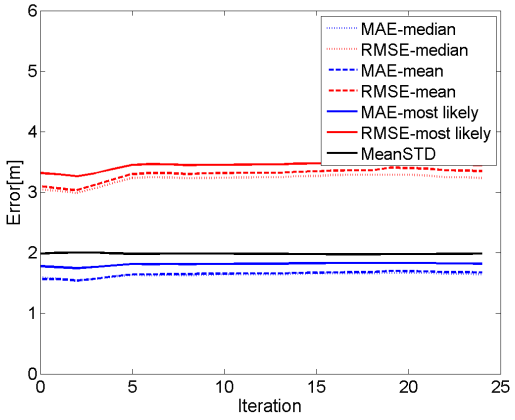
7.3.3 Without an EnKF

In the EnKF and EnKF-GMM, the overall performance of AZURE to predict hydraulic heads, improves. In this paragraph we will investigate to what extent we should ascribe those improvements to the EnKF/EnKF-GMM, or to the better estimates of the extent of the aquitard. Therefore the filter was also executed without performing the EnKF updates. That is, at each iteration the new proportions were evaluated, the transfers established and the update consists of adding or subtracting 12 to the cells whose component's type is shifted. The second measurement set was used. The results are shown in figure 7.9.

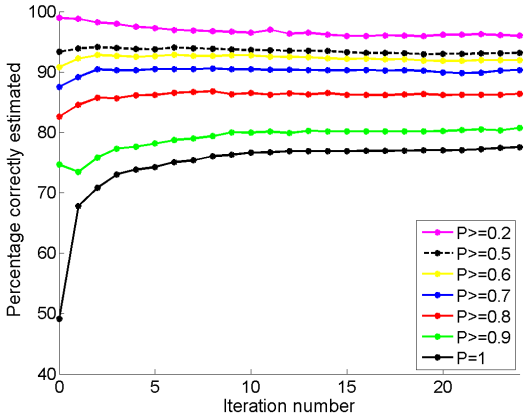
At the first iteration, the hydraulic head errors are roughly halved. After that, they remain more or less constant while the thickness error even slightly deteriorates again. Compared to the hydraulic head errors in scenario B of the EnKF (see figure 5.15a), there is a complete difference in magnitude. As the ensembles B and C are related and updated with the same observations, it follows that the influence of the varying extent of the aquitard on the hydraulic heads is significant. The correctly estimated percentages do increase, and, compared to the results obtained with measurement set 2 in the former subsection, the estimates are at least comparable. As the thickness errors are larger in this last experiment, it can be concluded that the EnKF in the EnKF-GMM has a positive influence on the estimation of the aquitard. However, we do need to emphasise that the filter divergence prohibits us to derive valuable PDFs from the updated ensemble.



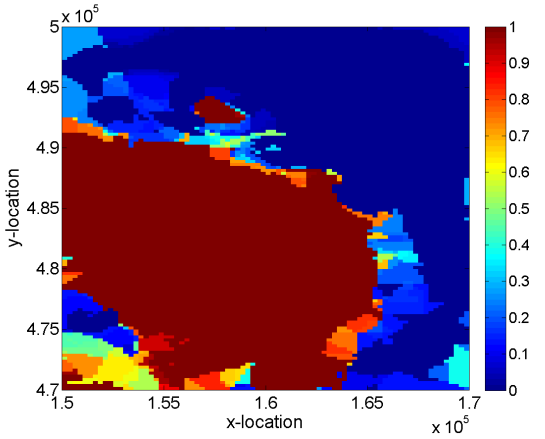
(a) Head error



(b) Thickness error.



(c) Percentage plot- percentage of correctly estimated facies with probability as indicated.



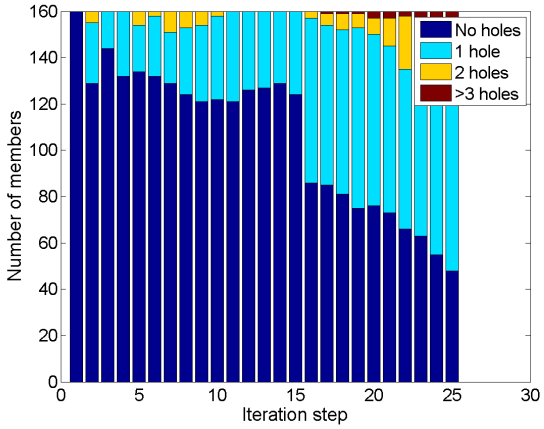
(d) Updated probabilities on aquitard presence.

Figure 7.9: Results experiment without EnKF

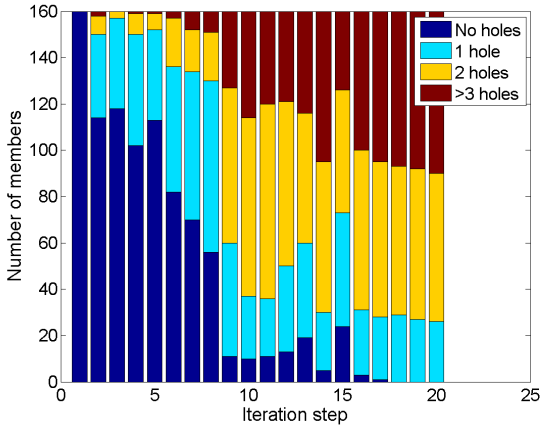
7.3.4 Holes in the aquitard

One of the assumptions made about the 'truth' is that no holes occur in the aquitard. Therefore none of the members of the prior ensemble contains holes, and in general a regular EnKF would also not create them. Therefore an ensemble updated with an EnKF will not contain members with holes. However, in the EnKF-GMM an indicator resampling algorithm is present. During the indicator resampling, holes can be created. As we outlined in subsection 5.2 and Appendix D, it becomes much more difficult to estimate the geological structure if holes are present since the correlation structure between hydraulic heads and thicknesses becomes more complex. In this paragraph we will show how holes in the aquitard (negatively) influence the results of the EnKF-GMM.

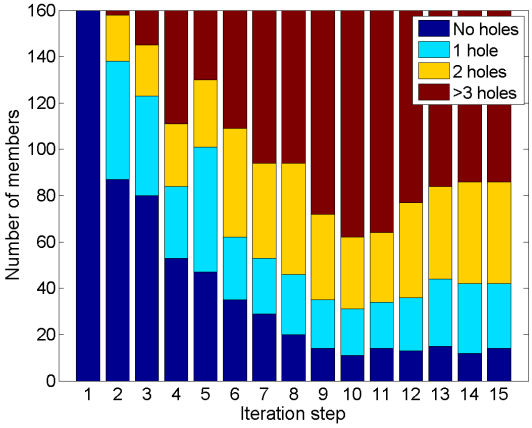
In figure 7.10 the frequency of the number of small holes among the members is shown for all measurement sets. Here a 'small hole' is defined as a cell where the aquitard is absent, surrounded by 8 cells where it is present. Almost all members in experiments 2 and 3 contain small holes.



(a) Set 1



(b) Set 2



(c) Set 3

Figure 7.10: Frequency of number of holes for measurement sets 1,2 and 3

Figure 7.11 studies the effect that holes have on the correctly estimated percentage. The results are only shown for measurement set 3, but the results for set 2 are similar. The percentage of cells that is correctly estimated by all members ($P = 1$) with 0 holes, is significantly better than the overall estimation that is bounded by the worst performer of the four groups. The other percentiles are in general also better. Only the 0.2 estimate is clearly worse. This can be explained by the smaller variance. The members without holes are those with the smallest overall percentage of aquitard presence. Therefore the variance among those members will be smaller than the variance of the entire ensemble. The result of this limited variance is that all percentiles are close to each other, both increasing the $P = 1$ performance as well as deteriorating the $P = 0.2$ performance.

