A Bayesian framework for validating and comparing models for prediction of surface displacements due to reservoir compaction

Restricted
SR.15.11680
June 2015

Restricted **SR.15.11680**

A Bayesian framework for validating and comparing models for prediction of surface displacements due to reservoir compaction

by

T. Park (GSUK PTD/TASE), S. Bierman (GSNL PTD/TASE)

Shell Global Solutions International B.V., Amsterdam

Further electronic copies can be obtained from the Global Information Centre.

Executive Summary

Model predictions of displacements at the surface of the earth due to reservoir compaction are subject to uncertainties due to the fact that:

- Some of the physical processes are not well understood, and multiple candidate models may be proposed which may be similar in their ability to explain historic data but different in their predictions of future displacements.
- For any given model, estimates of model parameters are uncertain.
- Model input, in particular spatio-temporally resolved estimates of pressure declines and rock porosity, are subject to uncertainties.

Additionally, measurements of displacements are subject to errors and part of the displacements may be caused by non-reservoir related processes which are not accounted for in the models.

A rational framework is required to enable quantification of uncertainties in model parameters and predictions, and to compare the relative ability of models to explain the variability in future measurements. For this purpose, we give an outline of a Bayesian statistical framework which is flexible enough to accommodate the use of prior information surrounding model parameters (e.g. prior knowledge from an understanding of physical processes or laboratory measurements), uncertainties in model input, and errors in measurements. A key advantage of the Bayesian statistical framework is that it offers a natural framework for estimating probability distributions for key quantities of interest such as future observations.

We make the following recommendations:

- When surface displacement measurements are used to constrain model parameters the available data should be partitioned into a training and test data set, denoted by D and $\tilde{D}$ respectively, such that measurements in $\tilde{D}$ were made at later dates than the measurements in D. The training data set should be used for estimation of posterior distributions of model parameters, whereas the predictive performance of models should be assessed using $\tilde{D}$.
- A quantitative measure such as Bayes factors, evaluated using the posterior predictive distributions and $\tilde{D}$, should be used to compare the relative predictive performance of models.
- Forecasts of future surface displacements should be probabilistic and care should be taken to communicate as accurately as possible the uncertainties surrounding these forecasts.
- We should asses if there is scope for making improvements to our predictions. This could be done by comparing the posterior distributions of variance components of model residuals with independent knowledge on errors in measurements and displacements due to processes other than reservoir compaction.

In this report, we provide an outline of the proposed Bayesian framework for model comparison and model validation as well as a simple illustration of the proposed methodology. To enable modeling and forecasting of surface displacements within the proposed framework future work it is required to implement algorithms, such as Markov Chain Monte Carlo methods, to estimate posterior distributions of model parameters.

Amsterdam, June 2015.

Table of Contents

List of Figures

1. Introduction

The load of the sediments above a gas bearing rock formation is supported partially by the rock matrix itself and partially by the pressurized fluid and gas within the pore space of the rock. The decrease in fluid volume in the rock associated with gas extraction results in pore volume reduction (compaction) at the reservoir level. The compaction of the reservoir results in a measurable amount of displacements, such as subsidence, at the surface. The Nederlandse Aardolie Maatschappij (NAM) is required to forecast future surface displacements above gas fields consequent on scenarios for future production.

Compaction is modeled as a function of pressure depletion, the compressibility of the rock and the thickness of the reservoir. A number of rock physics models are currently in use by NAM (NAM [2013], Mossop et al. [2011]), whereas other models may be proposed in future. The current NAM models are calibrated using measurements of surface displacements such as from optical levelling campaigns. Forecasts of the shape and maximum depth of the subsidence bowl differ substantially between these models. The current level of understanding of the rock physics is insufficient to confidently identify a single subsidence model, nor to eliminate any of the current models as possible candidates. Laboratory compaction measurements of core samples from the reservoir are available, but these data unfortunately cannot be used to determine with confidence whether strain is linearly or non-linearly related to pore pressure decline. In all current NAM models the uni-axial compaction coefficient of a reservoir section is estimated as a function of the porosity of the rock in that section. The relationship between compaction coefficients and porosity differs between the models, and is based on trend lines fitted to laboratory measurements on cores taken from the reservoir (NAM [2013]). The laboratory data indicate that uni-axial compaction coefficients tend to increase with increasing porosity. However, uni-axial compaction coefficients can only be estimated with low precision on the basis of porosity measurements, and the laboratory measurements might not be representative for overburden compressibility. Spatio-temporally resolved estimates of pore pressure decline are uncertain, though the uncertainty surrounding these estimates has not been quantified. In summary, model predictions of displacements at the surface of the earth due to reservoir compaction are subject to uncertainties due to the fact that:

