Jo Røislien

Random Field Models and Near Well Reservoir Characterization

Trondheim, April 2004

Doctoral thesis for the degree of doktor ingeniør

Norwegian University of Science and Technology
Faculty of Engineering Science and Technology
Department of Petroleum Engineering and Applied Geophysics

NTNU

URN:NBN:no-3476

Jo Røislien

Random Field Models and Near Well Reservoir Characterization

Dr. Ing. Thesis

Department of Petroleum Engineering and Applied Geophysics
Norwegian University of Science and Technology
2004

URN:NBN:no-3476

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree "Doktor ingeniør" (Dr.Ing) at the Norwegian University of Science and Technology (NTNU). The thesis research is part of the Uncertainty in Reservoir Evaluation (URE) project at NTNU, financed by petroleum companies and the Research Council of Norway.

I thank my supervisors, Ole Torsæter and Henning Omre for guidance and encouragement. Henning's comments and positive energy has been invaluable. Winter 2002/2003 I spent six months at Imperial College in London, writing a paper with professor Peter King, and summer 2003 I spent three months at Schlumberger-Doll Research Centre, where I got to work with Alberto Malinverno and his group. Both were great experiences, I will always be greatful for that.

I would also like to thank my friends in general, and the other students at the URE project; Jo Eidsvik, Stig-Kyrre Foss and Ole Petter Lødøen, in particular, for mathematical and non-mathematical quality time. I also thank the administration staff at the Department of Petroleum Engineering and Applied Geophysics, as well as the administration staff and the computer support group at the Department of Mathematical Sciences, where I spent most of my time.

Finally, a special thanks to my family and Ingrid. Were it not for your continuous support, this work would never be.

Trondheim, April 2004

Jo Røislien

Thesis Outline

The thesis consists of the following four parts:

Part 1: **T-distributed Random Fields – A Parametric Model for Heavy-Tailed Random Fields**
Jo Røislien, Henning Omre. Paper submitted for publication.

Part 2: **Continuous Near-Well Bayesian Linearized AVO Inversion**
Jo Røislien, Henning Omre. Report.

Part 3: **Discrete Representations of Continuous Random Fields**
Jo Røislien, Alberto Malinverno, Michael Prange, Nick Bennett. Report.

Part 4: **Prediction of Oil Production Conditioned on the Breakthrough Time**
Jo Røislien, Peter R. King, Eduardo Lopez, Sergey V. Buldyrev, H. Eugene Stanley, Shlomo Havlin. Paper submitted for publication.

All parts are self-contained, and can be read in any order.

Thesis Introduction

A brief introduction to the topic under study is given, together with a short summary of each of the four parts.

Reservoir Characterization

When describing a reservoir, several reservoir variables are of interest. These include discrete variables as rock type and fluid filling, as well as continuous variables as porosity and permeability. An important intermediate step is to determine the elastic material parameters P-wave velocity, S-wave velocity and density, alternatively some reparametrization of these. No direct observations are available, but indirect measurements can be provided. This results in a so-called inversion problem. Inverse problemse arise naturally when seeking to determine a cause given an effect. The opposite, determining an effect given a cause, is the forward problem. For a more on inverse problems, see Tarantola (1987), Kitanidis (2001) and Kolbjørnsen (2002).

In reservoir characterization the available information is typically sparse, and geophysical inversion problems are generally underdetermined and ill-posed. From a statistical point of view, as we are interested not only in a best guess, but also the attached uncertainty, Bayesian inversion is well-suited for the task. The solution in Bayesian inversion is a pos-

terior probability distribution, which formally is proportional to a prior distribution times the likelihood model. In the prior distribution, assumptions on the preferred solution can be introduced in the model, guaranteeing that a stable and unique solution to the inversion problem can be found. The likelihood model connects the data with associated observation error models to the variables of interest. Using the notation m for the elastic material parameters of interest, d for the available data, and letting $f(\cdot)$ denote a probability density function (pdf), the goal is to determine $f(m|d)$, which represents the pdf of m given d. Bayes' theorem reads