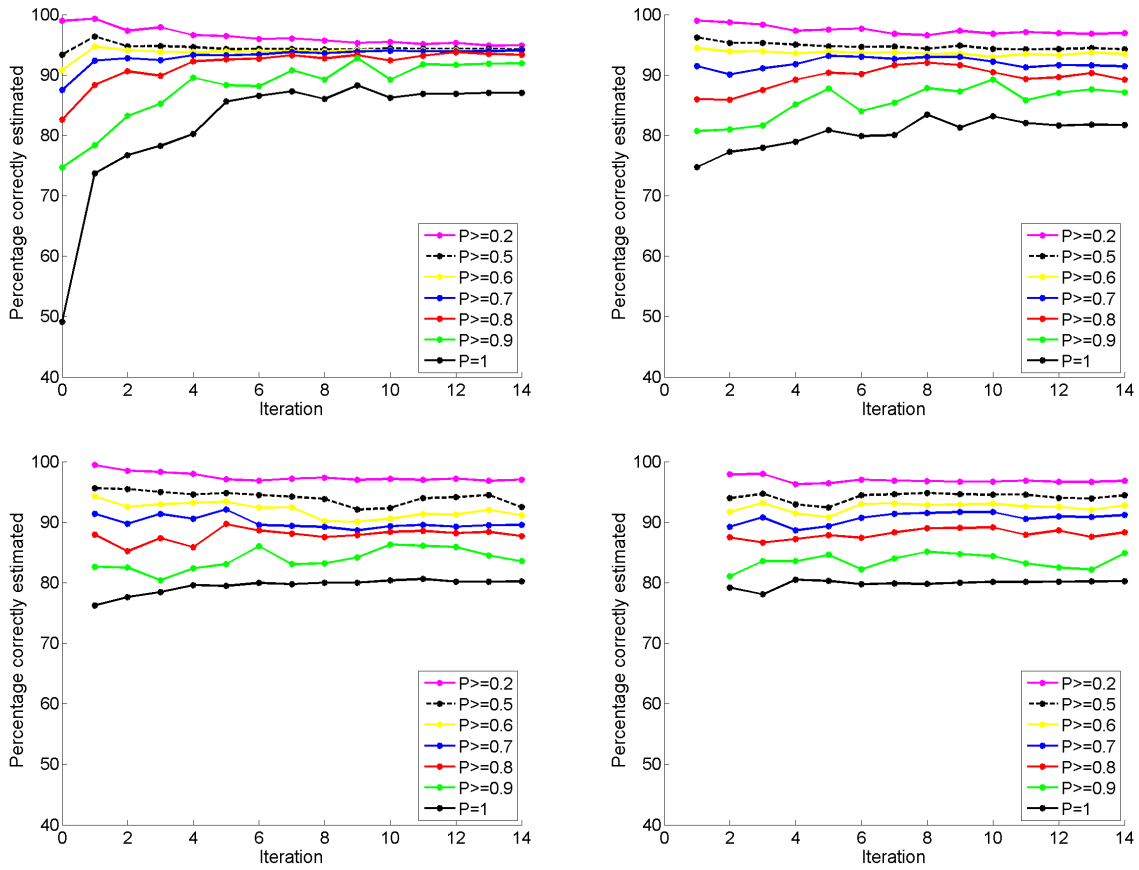


Figure 7.11: Percentage plot for measurement set 2 for subgroups distinguished by the number of small holes.

7.4 Analysis and discussion

In the previous subsection we presented the results of the EnKF-GMM applied to a synthetic model. The following problems occur:

1. Straight lines in the probability plots for the extent of the aquitard.
2. Updated probability plots are unstable.
3. Filter divergence.

In this subsection we will discuss the above problems and provide possible explanations and suggestions for future research. Furthermore, we will discuss the contribution of the EnKF-GMM to the research goals.

7.4.1 Straight lines in the probability plots on the extent of the aquitard

The first research goal is to estimate the extent of the aquitard. The EnKF-GMM shows a decent performance in doing so. Large uncertainties (probabilities close to 0.5) only occur at locations where they are expected. That is, at locations where no measurement are available nearby, or close to the boundaries of the domain where the hydraulic heads are prescribed. In figure 7.4 straight lines occur in the estimates. These lines are a result of the applied localisation techniques during the computation of the posterior probabilities based on the likelihood computation. Increasing the ensemble size or application of a more sophisticated localisation technique, can help to further improve the estimates. For instance, it may be better to introduce some sort of damping effect, that assigns a decreasing weight to measurements located further away. In this way, the transitions are less abrupt. Also, the hydraulic head-thickness correlation near the edges of the aquitard is not likely to be the same in all directions. The selection of a better localisation technique, however, is left for future research.

7.4.2 Updated probability plots are unstable

At later time steps, the probabilities on aquitard presence should stabilise. This means that the number of transfers should decrease to 0. This happens to some extent, but also at later time steps quite some transfers occur. Figure 7.12 shows the number of transfers made for each member per iteration step in the first experiment. The number of transfers does not diminish to 0. An explanation for this, is that the hydraulic head data does not contain enough information to be conclusive. The instability of the number of transfers forces us to question the accuracy of the estimates. In the results we have only presented the probabilities after the final update, but during the last five updates variations of 0.2 in probabilities occurred in most cells with a probability unequal to 0 or 1. In experiments 2 and 3 the number of transfers decreases further, although in neither of the experiments the

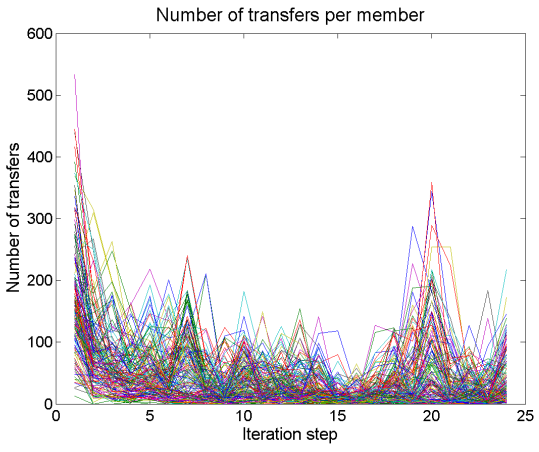


Figure 7.12: Number of transfers made in EnKF-GMM with measurement set 1.

number of transfers goes all the way to zero.

The large amount of transfers that remains, can be explained by the occurrence of holes in the aquitard. In the results we have seen how holes negatively influence all estimates. If a hole occurs, then the output of the groundwater flow model becomes insensitive to the component type of the neighbouring cells. This insensitivity can cause additional transfers at the later time steps. Although the first experiment shows the least amount of holes, the number of transfers is largest there. This is the result of many cells sharing the same single hydraulic head estimation as input for the likelihood computation. If a hole occurs near such an important location (in the sense that a lot of cells, also beyond the range of 2km, depend on it), then the insensitivity involves a lot of cells. An idea to resolve this issue, is to manually switch the component type of those holes to 'present' directly when the holes occur in the indicator resampling algorithm. In this way, the holes will be clogged. Further research must show whether this improves the estimates.

7.4.3 Filter divergence

Compared to the regular EnKF, the estimation of the mode of the EnKF-GMM is much better. After considering all results, we can conclude that a bimodal distribution much better fits the general shape of the the probability distribution of thicknesses than a unimodal log-normal distribution. Nonetheless, the thickness errors do hardly diminish. Again filter divergence occurs and the EnKF-GMM converges to an incorrect solution. We have already seen this behaviour in experiment A and B of the regular EnKF. Therefore this was to be expected, at least for measurement sets 1 and 2. However, filter divergence also occurs for set 3. In this case, enough informative hydraulic head observations may be available, but the occurrence of holes in the aquitard stimulates insensitivity of the hydraulic heads to thicknesses. It thereby also stimulates the filter divergence. Therefore, removing the holes might also improve the results with respect to the filter divergence.

No divergence can occur in the experiment where the EnKF-GMM is replaced by a simple recalculation of the probabilities on presence/absence. Therefore, although the errors in the hydraulic heads are significantly larger than in the other experiments, these results may be most valuable in practice.

7.4.4 Potential of the EnKF-GMM

In the remainder of this discussion we will ignore the filter divergence for a moment and discuss the value of the obtained posterior distributions, under the assumption that the state where the EnKF-GMM converges to, is the correct state. In other words, we will try to discuss the potential of the EnKF-GMM as a method on itself, instead of depreciating it based on a weak problem definition. In the EnKF-GMM, outliers occurred. In the results, these outliers were truncated. In the EnKF, there also were problems involving too large values, but no truncation was done. The difference is that in the application of the EnKF, the entire distribution was too wide in too many cells, resulting in large thicknesses in a lot of cells and members. A standard deviation could not be defined since it exceeded a magnitude of hundreds of metres. Here similar errors occur, but it only happens in few cells and it is better characterised as unrealistic outliers than a situation in which the entire PDF is physically unreasonable. Furthermore, we should note that roughly half of the outliers are situated in the interval $(-\infty, -9]$. These outliers largely influence the magnitude of the variance, and hence the updates, while AZURE is insensitive to

them. Furthermore, if the thickness reaches this insensitive region, we should consider switching its component type to absent. For example, in scenario 2 in a specific cell, there is one member with component type 'present', that has a logarithmic thickness of -155. This outlier boosts the value of the (co)variance that is used for the computation of the mean, median and mode, and for the EnKF-GMM updates. In the results we have truncated the outliers to obtain better PDFs, but it would be better to already intervent during the EnKF-GMM. In that way, the influence the outliers have on the updates can be reduced. This can either be done by a truncation, or by manually switching the component type.