- Some of the physical processes are not well understood, and multiple candidate models may be proposed which may be similar in their ability to explain historic data but different in their predictions of future displacements.
- For any given model, estimates of model parameters are uncertain.
- Model inputs, in particular spatio-temporally resolved estimates of pressure declines and rock porosity, are subject to uncertainties.

Additionally, measurements of displacements are subject to errors and part of the displacements may be caused by non-reservoir related processes which are not accounted for in the models.

In this report, we outline Bayesian statistical methodology which can be used as a framework to accommodate uncertainties in model parameters, model input and measurements. The key requirements of the Bayesian statistical framework are:

- Quantification of uncertainties in model parameters and model predictions. An explicit quantification of uncertainties in model predictions is essential since, given the relatively large uncertainties, model forecasts should be probabilistic rather than deterministic.
- Quantification of the relative ability of models to explain future observations.

- Flexibility in terms of the extent to which model predictions rely on calibration using historic surface displacement data, or on 'prior information' such as prior knowledge from an understanding of physical processes or laboratory measurements.

We provide an outline of the proposed Bayesian framework for model comparison and model validation as well as a simple illustration of the proposed methodology. To enable modeling and forecasting of surface displacements within the proposed framework future work is required to implement algorithms, such as Markov Chain Monte Carlo methods, to estimate posterior distributions and to take into account prior (before calibrating using geodetic data) knowledge surrounding model parameters.

2. A Bayesian framework for validating and comparing models

In this chapter we describe the overall concepts of Bayesian modelling which we believe are applicable both to the problem of building predictive surface displacement models and of validating existing models. We prefer a Bayesian statistical framework because of the pragmatic advantages that this offers, in particular its flexibility and ability to cope with complex problems, and because it offers a consistent framework for estimating probability distributions for key quantities of interest such as future observations (Gelman et al. [2003]). By using a Bayesian framework we are able to build and compare models which have either been calibrated to displacement data or built entirely from knowledge of physical processes. A combination of calibration using displacement data and the use of prior information from knowledge of physical processes is also possible. A general introduction to Bayesian inference and data modeling is given by Gelman et al. [2003]. .

Let $\mathcal{M}$ be a model for the prediction of surface displacement data, D, with a vector of model parameters θ. In a Bayesian framework we treat the parameters θ as random variables. This represents the fact that there will always be at least some uncertainty surrounding the true value of these parameters. Our aim is to use all available information, both a-priori information arising from knowledge of physical processes and subsidence data ,to estimate the *posterior distribution* of each parameter denoted by $p(\theta|D)$. The posterior distribution is defined as,

$$p(\theta|D) \propto p(D|\theta)p(\theta), \tag{2.1}$$

where $p(D|\theta)$ is the likelihood function of the data and $p(\theta)$ is the *prior* distribution of the parameters. The posterior distributions can be used to calculate quantities such as the expected value of the parameters as well as their confidence intervals (also referred to as 'credible intervals' within a Bayesian framework).

The form of the likelihood function must be chosen based on assumptions about the probability distribution of the data. The prior is an important quantity as this contains our belief about the model parameters *before* we look at the data. If we do not have strong beliefs about the range of values for the model parameters then we can use weak, or uninformative, prior probability distributions for the values that we believe these parameters can take. Such uninformative prior distributions will have a very small influence on the posterior distributions compared to the influence of the data. If however we have very good prior knowledge about the parameters then we can use a strong prior. A strong prior will have a much greater influence on the posterior distribution. If we have a strong prior it is also possible for us not to use any data to fit the model, in this case $p(\theta|D) = p(\theta)$ and our prior beliefs are the only inputs to the model.