$$f(m|d) = const \times f(d|m)f(m)$$

with normalizing constant $\left[\int f(d|m)f(m)\mathrm{d}m\right]^{-1}$. This Bayesian inversion approach is applied in reservoir characterization by Buland and Omre (2003). The prior pdf for the elastic material parameters, $f(m)$, is where prior knowledge of the model under study can be introduced. In the following, we assume that m is a continuous random field, see Christakos (1992) and Chiles and Delfiner (1999) for more on random fields. The likelihood function $f(d|m)$ provides the link between the observations and the elastic material parameters. This function contains the aforementioned forward model with associated observation error models, and this function will thus be defined through geophysical relations like the Zoeppritz equation. Such relations are often non-linear, see Sheriff and Geldart (1995). The observations themselves are weighted averages over space, so the likelihood model will include some spatial dependece.

Observations

When describing a reservoir, there are three types of information available: Seismic data, well observations and production history data.

There are several different kinds of seismic data. In this thesis, the focus is on surface seismics collected from vessels, in particular prestack AVO data. Seismic data generally exist on a large scale, often throughout the entire reservoir. The data, however, are sampled on a coarse grid, typically 25m×25m horisontally and 1ms vertically. Also, they are averaged in the vertical direction weighted by the wavelet, being of magnitude 10^1m at typical reservoir depth. Different types of seismic data are associated with different uncertainties and noise levels.

Well observations mainly consist of well logs. All well observations are generally rather sparse compared to seismic data, as they only exist along the well paths in the existing wells. However, the observations are sampled on a very fine grid along this path, typically every 0.15m. These measurements are much more detailed than seismic data, as they are averaged only with the size of the logging tool, typically being of order $10^0 - 10^{-1}$m, depending on the logging method. The uncertainty associated with these measurements is also generally very low.

Production history data are collected during the production process in production wells. Examples of such measurements are production rates, pressure data and fractional flow data. The observations will normally be crucially dependent on the recovery process actually used, and global features of the reservoir have impact. This global dependence make production data different from seismic data and well observations, which are of local type. Production data rarely suffice to characterize heterogeneous reservoirs, and a large amount of uncertainty still remains after history matching of geostatistical models.

Continuous Random Fields

A continuous random field $\{Z(x); x \in \mathcal{D} \subset \mathbb{R}^n\}$ is a function whose value for any $x \in \mathcal{D}$ is a random variable $Z(x) \in \mathbb{R}^1$. In geostatistics random fields is a natural way of describing continuously varying spatial phenomena, such as porosity, permeability and elastic material properties. A general random field can be described by the cumulative distribution

$$F_x(z) = F_{x_1,\ldots,x_n}(z_1,\ldots,z_n) = \text{Prob}\{Z(x_1) \leq z_1,\ldots,Z(x_n) \leq z_n\}$$

for any integer $n \geq 1$ and any configuration $\{x_i\}_{i=1}^n$ with $x_i \in \mathcal{D}$. Assuming they exist, the two first moments of the random field,

$$\begin{aligned} \mu(x) &= \text{E}\{Z(x)\} \\ \Sigma(x,y) &= \text{Cov}\{Z(x), Z(y)\}, \end{aligned}$$

can be expressed through $F_x(z)$ as Stieltjes integrals in $\mathbb{R}^1$ and $\mathbb{R}^2$, respectively, see Billingsley (1986). The most frequently used continuous random field model is the Gaussian random field (GRF), which is fully described by the two first moments, $\{\mu(x); x \in \mathcal{D}\}$ and $\{\Sigma(x,y); x,y \in \mathcal{D} \times \mathcal{D}\}$. Further, the corresponding multivariate distributions are simple, and so are estimation and inference. See Christakos (1992) and Chilés and Delfiner (1999) for more on random fields.

The correlation function for a random field is defined as

$$\rho(x,y) = \frac{\Sigma(x,y)}{\sigma(x)\sigma(y)}$$

with

$$\sigma^2(x) = \Sigma(x,x)$$

being the variance. The covariance function must be a positive definite function, i.e.