All errors that occur in the EnKF-GMM are related to the transfers of component types. There are two potential sources for the errors. First of all, a small typo in the Matlab code of the EnKF-GMM, and secondly the decision to also update the artificial distribution with an EnKF. The typo involves the computation of the artificial state vector. In the theoretical formulation of the EnKF-GMM in section 6, we have provided arguments for computing the artificial state vector in the most straightforward way. However, initially we intended to apply Cholesky transformations to the error vectors in a way similar to what was done by *Dovera and Della Rossa* (2011), but then extended to a cell-by-cell approach. In this approach a "Cholesky decomposed matrix-inverse Cholesky decomposed matrix-vector product", is included in the computation of the artificial state vector. However, in the implementation in Matlab this matrix-matrix-vector product was replaced by a point wise vector-vector product. In the point wise vector product, the first vector contains the sum of all elements in a row of the matrix product of the Cholesky decomposed matrix-inverse Cholesky decomposed matrix. What was implemented in the experiments is therefore a mixture somewhere between working with Cholesky decomposed matrices and simply translating the values with the difference in means.

Although the typo will almost certainly ruin the preservation of any correlation structure, it is not likely to be the main cause for the outliers. In the correct way of taking the matrix-vector product, a weighted sum is taken of all thicknesses in the area. Due to the typo here, we take the total weight times the value of the cell itself. As the thicknesses are supposed to be correlated, the cells are alike, and there is no reason why an outlier should be created. This explanation is still incomplete since also the hydraulic heads are part of the weighted sum. A consultation of the resulting matrices, however, has shown that those weights are negligible compared to those of the covariances between thicknesses. Secondly, we should note that this unfortunate error plays a negligible role during the first updates. As the correlation structure of both component types is initially the same, the matrix product is approximately the unitary matrix at the first iteration step. All deviations from the unitary matrix are caused by the finite sample size that is used to compute the covariance matrices. In case it would be the exact unitary matrix, the result is exactly the same whether we would have used the Cholesky decompositions, the translation of means or the result obtained with the typo (with the small influence of the covariance of head data left aside). However, as the filter continues, the matrix will not be unitary anymore and the errors will increase.

After interpretation of the results and analysing the product of the Cholesky-inverse Cholesky decomposed matrix, we would advise not to apply the Cholesky decompositions in this EnKF-GMM per cell. The ensemble sizes are too small and errors occur as a result of spurious correlations. Here they occur due to the strange sums of elements in a row, but if the matrix product was done correctly the results would still have been incorrect. Localisations only make sense in case we are still working with a reasonable ensemble size and not if the ensemble size is only 5 or 6. As mentioned before, the best solution would probably be to use a geostatistical

method to estimate the transferred cells and after each estimation do an additional confirming step for the computation of hydraulic heads. However, this is computationally expensive and additional research must show how large the benefits are.

The other possible cause for the outliers involves the decision of updating the artificial distributions. Additional simulations where no updates are performed should be done, but it is likely that not updating will improve the results.

7.4.5 Contribution to the research goals

Overall, we can conclude that the performance of the EnKF-GMM is much better than the regular EnKF. A bimodal distribution describes the uncertainties of the thicknesses much better. Furthermore, the extent of the aquitard can now be estimated in a straightforward manner. Similarly as in the regular EnKF, divergence occurs. This indicates that we should not aim at constructing entire probability distribution functions for the thicknesses, but only try to assign a weight to both modes. In general, the posterior most likely estimates are better than the prior most likely estimates, but no confidence level can be attached to the estimates as a result of the filter divergence. From a technical point of view, the EnKF-GMM could be further adjusted. The transition of the component types is somewhat troublesome, as it seems difficult to keep the correlation structure near the aquitard's edge from fading. Suggestions for a fine-tuning of the EnKF-GMM have been made, but additional research most show whether it will improve the results.

8 Conclusion and recommendations

In this thesis we have studied the possibility to use the calibration of the groundwater flow model AZURE to assess the uncertainty of the underlying geometry of the subsurface. The underlying geometry is implicit in the transmissivities of aquifers and resistances of aquitards, two dependent parameters in the groundwater flow model. In the current calibration routine their dependence is overlooked. A reformulation was proposed to include these dependencies.

After that we looked for a suitable calibration method that could provide us with probability distribution functions for the hydraulic parameters and indirect the thicknesses of aquitards and aquifers. Lots of similarities with data assimilation problems in the reservoir engineering were expected. Particularly methods from that field were therefore considered for application here. In reservoir engineering, many of the recently applied data assimilation techniques involve the application of an EnKF in combination with a parametrisation technique. Due to the 2.5 dimensionality of the AZURE model instead of a 2D or 3D groundwater flow model, none of the parametrisation techniques seemed helpful for our case. Therefore an EnKF without any parametrisation was selected to first get familiar with the problem and to discover the key difficulties of the problem.

For the EnKF, two ensembles were generated with varying amounts of uncertainty in the extent and thickness of the aquitard. Two different measurement sets also containing a varying amount of information were used for the updates. The EnKF was able to show some improvement of structures, but due to the large variance on a logarithmic scale, the gathered probability distributions were found to be physically unrealistic. A test case with an exactly known extent of the aquitard showed promising results with respect to the mean/median/most likely estimate. However, filter divergence was observed as a result of a weak problem definition. Apparently multiple configurations of the thickness can result in the same set of simulated hydraulic heads. The EnKF converges to one of these states instead of preserving a rather large range of uncertainties. Because of this filter divergence, the obtained probability distributions are not useful. However, in this situation with a known aquitard extent, the EnKF does not produce physically unrealistic state variables. This improved performance can be contributed to the absence of large variances in the thickness.

The influence of the large variances was eliminated using an EnKF-GMM. In this modified version of the EnKF, the posterior distribution of the thickness is allowed to be bimodal. One mode consists of realisations that assume that the aquitard is present, while the other mode contains realisations that assume absence in a specific cell. A new prior ensemble with a varying aquitard extent was tested in combination with three different measurement sets. Estimations about the extent of the aquitard were successfully made. Depending on the used measurement sets, a varying number of cells showed an estimation with a probability other than 0 or 1. The number of severe errors, e.g. wrong estimations with a large probability, remained modest.

The posterior probability distributions of the thickness showed less promising results. Filter divergence also occurred in the EnKF-GMM. The problem is not in the algorithm used, but in the definition of the problem. We emphasise that we have noted multiple times how limited the variances in hydraulic heads are. With respect to the true measurement set we can conclude that it is not suited to obtain useful estimates in all cells. In the sensitivity analysis it was shown that the influence of aquitard thickness on hydraulic heads is largest in areas where either the aquitard or the surrounding aquifers are thin. The true set lacks observation locations in these

areas. If in the future more locations from these areas are added to the existing set it may become possible to extract better estimates for the mean/median/mode. However, the problem of filter divergence is unlikely to disappear, since this occured for all three measurement sets.

From a technical point of view still several improvements can be made in the EnKF-GMM. The transfers of aquitard presence from/to absence will always remain a source of errors. However, the errors are likely to signficantly reduce in case an iterative EnKF with extra confirming steps and resampling by the application of geostatistical methods is performed. However, we would roughly need the double amount of computational power for this. If this is not a problem, then it is likely that the errors can be reduced to an acceptable level.

The posterior probability distributions of the aquitard were estimated to support a larger project that involves a complete feedback of AZURE to the different units of REGIS II. To really make a contribution to that project it is necessary to extend the data assimilation framework to include the uncertainties of all aquitards, and to work with the real data. The extension of the state vector by including the thicknesses of the other aquitards and additional measurement locations should not give any problems. However, securing the dependencies between layers is less straight forward. Vertical correlations will play a role and both the method to construct a prior ensemble as well as the component type resampling algorithm will have to be extended. It is recommended to let an experienced geologist have a look at both algorithms to help ensure spatial correlations and to include the true borehole data in the prior ensemble.

An application of the EnKF-GMM in a real scenario also involves the model error. In this thesis the AZURE model was a perfect representation of the true process, but in reality a model will always be a simplification. Therefore we need to include a non-zero model error to the model. Additional research is required to investigate the influence of this error. On the one hand, it may deteriorate the solution as updates can be partially based on model noise; on the other hand it may mitigate the outliers and thereby improve the estimates.

Furthermore we should note that although we have checked multiple scenarios, all test cases were based on the same aquitard. The interplay of thicknesses and hydraulic heads may be entirely different in areas with an aquitard of different size or average thickness. The measurement set in the area here is not sufficient, but if less boreholes and more hydraulic head measurements are available in other areas, this might be different. Therefore multiple (very) different test cases should be considered to really map in which situations hydraulic head data can improve the estimates of the underlying parameters.

Finally, we conclude that with the EnKF-GMM it may be possible to construct a complete probability distribution function for an aquitard, or at least an estimate on the extent of the aquitard. However, before this is possible additional research is required in several areas to fine-tune the method and to determine possible areas of application.