Having estimated the posterior distributions of the parameters we can predict future observations, $\tilde{D}$. Due to the uncertainty in estimating the parameters of the model the predictions will also be uncertain. We capture this using the *posterior predictive distribution* which is defined as,

$$p(\tilde{D}|D) = \int p(\tilde{D}|\theta)p(\theta|D)\mathrm{d}\theta. \tag{2.2}$$

In a similar way to the posterior distribution the posterior predictive distribution can be used to calculate the expected value of future observations as well as the uncertainty surrounding the expected values. In addition, the likelihood of the test data set given the predictions can be evaluated and may be used in statistical hypothesis testing.

In certain cases the posterior and posterior predictive distributions can be calculated analytically. However in many cases these distributions must be approximated using numerical methods such

as *Markov Chain Monte Carlo* simulation, see [Gelman et al. [2003]] for more information. This means that while in theory this framework can be applied to a very general choice of model and prior distributions there may be difficulties in practice.

As mentioned previously we have two sets of data, D and $\tilde{D}$. Where D is the data set which we can observe now and $\tilde{D}$ is the set of data which we will observe in the future. In order to compare and validate models without needing to wait for future observations we can partition our current data set into a training and a test set. The training set is used to evaluate the posterior distribution, $p(\theta|D)$, and both are used to calculate the posterior predictive distribution, $p(\tilde{D}|D)$. This is important as it enables us to evaluate and compare models based on their ability to predict future data rather than their fit to current data. The training and test data sets should not overlap and the test set should span a time period (epoch) after the measurements in the training set. If strong priors are available it is also possible that the training set will be empty and in this case the test set can contain all of the available data. In the case where uninformative or partly informative priors are used, the information present in the training data set is used to constrain the parameters.

Displacement data at the surface may consist of a combination of optical levelling surveys, Interferometric Synthetic Aperture Radar (InSAR) images or Global Positioning Systems (GPS). At the reservoir level, spatio-temporally resolved estimates of pressures or pressure declines are available. Furthermore, other information may be used such as a geomechanical influence matrix which governs displacements at the surface as a fuction of displacements at the reservoir level and which may itself be a function of parameters which are uncertain. The Bayesian framework is flexible and can accommodate both strong and weak prior information, and may assume individual variables to be either fixed constants or random variables. To illustrate these principles, we consider two scenarios:

1. Weak or no prior information is available: Parameters are constrained using a set of training data, and model predictions are evaluated against a test data set. Reservoir pressures are assumed to be known without error, although several scenarios for pressure estimates can be used. The geomechanical influence matrix is assumed to known without error.
2. Sufficiently strong prior information is available: No updating of prior information using the subsidence data is required. In this case it is not necessary to use any training data. Reservoir pressures are assumed to be known without error, although several scenarios for pressure estimates can be used. The geomechanical influence matrix is assumed to be known without error.

2.1. Illustration 1: Predictive Models With Data Calibration

We first consider a scenario where we only have weak prior information about the model parameters. This is represented by the Directed Acyclic Graph, DAG, shown in Figure 2.1. In this DAG we show the different components of the model and how they relate to each other. Variables contained in square boxes are treated as known and fixed while those in circles are random variables. In this scenario, the fixed quantities are the reservoir pressures, P, and the geomechanical influence matrix X. We could also treat P as a random variable however, in the absence of a full description of the uncertainty, we can also treat it as fixed and include uncertainty by considering a range of models with different fixed pressure scenarios. The model parameters, θ, can describe a wide range of models such as a linear relationship between pressure and subsidence or a time decay relationship. These model parameters have a prior distribution which is described by the fixed parameters, τ_θ. Another aspect of modelling of surface displacements is the covariance matrix of the observations, Σ_D. The covariances in the displacement data may have contributions from several different spatially uncorrelated, spatially correlated and temporally correlated factors, denoted

respectively by σ_s, σ_m and σ_t, which are considered separately and also have prior distributions specified by γ_s, γ_m and γ_t.

The arrows on the DAG show the relationships between the variables, those with an $=$ are deterministic relationships while those with a $\sim$ are stochastic relationships indicating that one variable is a random realisation from a distribution determined by the other.

Under this scenario the training data D is used to increase our knowledge of the model parameters thus giving the posterior predictive distribution a smaller variance. This smaller variance does not necessarily indicate a more accurate prediction and this should be established based on comparison with the test data.