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j \Sigma(x_i,x_j) > 0$$

for any integer $n \geq 1$, any set of weights $\{c_i\}_{i=1}^n$ with $c_i \in \mathbb{R}^1$, and any set of locations $\{x_i\}_{i=1}^n$ with $x_i \in \mathcal{D}$. It can be shown that the class of correlation functions coincide

with the class of positive definite functions where $\Sigma(x, x) = 1$. As the two first moments uniquely defines a GRF, the study of GRFs is mainly a study of correlation functions, making positive definiteness a fundamental property of interest.

Often some symmetries in the random field can be assumed. A random field is said to be stationary in the strict sense if all finite-dimensional distributions are invariant under translation, see Abrahamsen (1994) for a definition. Stationarity in the strict sense implies stationarity in the wide sense, defined by stationarity of the two first moments;

$$\begin{aligned} \mu(x) &= \mu \\ \Sigma(x, y) &= \Sigma(y - x). \end{aligned}$$

Consequently, stationarity in the wide sence is sufficient to have a stationary GRF. Note that generally, stationarity in the wide sense does not imply stationarity in the strict sense. The covariance function of a stationary random field, called a stationary covariance function, must have constant variance

$$\Sigma(y - x) = \sigma^2 \rho(y - x).$$

Further, a stationary random field in the wide sence is termed an isotropic random random field if the covariance function depends on distance alone;

$$\Sigma(x, y) = \Sigma(|y - x|)$$

Isotropy entails translation, rotation and reflection invariance. Other symmetries, as well as non-stationary random fields, can also be modelled. See the aforementioned references for more.

Contents of Thesis

An outline of the contents of all four parts follows.

In Part 1 T-distributed random fields (TRF) are adressed. The main contribution of this paper is the introduction of a new, analytically tractable, heavy-tailed continuous random field model, namely the TRF model. In Gunning (2002) a related random field model is discussed. Based on a motivating example from a well log from the Gullfaks field in the North Sea, it is demonstrated how the T-distribution is able to describe the variability in the geophysical data better than the frequently used GRF model. This flexibility is obtained by introducing a new parameter, the degrees of freedom. The qualities of TRFs are defined in detail, together with results for hiearchical representations, simulation and estimation of model parameters. It is shown that the TRFs are analytically tractable, with several properties equal, or similar, to the GRF. Also, the TRF includes the GRF as a limiting case when the degrees of freedom parameter goes to infinity. The heavy-tailed

effect is observable only cross-realization, and not in-realization, implying that several independent realizations of the random field is needed in order to formally estimate the degree of freedom parameter. Both the TRF and GRF models are estimated on a real well-log observations from the Gullfaks field, and the TRF model appears to be superior to existing Gaussian-based models, as well as Lévy-Stable random field models.

In Part 2 inversion of well log observations and seismic data is performed. The goal is the elastic parameters, P-wave velocity, S-wave velocity and density. The main contribution of this paper is that a spatially continuous approach is used in the inversion, allowing the information to be included in their exact spatial locations, not having to be forced into a grid representation. Also, scaling differences between well log observations and seismic data are implicitely taken care of. The paper is an extension of parts of the work done by Buland and Omre (2003). A priori, the the elastic parameters are defined as a log-Gaussian random field. This results in a posterior distribution for the log of the elastic parameters being approximately a continuous Gaussian random field. Thus, the result of the inversion is analytically available when performing a linearization of the well log observations. As the posterior pdf is given on a spatially continuous form, grid refining is easy. Also, prediction in any arbitrary location can be performed.

In Part 3 the problem of discretizing continuous random fields is adressed. The main contribution of this paper are general formulas for representation of any random field, by any set of basis functions. Assume a reparameterization of a continuous random field through a given set of spatial basis functions, and a corresponding set of random parameters. That is, the continuous spatial random field is split into a continuous spatial part, and a stochastic part. It is demonstrated how the stochastic properties of the continuous random field are inherited by the discrete set of random parameters, given the set of spatial basis functions. The results are demonstrated for different sets of basis functions, both orthogonal and non-orthogonal. Wavelet bases and results for GRFs are studied in more detail. See Walter and Shen (2001) for more on orthogonal systems.