A Representer method

The representer method is a variational data assimilation method that was first introduced by Bennett in 1992. The method applies a parametrisation technique that reduces the number of unknown state variables to the number of measurements which is usually significantly less. A short description of the representer method is now given:

The representer method minimises an objective function that is derived from the maximum a posteriori estimator or Bayesian maximum likelihood estimator. Correspondingly to this minimisation problem a set of corresponding Euler-Lagrange equations can be derived. The representer method iteratively solves these Euler-Lagrange equations. First an "optimal" parametrisation (meaning that the solution of the parametrised problem is equal to the solution of the full-size problem) is applied by the introduction of representer solutions. This reduces the number of independent parameters that have to be estimated to the number of measurements. A beneficial side-effect of the definition of this parametrisation, is the decoupling of the Euler-Lagrange equations such that the equations can be solved sequentially or even in parallel.

The parametrisation involves the expansion of the unknown state variables in a finite sum of representer functions and corresponding coefficients. These representer functions can in practice be seen as influence functions of the measurements on the state and parameter estimates. They predict for each measurement how the model parameters and model errors, and as a consequence the state variables, should change, while minimising increments of the objective function. The superposition of all these changes summed to a forward model prediction of the state, forms a new estimate. So the weighted sum of all adjoint variable representer functions, determine a new estimate for the adjoint variables. Similarly, an update for the model parameters can be obtained by adding the weighted sum to the prior of the parameters. For the state variables an additional correction term is necessary since the representer expansions assume a linear relationship between the states and the parameters, whereas this relationship is known to be nonlinear in reality. For small intervals around the point of linearisation, this would give a good estimate, but in the representer expansion the linearisation is determined around the prior means instead of around the last estimates. To control for these errors, a correction term is added in the update of the state variables.

Valstar (2001) derived formulae for these representer functions and their coefficients for the AZURE model. He showed that the representers equal the linearized cross-covariances between measurement predictions and state variables or model parameters. He there also proposed a computationally more efficient way to compute these functions.
Relevant to the research done here is to what extent the representer method gives accuracy estimates of the model parameters. *Valstar* (2001) showed that the approximate covariance of the parameters is given by

$$P_{\alpha_{k,l}}^{posterior} = P_{\alpha_{k,l}}^{prior} - \Psi_{k,p} \left((M(X_{representer}) + P_v)^{-1} \right)_{p,q} \Psi_{l,q} \tag{25}$$

where $P_{\alpha_{k,l}}$ is the (k,l) entry of the covariance matrix of the parameters,
$\Psi_{k,p}$ is the parameter representer for measurement p,
$M(\,\cdot\,)$ is the measurement operator,
$X_{representer}$ are the representer solutions of the state variables,
P_v is the covariance matrix of the measurement errors

B Probabilistic parameterization approach

In this Appendix, a summary of the proabablitstic parameterization approach of *Sebacher et al.* (2013) is given. The approach is extended to a 3D scenario, where three facies types can occur, each type representing a different layer.

Let (i, j, k) be an arbitrary cell location and define $A_l^{i,j,k}$ as the event in which in the grid the lithotype l occurs. Here $l \in \{1, 2, 3\}$ denotes respectively the lithotype of the upper aquifer, the aquitard and the lower aquifer. If at position (i, j, k) lithotype 1 occurs, then $P(A_1^{i,j,k}) = 1$ and $P(A_2^{i,j,k}) = P(A_3^{i,j,k}) = 0$. For a deterministic field, such as the true field, one can in this way produce three discrete fields corresponding to the different lithotypes, containing only ones and zeros. In reality, only at wells the true lithotype is known with probability 1 and if we move further away this certainty will dissipate with a certain correlation, resulting in continuous fields which are referred to as probability fields.

To model these probability fields, two Gaussian random fields y_1 and y_2 are generated and projected on $[0, 1]$ using a projection function. This function is given by $\phi_m : R \to [0, 1]$,

$$\phi_m(t) = \begin{cases} \frac{-|t|}{m} + 1 & \text{if } t \in [-m, m] \\ 0 & \text{if } |t| > m \end{cases}$$

where m is a truncation parameter. Figure B.1 shows the projection function.

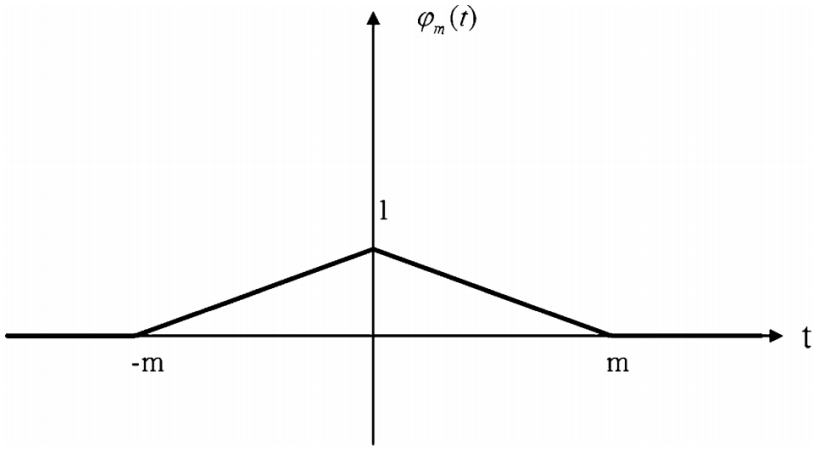


Figure B.1: Truncation map of *Sebacher et al.* (2013)

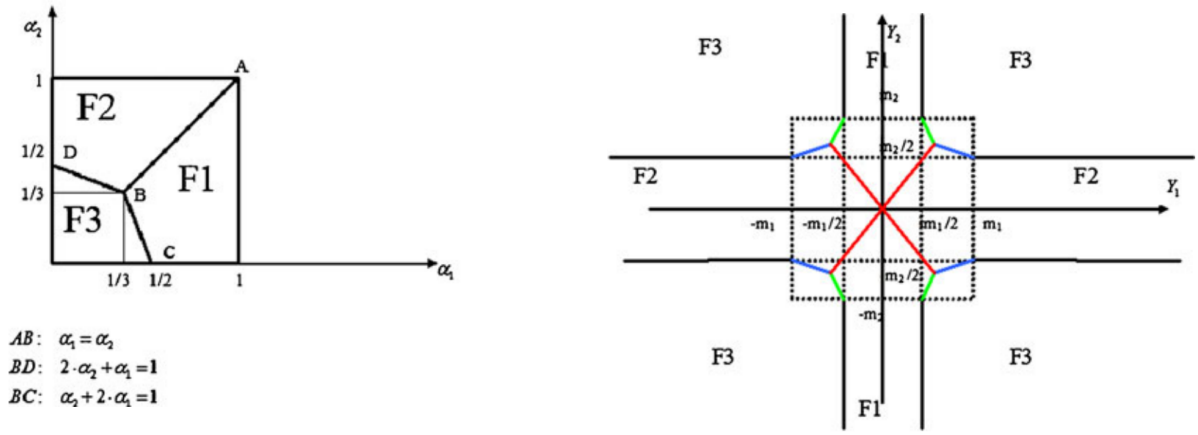
The probability fields α_1 and α_2 for the first two lithotypes are then defined by $\alpha_1 = \phi_{m_1}(y_1)$ and $\alpha_2 = \phi_{m_2}(y_2)$. The third probability field follows from the law of total probability by defining for each grid cell:

$$\alpha_3^{i,j,k} = \begin{cases} 1 - \left(\alpha_1^{i,j,k} + \alpha_2^{1,j,k}\right) & \text{if } 0 \leq \alpha_1^{i,j,k} + \alpha_2^{i,j,k} \leq 1 \\ 0 & \text{if } \alpha_1^{i,j,k} + \alpha_2^{i,j,k} > 1 \end{cases}$$

We note that the situation where $\alpha_1^{i,j,k} + \alpha_2^{i,j,k} > 1$ can indeed occur. In case this happens the sum of probabilities should be scaled back to 1.

The interpretation of the fields as probability on the occurrence of facies types now provides a natural way to assign a facies to a certain location (i, j, k). At a certain location (i, j, k) we assign lithotype $l \in \{1, 2, 3\}$ if $\alpha_l^{i,j,k} = \max\{\alpha_n^{i,j,k}, n = 1, 2, 3\}$ (with the convention that in case of equal probabilities the lowest lithotype number is chosen).

Having defined this decision rule, the rock-type rule map (similar as to figure 3.3) follows from this. As an intermediate step, figure B.2a shows the rock-type-rule map for (α_1, α_2) and B.2b for y_1, y_2. The threshold values of the truncation parameters m_1 and m_2 are not estimated as deterministic values but are considered uncertain with a prior mean and uncertainty range which should be chosen sufficiently large to ensure that the real proportions are in the interval. Then a correlation structure (variogram) needs to be determined and by applying an appropriate simulation technique to the borehole measurements the initial Gaussian random fields are created. Afterwards an EnKF can be applied to assimilate the head data.