As noted before, it is also possible to have strong prior information about some of the model parameters in which case there will be less updating of these priors by the information present in the data.

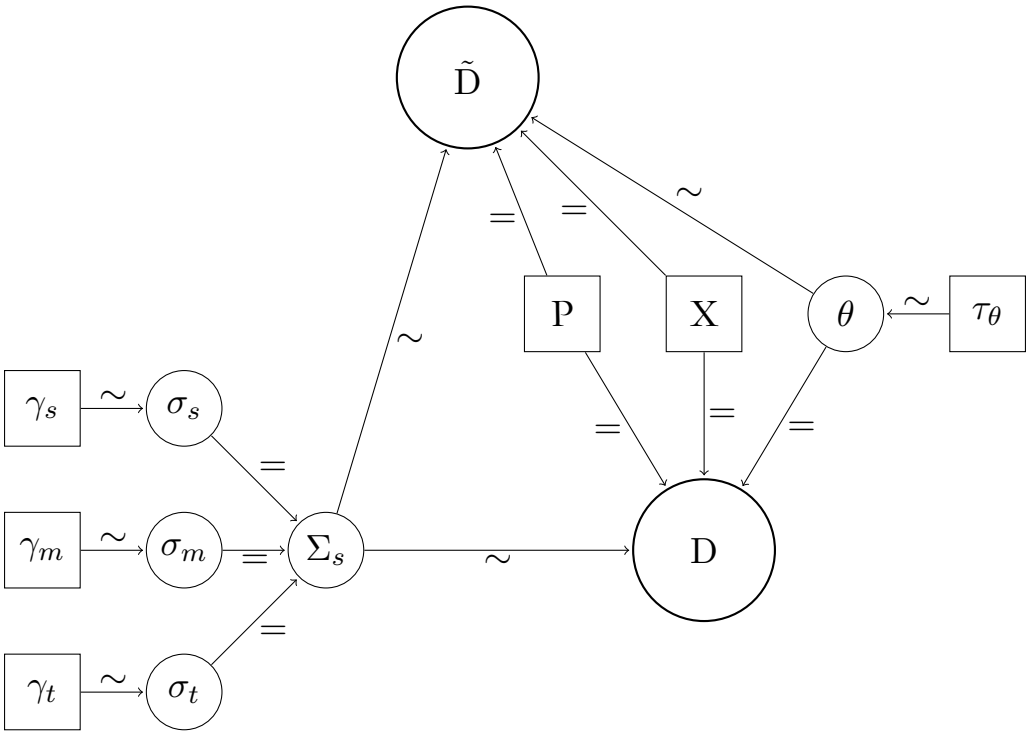

Figure 2.1.: Directed Acyclic Graph (DAG) representing a predictive model with weak prior information.

2.2. Illustration 2: Predictive Models With No Data Calibration

The second scenario which we consider is one where sufficiently strong prior information is available regarding all model parameters such that there is no requirement to constrain any of the parameters using the displacement data. This is represented by the DAG in Figure 2.2, which is similar to the DAG in Figure 2.1 with the exception that there is no relationship between model parameters and the training data. In this scenario the only inputs to the model are is the prior information and so the quality of the model relies entirely on the accuracy of that prior information.