In Paper 4 a data analysis on synthetic production data genereated from a percolation system is performed. The main contribution of this paper is the demonstration of to which extent the knowledge of the breakthrough time increases precision in forecast of oil production. This is done by estimation of the two first moments. Clear trends in the behaviour of the empirical mean and variance, when conditioning on different breakthrough times, are observed. This suggests that there might exist general results being independent of the given percolation system. The two first moments are also introduced into a more formal statistical setting; as parameters in a Gaussian distribution. This can be used to quantitatively compare conditional and unconditional production data, and thus quantifying the increased knowledge of future oil production the breakthrough time provides. See Stauffer and Aharony (1992) for an introduction to percolation theory.

In summary, Parts 1-3 contains the core of the thesis, focusing on near-well description of

reservoirs, and the modelling of elastic material parameters as continuous random fields. Part 2 applies the frequently used continuous GRF model in a Bayesian inversion procedure, resulting in an analytical, continuous GRF posterior distribution for the elastic material parameters given both well log observations and seismic data. However, it is a well-known fact that the GRF model is not capable of explaining all the variability in geophysical measurements. Part 1 introduces a new analytically tractable, continuous random field model with a more heavy-tailed pdf, the TRF model. When working with continuous random fields, at some point a discretization must take place. Part 3 adresses the problem of discretizing continuous random fields given a set of basis function. Results for any random field in general, and GRFs in particular, are provided. Finally, Part 4 is an empirical study of synthetic production data generated from a percolation system. It thus differs from the first three parts in several ways, mainly in fact that the full field is considered, and that continuous random fields is not the main concern.

References

[1] Abrahamsen, P. (1997) *A review of Gaussian random fields and correlation functions*, Technical report 878, Norwegian Computing Centre, Norway

[2] Billingsley, P. (1986) *Probability and Measure*, 2nd edition, John Wiley & Sons, New York

[3] Buland, A. and Omre, H. (2003) Joint AVO Inversion, Wavelet and Noise Level Estimation, *Geophysical Prospecting*, **6**, 531-550

[4] Chilés, J-P. and Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons

[5] Christakos, G. (1992), *Random Field Models in Earth Sciences*, Academic Press Inc., San Diego

[6] Gunning, J. (2002), On the Use of Multivariate Lévy-Stable Random Field Models for Geological Heterogeneity, *Mathematical Geology*, **34**, No.1

[7] Kitanidis, P.K (1996) On the Geostatistical Approach to the Inverse Problem, *Advances in Water Resources*, **19**, 333-342

[8] Kolbjørnsen, O. (2002) *Nonlinear Topics in the Bayesian Approach to Inverse Problems with Applications to Seismic Inversion*, Doctoral Thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim

[9] Sheriff, R.E. and Geldart, L.P. (1995) *Exploration Seismology*, Cambridge

[10] Stauffer, D. and Aharony, A. (1992) *Introduction to Percolation Theory*, Taylor and Francis, London, 2nd edition

[11] Tarantola, A. (1987) *Inverse Problem Theory*, Elsevier Sci., New York

[12] Walter, G.G. and Shen, X. (2001) *Wavelets and Other Orthogonal Systems*, Chapman & Hall/CRC

T-distributed Random Fields: A Parametric Model for Heavy-tailed Random Fields

Jo Røislien
Department of Petroleum Engineering and Applied Geophysics
Norwegian University of Science and Technology

Henning Omre
Department of Mathematical Sciences
Norwegian University of Science and Technology

Abstract

Histograms of observations from spatial phenomena are often found to be more heavy-tailed than Gaussian distributions, which makes the Gaussian random field model unsuited. A T-distributed random field model with heavy-tailed marginal probability density functions is defined. The model is a generalization of the familiar Student-T distribution, and it may be given a Bayesian interpretation. The increased variability appears cross-realizations, contrary to in-realizations, since all realizations are Gaussian-like with varying variance between realizations. The T-distributed random field model is analytically tractable and the conditional model is developed, which provides algorithms for conditional simulation and prediction, so-called T-kriging. The model compares favourably with most previously defined random field models. The Gaussian random field model appears as a special, limiting case of the T-distributed random field model. The model is particularly useful whenever multiple, sparsely sampled realizations of the random field are available. The properties of the T-distributed random field model is demonstrated on well log observations from the Gullfaks field in the North Sea.