(a) Truncation map for the probability fields

(b) Truncation map for the Gaussian Random fields

Figure B.2: Rock-type-rule mappings. From: *Sebacher et al.* (2013)

C Derivation of Kalman Filter equations

In the main text the Kalman Filter equations are enlisted. In this appendix we will provide the proof to equations 15-17. Several different derivations of these formulae have been published. Here a derivation that respects Kalman's original concept is given, using orthogonal projections on the vector space of random variables. (*Kalman* (1960), *Terejanu* (online))

Let $Y(t)$ be the vector space that contains all linear combinations of observations $\mathbf{y}(t_0), ...\mathbf{y}(t)$ with real coefficients; that is

$$Y(t) = \left\{ \mathbf{z} \;\middle|\; \mathbf{z} = \sum_{i=t_0}^{t} a_i \mathbf{y}_i,\ \forall a_i \in \mathbb{R} \right\}$$

Kalman's proof is based on the following theorems involving orthogonal projections (which are not proven here, *Kalman* (1960)):

1. Let $\{\mathbf{x}_t\}$, $\{\mathbf{y}_t\}$ be random processes with zero mean. We observe $\mathbf{y}_{t_0}, ..., \mathbf{y}_t$. If either 1) the random processes $\{\mathbf{x}_t\}$, $\{\mathbf{y}_t\}$ are Gaussian; or 2) the optimal estimate is restricted to be a linear function of the observed random variables and $L(\epsilon) = \epsilon^2$ a loss function that is positive and a non-decreasing function of the estimation error $\epsilon = \mathbf{x}_t - \hat{\mathbf{x}}$; then

$$\begin{aligned} \mathbf{x}^*_{t_1} &= \text{the optimal estimate of } \mathbf{x}_{t_1} \text{ given } \mathbf{y}_{t_0}, ..., \mathbf{y}_t \\ &= \text{the orthogonal projection } \bar{\mathbf{x}}_t \text{ of } \mathbf{x}_{t_1} \text{ on } Y(t). \end{aligned}$$

2. Two vectors $\mathbf{u}, \mathbf{v} \in Y(t)$ are orthogonal if their correlation is zero.

3. Any vector $\mathbf{x}$ can be uniquely decomposed in two parts $\bar{\mathbf{x}} \in Y(t)$ and $\tilde{\mathbf{x}} \perp Y(t)$ such that $\mathbf{x} = \bar{\mathbf{x}} + \tilde{\mathbf{x}}$

At time t we may assume that $Y(t-1)$ is known and that $\mathbf{y}_t$ is measured. For $\mathbf{y}_t$ there exists a unique decomposition $\mathbf{y_t} = \bar{\mathbf{y}}_\mathbf{t} + \tilde{\mathbf{y}}_\mathbf{t}$ such that $\bar{\mathbf{y}}_\mathbf{t} \in Y(t-1)$ and $\tilde{\mathbf{y}}_\mathbf{t} \perp Y(t-1)$. The component $\tilde{\mathbf{y}}_\mathbf{t}$ generates a linear manifold $Z(t)$. By definition, $Z(t)$ is a linear manifold of $Y(t)$ if it is a non-empty subset of $Y(t)$ and there exists an element $y \in Y(t)$ such that $Z(t)+y = \{z+y | z \in Z(t)\}$ is a vector subspace of Y(t). Here, $Z(t)$ is not empty, as $\tilde{\mathbf{y}}_\mathbf{t} \in Z(t)$ and take $y = \bar{\mathbf{y}}_\mathbf{t}$; then $\bar{\mathbf{y}}_\mathbf{t} + \tilde{\mathbf{y}}_\mathbf{t}$ such that $\bar{\mathbf{y}}_\mathbf{t} \in Y(t-1) \subseteq Y(t)$. Furthermore, $Y(t) = Y(t-1) \cup Z(t)$ and every vector in $Z(t)$ is orthogonal to every vector in $Y(t-1)$.

At each time step t, denote $\mathbf{x}^f_t$ as the forecast-state vector and $\mathbf{x}^a_t$ as the analysed state vector. At time t we can assume that $\mathbf{x}^a_{t-1}$ is known and therefore $\mathbf{x}^f_t$ can be computed by forecast equation 13. The new updated optimal state $\mathbf{x}^a_t$ is then given by:

$$\begin{aligned} \mathbf{x}^a_t &= E[\mathbf{x}_t | Y(t)] \\ &= E\left[\mathbf{x}_t | Y(t-1)\right] + E\left[\mathbf{x}_t | Z(t)\right] \\ &= \mathbf{x}^f_t + E\left[\mathbf{x}_t | Z(t)\right] \end{aligned} \tag{26}$$

Now assume that the last term can be rewritten as a linear operation on the random variable $\tilde{\mathbf{y}}(t)$. This introduces the Kalman gain K_t:

$$E[\mathbf{x}_t | Z(t)] = K_t \tilde{\mathbf{y}}_t \tag{27}$$

where $\mathbf{y}_t = \tilde{\mathbf{y}}_t + \bar{\mathbf{y}}_t$. Note that $\bar{\mathbf{y}}_t$ is the orthogonal projection of $\mathbf{y}_t$ on $Y(t-1)$. By theorem (1) it follows that $\bar{\mathbf{y}}_t$ is the optimal estimate of $\mathbf{y}_t$ given $Y(t-1)$. It follows that:

$$\begin{aligned}
\bar{\mathbf{y}}_t &= E[\mathbf{y}_t \mid Y(t-1)] \\
&= E[H_t\mathbf{x}_t + \mathbf{v}_t \mid Y(t-1)] \\
&= E[H_t\mathbf{x}_t \mid Y(t-1)] + E[\mathbf{v}_t \mid Y(t-1)] \\
&= H_t\mathbf{x}_t^f
\end{aligned} \tag{28}$$

where the first equality follows by substitution of equation 12. The last term vanishes since $\mathbf{v}_t$ is independent of $Y(t-1)$ and has mean zero. Substituting the result of 28 into the expression for $\mathbf{x}_t^a$ yields equation 14:

$$\mathbf{x}_t^a = \mathbf{x}_t^f + K_t\left(\mathbf{y}_t - H_t\mathbf{x}_t^f\right) \tag{29}$$

Next we need to find an explicit expression for the Kalmain Gain. For this we first consider the estimation errors $\tilde{\mathbf{x}}_t^f$ (forecast) and $\tilde{\mathbf{x}}_t^a$ (analysed) and covariance matrices P_t^f(forecast) and P_t^a(analysed). For the Kalman Filter it is assumed that M is a linear model. Therefore it can be represented by a matrix M.

Forecast error:

$$\begin{aligned}
\tilde{\mathbf{x}}_t^f &= \mathbf{x}_t - \mathbf{x}_t^f \\
&= M_t\mathbf{x}_{t-1} + \mathbf{w}_{t-1} - M_t\mathbf{x}_{t-1}^a \\
&= M_t\tilde{\mathbf{x}}_{t-1}^a + \mathbf{w}_{t-1}
\end{aligned} \tag{30}$$

Error of analysed state:

$$\begin{aligned}
\tilde{\mathbf{x}}_t^a &= \mathbf{x}_t - \mathbf{x}_t^a \\
&= \mathbf{x}_t - \left(\mathbf{x}_t^f + K_t(\mathbf{y}_t - H_t\mathbf{x}_t^f)\right) \\
&= \mathbf{x}_t - (I - K_tH_t)\mathbf{x}_t^f - K_t\mathbf{y}_t \\
&= \mathbf{x}_t - (I - K_tH_t)\mathbf{x}_t^f - K_t\left(H_t\mathbf{x}_t + \mathbf{v}_t\right) \\
&= (I - K_tH_t)(\mathbf{x}_t - \mathbf{x}_t^f) - K_t\mathbf{v}_t \\
&= (I - K_tH_t)\left(M_t\mathbf{x}_{t-1} + \mathbf{w}_{t-1} - M_t\mathbf{x}_{t-1}^a\right) - K_t\mathbf{v}_t \\
&= (I - K_tH_t)\left(M_t\tilde{\mathbf{x}}_{t-1}^a + \mathbf{w}_{t-1}\right) - K_t\mathbf{v}_t
\end{aligned} \tag{31}$$

Forecast error covariance matrix:

$$\begin{aligned}
P_t^f &= E\left[\tilde{\mathbf{x}}_t^f(\tilde{\mathbf{x}}_t^f)^T\right] \\
&= E\left[\left(M_t(\mathbf{x}_{t-1}-\mathbf{x}_{t-1}^a)+\mathbf{w}_{t-1}\right)\left(M_t(\mathbf{x}_{t-1}-\mathbf{x}_{t-1}^a)+\mathbf{w}_{t-1}\right)^T\right] \\
&= E\left[M_t(\mathbf{x}_{t-1}-\mathbf{x}_{t-1}^a)\left(M_t(\mathbf{x}_{t-1}-\mathbf{x}_{t-1}^a)\right)^T\right] + E\left[M_t(\mathbf{x}_{t-1}-\mathbf{x}_{t-1}^a)\mathbf{w}_{t-1}^T\right] + \\
&+ E\left[\mathbf{w}_{t-1}M_t(\mathbf{x}_{t-1}-\mathbf{x}_{t-1}^a)^T\right] + E\left[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T\right] \\
&= E\left[M_t\tilde{\mathbf{x}}_{t-1}^a\left(M_t\tilde{\mathbf{x}}_{t-1}^a\right)^T\right] + E\left[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T\right] \\
&= M_tE\left[\tilde{\mathbf{x}}_{t-1}^a{\tilde{\mathbf{x}}_{t-1}^a}^T\right]M_t^T + E\left[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T\right] \\
&= M_tP_{t-1}^aM_t^T + Q_{t-1}
\end{aligned} \tag{32}$$