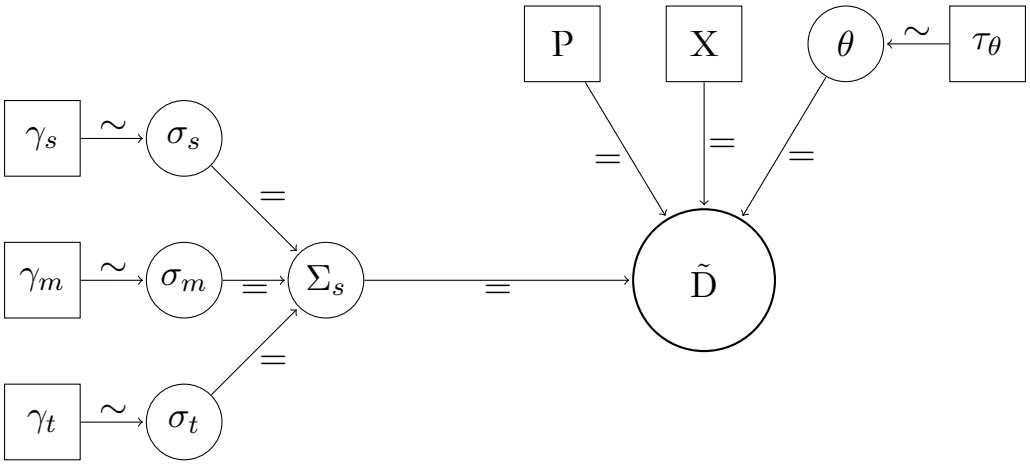

Figure 2.2.: Directed Acyclic Graph (DAG) representing a predictive model with strong prior information and no updating of model parameters using a training data set (see text).

3. Model Comparison

As stated previously, the validity of a model should be assessed based on how well it is able to predict the measured variation in displacements in the test data set. An important advantage of the proposed Bayesian modeling scheme is that this explicitly takes into account the amount of information that is present in the displacement data to support model complexity. Where little information is present in the training data to constrain model parameters, the uncertainty in model predictions will increase and therefore the likelihood of the test data set given the model predictions will decrease.

Models should be compared in terms of their ability to describe the observed variation in all of the available displacement data, using a formal statistical test. The criterium for statistical testing we propose is the *Bayes factor*. Given two candidate models, $\mathcal{M}_i$ and $\mathcal{M}_j$ the comparison criteria is given by,

$$c\left(\mathcal{M}_i, \mathcal{M}_j\right) = \frac{p(\tilde{D}|\mathcal{M}_i)}{p(\tilde{D}|\mathcal{M}_j)} \tag{3.1}$$

where $p(\tilde{D}|\mathcal{M}_i)$ is the likelihood of the test data under the posterior predictive distribution found by fitting model $\mathcal{M}_i$. If the comparison criteria has a value greater than 1 this indicates that model i is stronger, whereas a value of less than 1 indicates that model j is stronger. We recommend comparing each candidate model to a simple null model, such as a one parameter model for the average observed displacement. It is also possible to estimate the probability distributions of the Bayes factors, and these distributions can form the basis for statistical testing.

This comparison framework is flexible and can accommodate situations with strong or weak prior information about some or all of the parameters. Also, it is possible to test fully probabilistic predictions of models without knowledge of the model form or how it was calibrated ('black-box models'). The only requirement for such black-box models is that predictions are probabilistic and that the model was not calibrated using data in the test set.

We note that, using the proposed statistical testing method alone, it is not possible to confidently asses if we have the correct model as it is always possible that another equally good or better model exists which we have not yet considered.

4. Assessing the fit of the model

It is important to evaluate the ability of models to explain the measured variation in the test data set (the 'fit' of the model to the data). There are many different aspects of model fit which we can consider and these cannot be adequately captured in one measure. We recommend the following for the assessment of model fit:

1. Assess whether the distribution of the test data agrees with the posterior predictive distribution. We recommend to use the posterior predictive p-values as defined by Gelman et al. [2003] pages 145-146:

   > The Bayesian p-value is defined as the probability that the replicated data could be more extreme than the observed data, as measured by the test quantity.

2. Compute standardised residuals, r. If the expected value of the prediction for test data point $\tilde{D}_i$ is μ_i and its standard deviation is σ_i, then the standardised residuals are defined as,

   $$r_i = \frac{\tilde{D}_i - \mu_i}{\sigma_i}. \tag{4.1}$$

   For example, when the posterior predictive distribution is normally distributed, $r_i \sim N(0, 1)$. The standardised residuals can then be used to identify outliers and can also compared to the expected distribution using a quantile-quantile ('qq') plot.

3. Compare posterior distributions of variance components with independent knowledge on errors in measurements and displacements due to processes other than reservoir compaction. If the estimated variances are larger than the deviations as expected through measurement errors and non-reservoir related displacements, this would indicate that there is room for improvement and that other models may be able to give a better explanation of the test data set.