URN:NBN:no-3476

1 Introduction

In recent years, a series of papers have been presented on parametric stochastic models for continuous random fields (RF) with non-Gaussian marginal probability density functions (pdf). Gaussian RF (GRF) models are of course most widely used in practice, see Chilès and Delfiner (1999). The normal-score approach makes a univariate ϕ-transform of the random variables (RV) into a Gaussian marginal pdf, for then to assume a GRF model. The normal-score model, ϕ^{-1}-GRF model, is widely used as a model for RF with non-Gaussian marginal pdfs, see Chilès and Delfiner (1999). The advantage of the ϕ^{-1}-GRF model is its simplicity and flexibility, but its disadvantage is the lack of analytical tractability. In the current paper RFs with symmetric, unimodal marginal pdfs, heavy-tailed or not, are considered. In a recent paper, Gunning (2002), the use of Levy-Stable RFs (LSRF) as models for such RFs is discussed. Gunning presents an interesting discussion on desirable features of stochastic models for continuous RFs, and the LSRF is evaluated with respect to these features. The current paper can be seen as an extension of Gunnings work. We elaborate on his list of features, but suggest another class of RF models. Our list of desirable features of parametric stochastic models for continuous RFs is as follows:

- *Fully specified probabilistic model*
  This is a requirement for simulation of the RF. RF models based on contrasts and increments often leave parameters in the trend unspecified, and hence they are unsuited for simulation.

- *Permutation invariance*
  The RF model must be exchangeable in its components. This is a requirement for RF models, see Yaglom (1962).

- *Probabilistic consistency*
  Finite dimensional marginal pdfs of different dimensions defined by the RF model must be such that they coinside when non-common dimensions are integrated out. This is another requirement for RF models, see Yaglom (1962).

- *Marginal invariance*
  Finite dimensional marginal pdfs defined by the RF model should all belong to the same parametric class of pdfs. This makes it easier to ensure probilistic consistency, and simplifies analytical work.

- *Additivity closed*
  Often, observations are collected as averages over sampling volumes, so-called sample support. The objective is frequently to determine spatial averages over a given volume. If spatial averages define an integrated RF which belongs to the same parametric class as the initial RF, this will normally simplify the analytical work.

- *Analytical tractability*
  The objective is usually to determine the conditional RF model conditioned on a set of observations in arbitrary locations. In order to define this conditional RF model, the non-observed dimensions of the RF model must be integrated out. This is very difficult to do in RF models unless the model is analytically tractable. Numerical integration in such high dimensions is virtually impossible. Analytical tractability will also make it easier to define efficient simulation algorithms for the RF.

- *General model*
  It is of course desirable that the parametric RF model is general in the sense that the admissible range of parameter values define a large variety of RF models. It would of course be convenient if marginal pdfs of all kinds could be modeled. Moreover, it would be fine if certain parameter combinations provide familiar RF models, the GRF model in particular.

- *Model parameter inference*
  The parametric RF model is by definition fully specified by a set of parameters. It is of course desireable that the number of parameters is low, and that the parameters are interpretable. Most important, however, is that reliable estimators can be defined based on a set of observations from the RF.

- *Diminishing spatial dependence*
  The random variables in two locations should approach independence as their inter-distance increases. This property is related to the ergodic characteristics of the RF, and will make estimators for model parameters more reliable. Without diminishing spatial dependence one may not be able to define consistent estimators for the model parameters based on observations from one single realization of the RF.

GRF models are by far the most used in practice, and the GRF is normally defined as follows:

Definition 1 ***Gaussian Random Field (GRF)***
A RF $\{Z(x); x \in \mathcal{D} \subset \mathbb{R}^n\}$ *is termed a GRF if*

$$Z = [Z(x_1), \ldots, Z(x_m)]^T \quad \sim \quad N_m(\mu, \Phi)$$

with pdf

$$f(z) \quad = \quad \frac{1}{(2\pi)^{m/2}} |\Phi|^{-1/2} \exp\left\{ -\frac{1}{2}(z-\mu)^T \Phi^{-1} (z-\mu) \right\}$$

for all configurations $(x_1, \ldots, x_m) \in \mathcal{D} \times \ldots \times \mathcal{D}$ *and all* $m \in \mathcal{N}_+$ *where* $N_m(\mu, \Phi)$ *represents the multivariate Gaussian distribution with parameters* (μ, Φ) *of proper dimensions. The*