Analysed error covariance matrix:

$$\begin{aligned}
P_t^a &= E\left[\tilde{\mathbf{x}}_t^a\tilde{\mathbf{x}}_t^{aT}\right] \\
&= E\left[\left((I-K_tH_t)\left(M\tilde{\mathbf{x}}_{t-1}^a+\mathbf{w}_{t-1}\right)-K_t\mathbf{v}_t\right)\left((I-K_tH_t)\left(M_t\tilde{\mathbf{x}}_{t-1}^a+\mathbf{w}_{t-1}\right)-K_t\mathbf{v}_t\right)^T\right] \\
&= (I-K_tH_t)E\left[M_t\tilde{\mathbf{x}}_{t-1}^a(M_t\tilde{\mathbf{x}}_{t-1}^a)^T+\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T\right](I-K_tH_t)^T + K_tE[\mathbf{v}_t\mathbf{v}_t^T]K_t^T \\
&= (I-K_tH_t)\left(M_tE\left[\tilde{\mathbf{x}}_{t-1}(\tilde{\mathbf{x}}_{t-1}^a)^T\right]M_t^T+E\left[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T\right]\right)(I-K_tH_t)^T + K_tR_tK_t^T \\
&= (I-K_tH_t)\left(M_tP_{t-1}^aM_t^T+Q_{t-1}\right)(I-K_tH_t)^T + K_tR_tK_t^T \\
&= (I-K_tH_t)P_t^f(I-K_tH_t)^T + K_tR_tK_t^T
\end{aligned} \tag{33}$$

To derive the Kalman Gain we will consider the the residual $\mathbf{x}_t - E[\mathbf{x_t}|Z(t)]$. Note that $E[\mathbf{x_t}|Z(t)]$ represents the optimal estimate of $\mathbf{x}_t$ given $Z(t)$ and hence by theorem 1 it is the orthogonal projection of x_t on $Z(t)$. Then, $\mathbf{x}_t - E[\mathbf{x_t}|Z(t)] \perp Z(t)$. Since $\tilde{\mathbf{y}}_t \in Z(t)$, also $\mathbf{x}_t - E[\mathbf{x_t}|Z(t)] \perp \tilde{\mathbf{y}}_t$. It then follows that:

$$\begin{aligned}
0 &= E[(\mathbf{x}_t - E[\mathbf{x}_t|Z(t)])\tilde{\mathbf{y}}_t^T] \\
&= E[(\mathbf{x}_t - K\tilde{\mathbf{y}}_t)\tilde{\mathbf{y}}_t^T] \\
&= E[\mathbf{x}_t\tilde{\mathbf{y}}_t^T] - K_tE[\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T]
\end{aligned}$$

Further, $\mathbf{x}_t$ can be decomposed in $\mathbf{x}_t = \bar{\mathbf{x}}_t + \tilde{\mathbf{x}}_t$ where $\bar{\mathbf{x}}_t \in Y(t-1)$. Any vector in $Y(t-1)$ is perpendicular to any vector in $Z(t)$ and thus it follows that $\bar{\mathbf{x}}_t \perp \tilde{\mathbf{y}}_t$. Furthermore, $\bar{\mathbf{x}}_t = E[\mathbf{x}_t|Y(t-1)] = \mathbf{x}_t^f$ and hence $\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_t^f$.

$$\begin{aligned}
0 &= E[(\bar{\mathbf{x}}_t + \tilde{\mathbf{x}}_t)\tilde{\mathbf{y}}_t^T] - K_tE[\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T] \\
&= E[\bar{\mathbf{x}}_t\tilde{\mathbf{y}}_t^T] + E[\tilde{\mathbf{x}}_t\tilde{\mathbf{y}}_t^T] - K_tE[\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T] \\
&= E[\tilde{\mathbf{x}}_t\tilde{\mathbf{y}}_t^T] - K_tE[\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T] \\
&= E[\tilde{\mathbf{x}}_t^f\tilde{\mathbf{y}}_t^T] - K_tE[\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T]
\end{aligned} \tag{34}$$

Now note that $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \bar{\mathbf{y}}_t = H(\mathbf{x}_t) + \mathbf{v}_t - H(\mathbf{x}_t^f) = H(\mathbf{x}_t - \mathbf{x}_t^f) + \mathbf{v}_t = H(\tilde{\mathbf{x}}_t^f) + \mathbf{v}_t$ and $\tilde{\mathbf{x}}_t^f = \mathbf{x}_t - \mathbf{x}_t^f = M(\mathbf{x}_{t-1}) + \mathbf{w}_{t-1} - M(\mathbf{x}_{t-1}^f) = M(\tilde{\mathbf{x}}_{t-1}^f) + \mathbf{w}_{t-1}$. Using these results, we can rewrite (34). First term:

$$\begin{aligned} E[\tilde{\mathbf{x}}_t^f \tilde{\mathbf{y}}_t^T] &= E[(M_t\tilde{\mathbf{x}}_{t-1}^f + \mathbf{w}_{t-1})(H_t\tilde{\mathbf{x}}_t^f + \mathbf{v}_t)^T] \\ &= E[(M_t\tilde{\mathbf{x}}_{t-1}^f) + \mathbf{w}_{t-1})(H_t(M_t\tilde{\mathbf{x}}_{t-1}^f + \mathbf{w}_{t-1}) + \mathbf{v}_t)^T] \\ &= E[M_t\tilde{\mathbf{x}}_{t-1}^f(\tilde{\mathbf{x}}_{t-1}^f)^T M_t^T H_t^T] + E[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T H_t^T] \\ &= (E[M_t(\tilde{\mathbf{x}}_{t-1}^f)(\tilde{\mathbf{x}}_{t-1}^f)^T M_t^T] + E[\mathbf{w}_{t-1}\mathbf{w}_{t-1}^T])H_t^T \\ &= P_t^f H_t^T \end{aligned} \tag{35}$$

Second term:

$$\begin{aligned} K_t E[\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t^T] &= K_t E[(H_t\tilde{\mathbf{x}}_t^f + \mathbf{v}_t)(H_t\tilde{\mathbf{x}}_t^f + \mathbf{v}_t)^T] \\ &= K_t\left(E[H_t\tilde{\mathbf{x}}_t^f(\tilde{\mathbf{x}}_t^f)^T H_t^T] + E[\mathbf{v}_t\mathbf{v}_t^T] + E[H_t\tilde{\mathbf{x}}_t^f\mathbf{v}_t^T] + E[\mathbf{v}_t(\tilde{\mathbf{x}}_t^f)^T H_t^T]\right) \\ &= K_t\left(H_t P_t^f H_t^T + R_t + 0 + 0\right) \end{aligned} \tag{36}$$

Substituting (35) and (36) in (34) yields:

$$0 = P_t^f H_t^T - K_t\left(H_t P_t^f H_t^T + R_t\right) \tag{37}$$

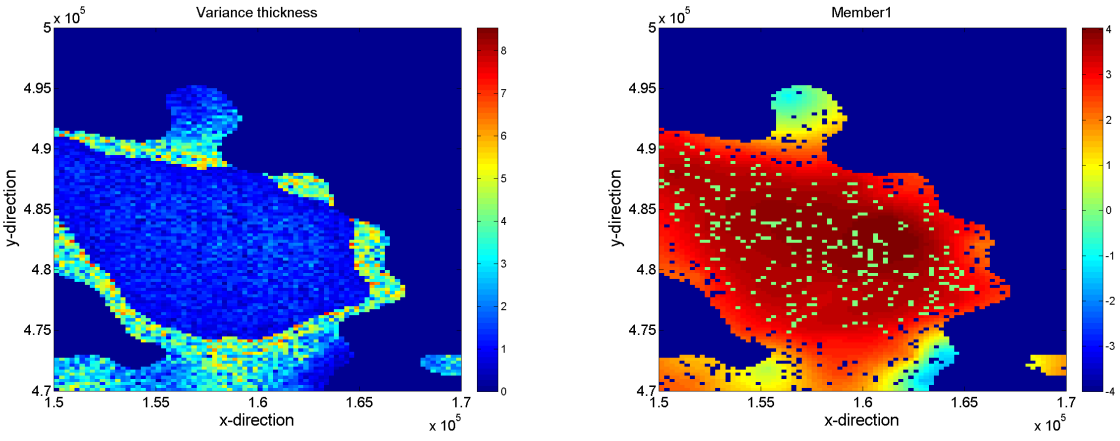
from which the Kalman gain follows:

$$K_t = P_t^f H_t^T\left(H_t P_t^f H_t^T + R_t\right)^{-1} \tag{38}$$

D Holes in the aquitard

In this thesis we have assumed that no holes occur in the reference scenario. In this Appendix we will show how holes in the aquitard influence the simulated hydraulic heads. To that use, an ensemble was constructed that takes the reference scenario of ensemble B as a starting point. The members were constructed by randomly selecting a thousand cells in the domain. If in the reference scenario the aquitard was thinner than 30 metres, a hole was created. Otherwise, the thickness was reduced to one metre. In case the aquitard was already absent, nothing was adjusted. The output is somewhat similar to what is simulated using the combination of SGS and SIS. SIS creates holes near the edges and SGS causes the thicknesses to locally vary to a large extent. Figure D.1a and D.1b show the variances in the ensemble thickness and the thickness of the first member on a logarithmic scale.