4. Assess whether model residuals have a spatial or temporal signature. Using the posterior predictive distributions it is possible to test whether there is evidence that certain aspects of 'model misfit', such as a correlation between residuals and a position on a geographical cross-section of the area of interest, could have arisen by chance under the candidate model and estimated variance components (a Bayesian posterior p-value may be defined on an appropriate posterior predictive distribution of a quantity of interest).

5. Illustrative Example

In order to illustrate our proposed model comparison and validation method we present an illustrative hypothetical example for a model with $N+1$ parameters θ_n ($n = 1, 2, ..., N+1$) which predicts surface displacements Y_i at a set of k locations ($i = 1, 2, ..., k$) using an $[k \times m]$ influence matrix X where element $X_{i,j}$ describes the influence of reservoir section j ($j = 1, 2, ...m$) on surface location i (for a given set of epochs). Let $z_i = \Sigma_{j=1}^{m} X_{i,j}$ be the sum of the influences of reservoir sections j on surface location i, and $\mathbf{z}$ be the $[k \times 1]$ column vector of influence measures. Furthermore, let Z be a $[k \times N+1]$ design matrix with a value of 1 for all elements in in the first column (to model the intercept), $\mathbf{z}$ for the second column, $\mathbf{z}^2$ for the third, etc. until $\mathbf{z}^N$ for the $(N+1)^{th}$ column. We consider a model of the form,

$$Y = Z\theta + \epsilon, \tag{5.1}$$

where θ is the vector of $N+1$ parameters to be estimated, and $\epsilon \sim N(0, \sigma^2 I_k)$ an vector of model residuals with zero mean and variance of σ^2 (I_k denotes the $[k \times k]$ identity matrix).

The true underlying model from which we simulate our data is of order $N = 2$, thus using the first three columns of X (intercept, linear in $\mathbf{z}$ and quadratic in $\mathbf{z}$). The values used in the simulation are: $k = 200$ measurements, $\theta_{n=1} = 0$ (intercept) $\theta_{n=2} = 0.8$ (linear term), $\theta_{n=3} = 1$ (quadratic term), and $\sigma^2 = 25$ (variance of model residuals). We partition the simulated data in two parts such that the first half contains the lowest values for $\mathbf{z}$. The first half of this partition is used as the training set and the second half as the test set. We place flat priors on all parameters. In this example, the posterior predictive distributions can be calculated analytically. We fit a range of polynomial models of different orders $N = 0, 1, \ldots, 5$ and assess the predictive capabilities of each model. We use the model of order $N = 0$ as our null model (intercept only) to which each other model is compared. Bayes factors and posterior distributions and posterior predictive distributions are estimated. Results are shown in Figures 5.1 and 5.2. The model with the highest Bayes factor is of order $N = 2$ in accordance with the true underlying model (top left panel in Figure 5.1). Bayes Factors for more complex models with higher orders of N are however similar so we may need to consider the variance of the Bayes factor to judge if this is significant. The variance of the Bayes Factor is a function of the variance in the estimated likelihoods of the models under consideration. For example, the posterior variance of the model of order $N = 2$ is given in Figure 5.2. The model of order $N = 1$ is clearly not able to explain the measurements in the test data set as most of the measurements fall outside of the 95% confidence bounds of the posterior predictive distributions (top right panel in Figure 5.1). The fit of the model of order $N = 2$ appears to be adequate since the 95% bounds on the posterior predictive distributions of the model contain most of the measurements, and the residuals closely follow the expected normal distribution. In practice, other aspects of the fit of this model will have to be assessed, in particular evidence of the existence of spatial and/or temporal patterns in residuals. In the Bayesian framework, it may be possible to quantitatively assess the statistical significance of such patterns using the posterior predictive distributions.

The posterior distributions of the parameters of the model of order $N = 2$ are given in Figure 5.2. Additional statistical significance tests are possible using these posterior distributions. For example, it is straightforward to compute the posterior probability that parameters are significantly different from zero. In this case, there is sufficient information in the data to constrain the quadratic term (parameter θ_3) such that there is strong evidence that it is larger than zero. The posterior distribution for the estimated variance can be compared with independent knowledge on errors in measurements and signals in displacement data due to non-reservoir related processes to assess if there is potential to improve the model.