GRF is parametrized $G_x(\mu_x, \phi_{xx})$ *where*

$$\begin{aligned} \mu_x &: \{\mu(x); x \in \mathcal{D}\} \\ \phi_{xx} &: \{\phi(x', x''); (x', x'') \in \mathcal{D} \times \mathcal{D}\} \end{aligned}$$

with μ_x *the expectation function over* $\mathcal{D}$*, and* ϕ_{xx} *a positive definite covariance function over* $\mathcal{D} \times \mathcal{D}$*.* ∎

The GRF model meets most of the desirable features listed above. General model is the only feature really being violated. The GRF model constitutes an extreme case among RF models in the sense that it maximizes entropy, i.e. it is the smoothest model possible given the model parameters. This is also related to the central limit theorem – everything averages out to the GRF. The GRF model exhibits two unfavorable features in particluar: The Gaussian marginal pdf has extremely light tails, and the prediction variance is only dependent on the location configuration of the observations – not on the values actually observed.

The objective of the current paper is to define a RF model with symmetric, unimodal marginal pdfs with large flexibility in tail behaviour. The T-distributed RF (TRF) model is introduced and its properties explored. The TRF model can also be given a Bayesian interpretation as an extension of the work in Kitanidis (1986), Omre (1987) and Hjort and Omre (1994).

The paper is organized as follows. The next section contains a motivation for the study. Section 3 contains the definition of the TRF and a discussion of its properties. In Section 4 T-kriging is introduced, while Section 5 contains definitions of estimators for the model parameters of the TRF model. A case study on the real data explored in Section 2 is included in Section 6. Lastly, the two last sections contain a discussion of the characteristics of the TRF model and some concluding remarks.

2 Motivation

Consider the density log from a well in the Gullfaks field in the North Sea, see the left display of Figure 1. The well seems to penetrate several horisontal sedimentary layers with varying properties, and a segmentation into layers is made in the right display of Figure 1. The layered structure is assumed to be caused by abrupt changes in the depositional processes. The list of layer averages and empirical variances of the nine layers is presented in Table 1. If a classification of the layers based on the average values is made, layer 1, 2, 3, 5, 7 and 9 will be pooled in one class. One could of course have used empirical variances for further classification, but we have chosen not to do so here, because empirical variances are often very sensitive to individual samples. In Figure 2 the histograms of the

six pooled layers are displayed, and they look fairly dense with light tails. The pooled histogram from the six layers in the left display of Figure 3, however, appears as peaked with relatively heavy tails. In the right display of Figure 3, a kernel-smoothed pooled histogram is presented together with maximum likelihood adapted Gaussian and Student-T pdfs. Note that the Student-T pdf reproduces the histogram much more reliably than the Gaussian one.

To summarize, by compiling relatively dense, light-tailed histograms from the individual layers, one obtains a composite peaked, heavy-tailed histogram. This change in shape appears since the layer-histograms have different variances. Recall that identical centering is ensured by the initial classification. Statisticians are aware that the composite of Gaussian RVs, identically centered with varying variances, may appear as Student-T distributed, se Walpole and Myers (1993). In the current study, the well logs in the individual layers are considered to be GRFs, equally centered but with different variances, which appears as outcomes from a unifying TRF.

The evaluation above is made in a well with numerous observations in each layer. If only a few observations in each layer are available, as for example core plug samples, one may wish to interpolate or simulate along the entire well trace and extend this into the three-dimensional reservoir. The layers may be classified with respect to averages, while further classification based on empirical variances may be considered unreliable. The composite histogram may appear Student-T like, hence a RF model with Student-T marginal pdf would be prefered. Based on this RF model, interpolation and simulation in the individual layers conditional on the few available obervations in that particular layer can be made. This setting constitutes the challenge of the study.

3 T-distributed Random Fields

In order to define a suitable RF model for the data presented in Section 2, a survey of multivariate T-distributed RVs is presented in the first subsection. In the subsequent subsections, TRFs, simple and conditional, are defined.