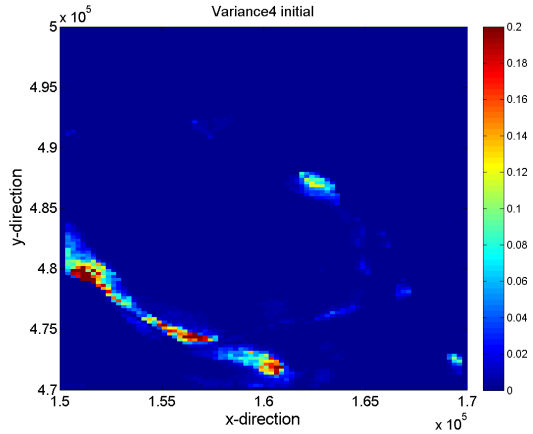
The ensemble was simulated forward in time for one year. The hydraulic head variances in the surrounding aquifers after this year are shown in figures D.2a and D.2b. It shows that the hypothesis that there are no hydraulic head variances left due to occurrence of holes, is incorrect. Large variances occur in multiple cells. However, the results do illustrate that it is easier to achieve the research goals if it is assumed that no holes can occur in the aquitard. Namely, there is a difference in variance of the simulated hydraulic heads between this ensemble and the others. The occurrence of the holes makes the hydraulic head patterns less smooth. At the locations of the holes and decreased thicknesses we can observe irregularities. Figures D.2c and D.2d show the hydraulic heads in the fifth aquifer for the reference situation of ensemble B and the first member with holes as defined in figure D.1b. Several dark spots can be observed in the left figure and transitions are more abrupt. The influence of the holes can be investigated by studying the difference in hydraulic heads generated with the first member and the reference case. Figure D.2e shows a close-up of this hydraulic head difference in a small part of the domain. In figure D.2f we have zoomed in on the aquitard thickness of the first member in the same area of the domain. It follows that the location of the irregularities in the hydraulic heads are directly related to the location of the holes.



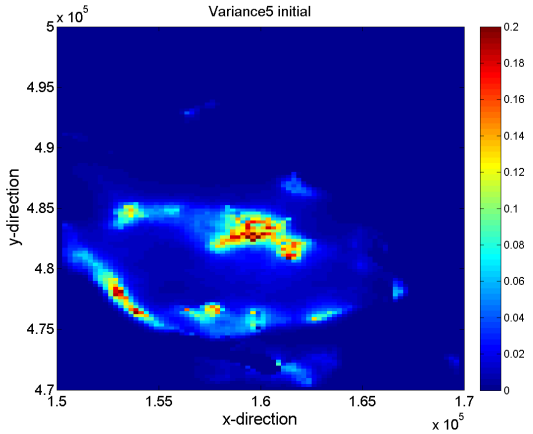
(a) Variance of thickness on a logarithmic scale

(b) Thickness of first member on a logarithmic scale

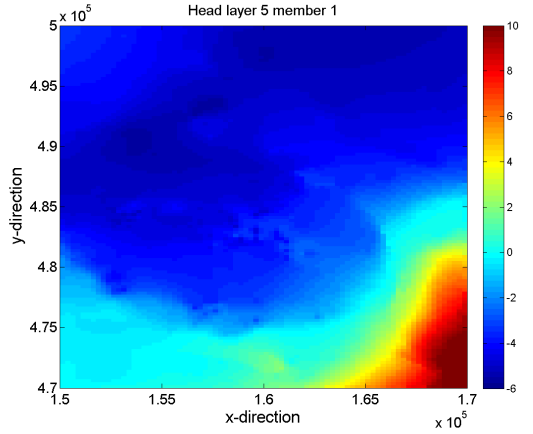
Figure D.1: Aquitard with holes (1)



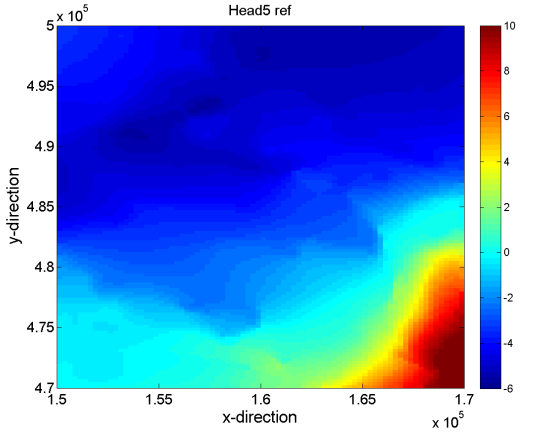
(a) Hydraulic head variances in the fourth aquifer [m^2]



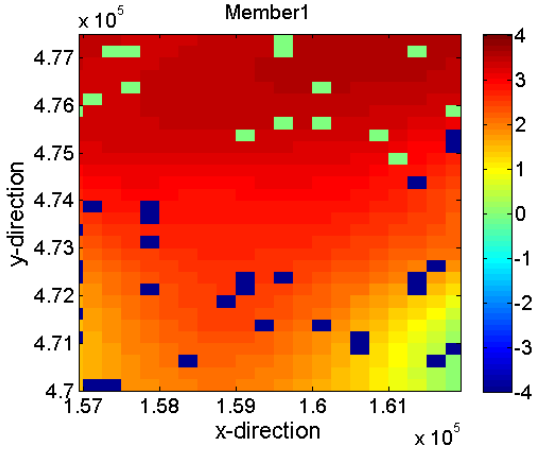
(b) Hydraulic head variances in the fifth aquifer [m^2]



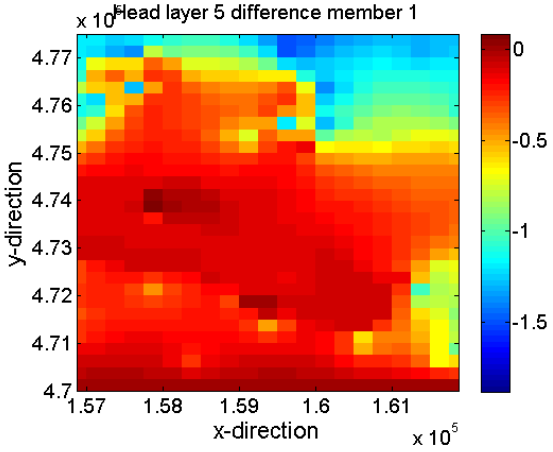
(c) Hydraulic heads in the fifth aquifer for the first member [m]



(d) Hydraulic heads in the fifth aquifer for the reference case [m]



(e) Close-up of aquitard thickness of the first member



(f) Close up of the difference in hydraulic heads between the first member and the reference case [m]

Figure D.2: Aquitard with holes (2)

E MAE and RMSE

To measure errors several different quantities are available. In this thesis, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used. The MAE is given by:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\mathbf{x}_i^{est} - \mathbf{x}_i\right| = \frac{1}{n}\sum_{i=1}^{n}|\tilde{\mathbf{x}}_i| \tag{39}$$

The RMSE is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^{est} - \mathbf{x}_i\right)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{x}}_i^2} \tag{40}$$

where x_i^{est} is the estimation for the i^{th} cell or member, and x_i the true state.

The MAE is a standard way to express the magnitude of errors, but it does not measure the variability of the error; it gives the first moment of the absolute error:

$$\mathrm{E}[|\tilde{\mathbf{x}}|] = \frac{1}{n}\sum_{i=1}^{n}|\tilde{\mathbf{x}}_i| = \mathrm{MAE} \tag{41}$$

The RMSE on the other hand, does provide information regarding the variability of the error. It is related to the standard deviation of the error in case of an unbiased estimator:

$$\mathrm{Var}[\tilde{\mathbf{x}}] = \mathrm{E}\left[\tilde{\mathbf{x}}_i - \mathrm{E}[\tilde{\mathbf{x}}]^2\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\mathbf{x}}_i - \mathrm{E}[\tilde{\mathbf{x}}]\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{x}}_i^2 = \mathrm{RMSE}^2 \tag{42}$$

where the before last equality only holds in case of an unbiased estimator. If the estimator is not unbiased, it still tells something about the variability of the errors.