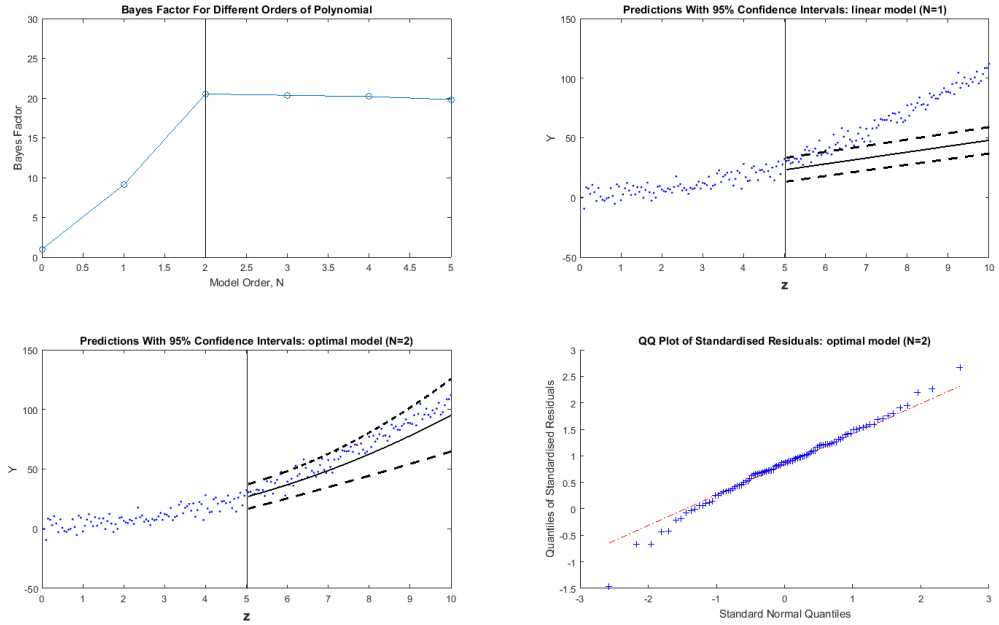

Figure 5.1.: Results of our analysis for an illustrative hypothetical example. Bayes factors for a range of models of increasing complexity $N = 0, 1, \ldots, 5$ are given in the top left panel. The fit of the posterior predictive distributions of measurements in the test data set $P(\tilde{D}|D)$ given the model of order $N = 1$ (calibrated using the training data set D) are given in the top right panel. The fit of the posterior predictive distributions of measurements in the test data set $P(\tilde{D}|D)$ given the model of order $N = 2$ is given in the bottom left panel. The plot in the bottom right panel shows a quantile-quantile plot comparing the standardised residuals to a standard normal distribution.

As noted earlier in this report, it will be necessary to take spatial and/or temporal covariances in errors in measurements into account in the application of the Bayesian framework to a genuine model for reservoir compaction and displacement measurements. This will be important in particular to be able to estimate realistic posterior predictive distributions. To enable modeling and forecasting of surface displacements within the proposed framework, future work is required to implement algorithms which allow the estimation of posterior distributions such as Markov Chain Monte Carlo methods (see e.g. Gamerman and Lopes [2006] for a general reference and Yan et al. [2007] for an example of an algorithm for drawing values from the posterior distributions in the presence of spatio-temporal covariances in the model residuals).