3.1 Multivariate T-distriubted random variables

The multivariate T-distributed RV is a generalization of the Student-T distributed RV introduced in most introductory courses in statistics. The exposure here is mainly based on results in Cornish (1954), Johnson and Kotz (1972) and Welsh (1996).

Definition 2 ***T-distributed random variable***

A RV $Z \in \mathbb{R}^m$ is multivariate T-distributed

$$Z \sim T_m(\mu, \Omega, \nu)$$

if its pdf is

$$f(z) = \frac{\Gamma(\frac{\nu+m}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{m/2}}|\Omega|^{-1/2}\left[1 + \frac{1}{\nu}(z-\mu)^T\Omega^{-1}(z-\mu)\right]^{-\frac{\nu+m}{2}}$$

where $\Gamma(x)$ is the gamma function, and with $\mu \in \mathbb{R}^m$ a centering vector, $\Omega \in \mathbb{R}^m \times \mathbb{R}^m$ a positive definite scale/dependence matrix, and $\nu \in \mathbb{R}_+$ the degrees of freedom. ∎

This definition specifies a spherical-symmetric pdf centered at μ with Ω controlling scale and multivariate dependence, while ν controls the tail behaviour, see Mardia et al. (1979). In Figures 4 and 5 $T_2(0, \Omega, \nu)$ are displayed for varying values of Ω and ν. Note that all bivariate pdfs have spherical contour lines, but the general shape appear with varying peakedness and heavy-tailedness.

The properties of the multivariate T-distributed RV are summarized below:

Result 1 ***Properties of T-distributed random variables***
Let $Z \in \mathbb{R}^m$ have pdf as in Definition 2, then the following properties can be demonstrated:

1. *Special cases:*

$$\begin{array}{rcl} T_m(\mu, \Omega, \nu) & \stackrel{\nu\to\infty}{\longrightarrow} & N_m(\mu, \Omega) \\ T_m(\mu, \Omega, 1) & = & C_m(\mu, \Omega) \end{array}$$

   where $N_m(\mu, \Omega)$ and $C_m(\mu, \Omega)$ are the multivariate Gaussian and Cauchy distributions, respectively.

2. *Moments:*

$$\begin{array}{rcll} \mathrm{E}\{Z\} & = & \mu & ; \ \nu \geq 2 \\ \mathrm{Cov}\{Z\} & = & \frac{\nu}{\nu-2}\Omega & ; \ \nu \geq 3 \end{array}$$

   while for ν less than the specified values the moments are infinite.

3. *Linear transform:*
   Let A be an arbitrary known $(k \times m)$-matrix, and

$$Z_A = AZ,$$

   then

$$Z_A \sim T_k(A\mu, A\Omega A^T, \nu).$$

   Note that this entails that all marginal pdf s of Z are T-distributed as well.

4. *Conditional distributions:*
   Let Z_A be defined as in Point 3. Then

$$[Z|Z_A = z_A] \sim T_m\left(\mu_{\cdot|z_A}, \Omega_{\cdot|z_A}, \nu + k\right)$$

   where

$$\begin{aligned} \mu_{\cdot|z_A} &= \mu + \Omega A^T (A\Omega A^T)^{-1}(z_A - A\mu) \\ \Omega_{\cdot|z_A} &= \xi(z_A)\left[\Omega - \Omega A^T (A\Omega A^T)^{-1} A\Omega\right] \end{aligned}$$

   with

$$\xi(z_A) = \frac{1}{1+\frac{k}{\nu}}\left[1 + \frac{1}{\nu}(z_A - A\mu)^T (A\Omega A^T)^{-1}(z_A - A\mu)\right]$$

   Note that this entails that all pdfs conditional on arbitrary linear combinations are T- distributed.

5. *Non-independence:*
   Let Z be arbitrarily decomposed as follows:

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix},\ \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},\ \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

   with $Z_1 \in \mathbb{R}^{m_1}$ and $Z_2 \in \mathbb{R}^{m_2}$ and $m = m_1 + m_2$, then

$$T_m(\mu, \Omega, \nu) \neq T_{m_1}(\mu_1, \Omega, \nu) \times T_m(\mu, \Omega, \nu)\ ;\ \nu < \infty$$

   even for the particular case Ω_{12} and Ω_{21} being matrices containing only zeros. Note that Z_1 and Z_2 will not be independent even if Ω is a block matrix for $\nu < \infty$. For $\nu \to \infty$ independence will be obtained in the limit for this case, since then the multi-Gaussian case is reached.