The combination of the RMSE and MAE can be used to assess both the uncertainty and variance among the errors. In general, the following relation holds:

$$\mathrm{MAE} \leq \mathrm{RMSE} \leq \sqrt{n}\mathrm{MAE} \tag{43}$$

where MAE = RMSE in case all errors are equal and RMSE = $\sqrt{n}$MAE in case all errors but one are zero. (*Willmott and Matsuura* (2005)) This shows that the larger distance there is between the MAE and RMSE, the larger differences in error there is. *Willmott and Matsuura* (2005) make a strong case in showing that the RMSE on itself should not be used as an error measure to compare between models/experiments, so here we will solely use the RMSE to say something about the variance of the error.

F Cholesky transformations

F.1 Principle

Cholesky transformations can be used to correlate and decorrelate variables. Given a covariance matrix P, it can be factorised uniquely into a product $P = LL^T$, where L is a lower diagonal matrix having positive diagonal entries.

Consider the vector $X_k = \mathbf{x}_t^f - \mu_t^k$, where $\mathbf{x}_t^f$ is the state vector of members with component type k and μ_t^k the mean of these members. As the mean of X is 0 (unbiased error) it follows that that the covariance matrix P_k can be defined as:

$$P_k = E[X_k X_k^T]$$

As P_k is a covariance matrix it is theoretically known to be positive definite. Therefore a Cholesky decomposition can be performed to obtain L_k such that $P_k = L_k L_k^T$.

Now consider $Y = L_k^{-1} X_k$, a random vector. Its mean and covariance matrix are given by:

$$\begin{aligned} E[Y] &= E[L_k^{-1} X_k] = L_k^{-1} E[X_k] = 0 \\ E[YY^T] &= E[(L_k^{-1} X_k)(L_k^{-1} X_k)^T] = E[L_k^{-1} X_k X_k^T (L_k^{-1})^T] = L_k^{-1} E[X_k X_k^T](L_k^{-1})^T \\ &= L_k^{-1} P_k (L_k^{-1})^T = L_k^{-1} L_k L_k^T (L_k^{-1})^T = I \end{aligned}$$

If now, the random vector Y were to be multiplied with L_l, we can define $Z = L_l Y$. It follows that:

$$E\left[ZZ^T\right] = E\left[(L_l Y)(L_l Y)^T\right] = E\left[L_l Y Y^T L_l^T\right] = L_l E\left[YY^T\right] L_l^T = LL^T = P_l$$

Hence, the random vector Z has covariance matrix P_l.

F.2 Discussion

The transformation of the component type of a cell is done by:

$$(\mathbf{x}_t^{m,f'})_j = \mu_j^l + L_j^l (L_j^k)^{-1} (F_k \circ ((\mathbf{x}_t^m)_j - \mu_j^k))) \tag{44}$$

$$\hat{\mathbf{x}}_t^{m,f'} = \hat{\mu}_j^l + \hat{L}_j^l (\hat{L}_j^k)^{-1} (\hat{\mathbf{x}}_t^{m,f} - \hat{\mu}_j^k)) \tag{45}$$

where $\hat{\mathbf{x}}_t^{m,f'}$ is the partial auxiliary state vector. It is partial since only those elements that contain the old component type k are considered. Further, $\hat{\mu}_j^l$ is the new component mean and $\hat{\mu}_j^l + \hat{L}_j^l (\hat{L}_j^k)^{-1} (\hat{\mathbf{x}}_t^m - \hat{\mu}_j^k))$ the deviation after transformation from the new component mean. The matrices $\hat{L_j}^l$ and $\hat{L_j}^k$ are the lower triangular Cholesky decompositions of a sub-matrix of the component covariance matrices P_j^l and P_j^k, defined for the same elements as the partial state vectors. We note that this means that we are working with a different covariance matrix for each member since each member has in general a different spatial composition of component types.

The Cholesky decomposition of the component covariance matrices only exists if a matrix is positive definite. In case of a covariance matrix this is theoretically known to be the case.

For finite ensemble approximations, however, the sample covariance matrix can appear negative definite. This issue can be solved by adding a small constant to the diagonal of the sample covariance matrices, with the size of the measurement error. The influence of this constant is that it attenuates the relations between variables slightly.

In theory, as the correlation structure of both component types is the same initially, the product should be a unit matrix. However, taking the product of the adjusted decomposed Cholesky matrix and inverse matrix sometimes shows rather differently. Especially if one component type has a small probability, gross errors are introduced. It is possible that the attempts to maintain the correlation structure have an adverse effect and destroy the correlation structure instead. Therefore it is better to use either a simple translation of the means, or, if enough computational power is available, geostatistical methods.

List of Figures

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B. (2009): The Ensemble Kalman Filter in Reservoir Engineering- a Review. *SPE Journal 14*(03): 393-412

Armstrong, M., Galli, A., Beucher, H., Loc'h, G., Renard, D., Doligez, B., Eschard, R., Geffroy, F. (2011): *Plurigaussian Simulations in Geosciences.* Springer-Verlag Berlin Heidelberg

Burgers, G., van Leeuwen, P.J., Evensen, G. (1998): Analysis Scheme in the Ensemble Kalman Filter. *Mon Weather Rev 126*: 1719–1724

Delft, G. van, Serafy, G.Y., Heemink, A.W. (2009): The ensemble particle filter (EnPF) in rainfall-runoff models. *Stoch Environ Res Risk Assess 23*:1203-1211

Dovera, L., Della Rossa, E.(2011): Multimodal ensemble Kalman filtering using Gaussian mixture models. *Comput Geosci 15*:307-323

Evensen, G. (1994): Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99(C5), 10143–10162

Evensen, G. (2003): The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean. Dyn. 53*, 343-367

Harbough, A.W.(2005): MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—the Ground-Water Flow Process , *U.S. Geological Survey Techniques and Methods 6-*A16

Hekman, A., Manen, H. van, Lange, W. de (2014, January 24): AZURE: Innoveren in de 'gouden driehoek'. Retrieved from http://www.vakbladh2o.nl/

Hoteit. I., Luo. X., Pham, D. (2012): Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters. *Mon. Wea. Rev.*, 140, pp 528–542.

Kalman, R.E. (1960): A New Approach to Linear Filtering and Prediction Problems. Transcript retrieved from http://enkf.nersc.no/Publications/kalman1960.pdf on (2014, August 18).

Kitandis, P.K.(1997): *Introduction to Geostatistics.* Cambridge: Cambridge University Press. ISBN:0521583128

Liu, N., Oliver, D.S. (2005): Ensemble Kalman filter for automatic history matching of geologic facies. *Journal of Petroleum Science and Engineering 47*: 147-161

NERC (2014, September 17): How Groundwater Moves. Retrieved from: http://www.cieh.org/uploadedFiles/Core/Policy/Environmental_protection/Water/Private_water_supplies/C2.pdf

Oliver, D.S., Chen, Y. (2011): Recent progress on reservoir history matching: a review. *Comput Geosci 15*:185-221

Sebacher, B., Hanea, R., Heemink, A. (2013): A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). *Comput Geosci 17*:813–832

Seiler, A., Rivenaes, J.C., Aanonsen, S.I., Evensen, G., StatoilHydro ASA (2009): Structural Uncertainty Modelling and Updating by Production Data Integration. In SPE/EAGE Reservoir Characterization and Simulation Conference.

Remy, N., Boucher, A., Wu, J. (2009): *Applied Geostatistics with SGeMS.* Cambridge: Cambridge University Press. ISBN:9780521514149

Soares, A. (2001): Direct Sequential Simulation and Cosimulation. *Mathematical Geology 33*(8): 911-926

Terejanu, G.A. (2014, August 18): Discrete Kalman Filter Tutorial. Retrieved from http://www.cse.sc.edu/ terejanu/files/tutorialKF.pdf

Valstar, J.R.(2001): *Inverse modeling of groundwater flow and transport.* ISBN:9064640629

Vernes, R.W., van Doorn, Th. H.M. (2005): TNO-Rapport, Van Gidslaag naar Hydrogeologische eenheid, NITG 05-038-B

Wen, X.-H., Chen, W.C. (2005): Real-time reservoir model updating using Ensemble Kalman Filter. *SPE 92991, SPE Reservoir Simulation Symposium 2005, 31 January-2 February Houston*

Willmott, C.J., Matsuura, K. (2005): Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res Vol. 30*: 79–82

Zhang, Y., Oliver, D.S. (2011): Evaluation and error analysis: Kalman gain regularization versus covariance regularization. *Comput Geosci 15*:489-508