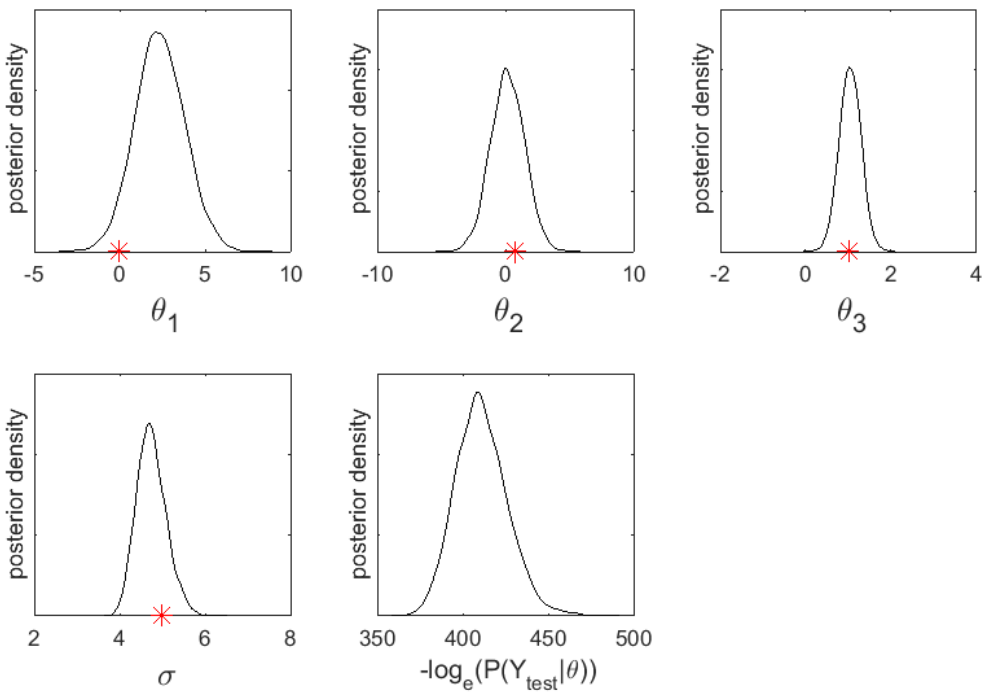

Figure 5.2.: Posterior distributions of the parameters and the negative log likelihood for the model of order $N = 2$ for our illustrative example. True values for the parameters, used to generate the data, are indicated using red asterisks.

References

Dani Gamerman and Hedibert F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, 2006. ISBN 1584885874.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2003. ISBN 158488388X.

A Mossop, O van der Wal, and R van Eijs. Subsurface technical report: Subsidence modelling of ameland fields. Technical report, Nederlandse Aardolie Maatschappij, 2011.

NAM. Technical addendum to the winningsplan groningen 2013 subsidence, induced earthquakes and seismic hazard analysis in the groningen field. Technical report, NAM, 2013.

J. Yan, M. Cowles, S. Wang, and M. Armstrong. Parallelizing mcmc for bayesian spatiotemporal geostatistical models. *Statistics and Computing*, 17:323–335, 2007.

Bibliographic information

Classification	Restricted
Report Number	SR.15.11680
Title	A Bayesian framework for validating and comparing models for prediction of surface displacements due to reservoir compaction
Subtitle	
Authors	T. Park (GSUK PTD/TASE), S. Bierman (GSNL PTD/TASE)
Keywords	subsidence, gas filed, bayesian statistics, model comparison
Issue Date	June 2015
Period of work	November 2014 - June 2015
US Export Control	Non US — Public Domain
WBSE Code	ZZPT/022595/14010183
Reviewed by	P. Jonathan (GSUK PTD/TASE)
Approved by	P. Jonathan (GSUK PTD/TASE)
Sponsoring Company/ Customer	Nederlandse Aardolie Maatschappij (NAM)
Spons./Cust. Address	Schepersmaat 2, 9405 TA, Assen, The Netherlands
Issuing Company	Shell Global Solutions International B.V., Amsterdam P.O. Box 38000 1030 BN Amsterdam The Netherlands

Report distribution

Electronic distribution (PDF)

Name, Company, Ref. Ind.	
PT Information Services, PTT/TIKE, PT-Information-Services@Shell.com	PDF
Bierman Stijn M, GSNL-PTD/TASE	PDF
Doornhof Dirk, NAM-PTU/E/Q	PDF
Jansen Jeroen M, NAM-UIO/T/H	PDF
Jonathan Philip, GSUK-PTD/TASE	PDF
Mossop Antony P, NAM-PTU/E/Q	PDF
Park Timothy A, GSUK-PTD/TASE	PDF
van Boom Ruud, NAM-UIO/T/DL	PDF
van der Wal Onno, NAM-PTU/E/Q	PDF
van Eijs Rob MHE, NAM-PTU/E/Q	PDF

Principal typesetting performed with LaTeX system (MiKTeX) using the PTreport2015.cls class based on the KOMA-Script with the necessary modifications to match the Shell Projects and Technology house style.

The PTreport2015.cls class file is written and maintained by IdlTeX (www.idltex.com). All rights reserved.