6. *Decomposition:*
   Let H and U be independent RVs, then

$$Z = \mu + H^{-\frac{1}{2}}\Omega^{1/2}U$$

   with

$$\begin{aligned} \nu H &\sim \chi^2(\nu) \\ U &\sim N_m(0, I_m) \end{aligned}$$

   where $\chi^2(\nu)$ is a univarite chi-squared distribution with ν degrees of freedom and I_m is a $(m \times m)$ unit-diagonal matrix.

7. *Hierarchical representation:*
Let Z conditional on the random parameters (β, ϕ^2) be

$$[Z|\beta, \phi^2] \sim N_m(G\beta, \phi^2\Phi_0^Z)$$

where G is an arbitrary known $(m \times k)$ matrix, β is a random $(k \times 1)$ vector, ϕ^2 is a univariate RV and Φ_0^Z is a known positive definite $(m \times m)$ correlation matrix, with

$$\begin{aligned} [\beta|\phi^2] &\sim N_k\big(\mu_\beta, \phi^2\Phi_0^\beta\big) \\ \phi^2 &\sim IG\left(\frac{\nu}{2}, \frac{\nu\omega^2}{2}\right) \end{aligned}$$

where μ_β and Φ_0^β are expectation vector and correlation matrix of appropriate dimensions, and $IG(\kappa, \lambda)$ represents the inverse gamma pdf

$$f(\phi^2) = \frac{1}{\Gamma(\kappa)}\lambda^\kappa \left(\frac{1}{\phi^2}\right)^{\kappa+1} e^{-\frac{\lambda}{\phi^2}} \; ; \; \kappa, \lambda > 0, \phi^2 > 0.$$

Then

$$Z \sim T_m(\mu, \Omega, \nu)$$

where

$$\begin{aligned} \mu &= G^T\mu_\beta \\ \Omega &= \omega^2\Omega_0 = \omega^2[\Phi_0^Z + G^T\Phi_0^\beta G]. \end{aligned}$$

Note that this representation can be interpreted in a Bayesian setting with (β, ϕ^2) being random hyperparameters.

3.2 T-distributed random field

The TRF is defined by the multivariate T-distributed RVs along the lines of the GRF in Definition 1.

Definition 3 ***T-distributed Random Field (TRF)***
A RF $\{Z(x); x \in \mathcal{D} \subset \mathbb{R}^n\}$ is termed a TRF if

$$Z = [Z(x_1), \ldots, Z(x_m)]^T \sim T_m(\mu, \Omega, \nu)$$

for all configurations $(x_1, \ldots, x_m) \in \mathcal{D} \times \cdots \times \mathcal{D}$ and all $m \in \mathcal{N}_+$ where $T_m(\mu, \Omega, \nu)$ represents the multivariate T-distribution with parameters (μ, Ω, ν). The TRF is parametrized $T_x(\mu_x, \omega_{xx}, \nu)$ where

$$\begin{aligned} \mu_x &: \{\mu(x); x \in \mathcal{D}\} \\ w_{xx} &; \{\omega(x', x''); (x', x'') \in \mathcal{D} \times \mathcal{D}\} \end{aligned}$$

with μ_x the centering function over $\mathcal{D}$ and ω_{xx} a positive definite scale/dependence function over $\mathcal{D} \times \mathcal{D}$, and ν the degrees of freedom. ∎

Note that from the definition of TRFs and Result 1.1 it is obvious that a TRF tends towards a GRF whenever $\nu \to \infty$. T-distributed RVs are in Result 1.3 shown to be closed under linear operations. This entails that a differential TRF, if it exists, also will be a TRF. The same holds for an integrated TRF.

In Figures 6 through 9 realizations of $T_x(0, \omega_{xx}, \nu)$ with varying ν are displayed. The scale/dependence function ω_{xx} has form

$$\omega(x', x'') = \exp\left\{-\alpha(x'' - x')^2\right\}$$

with $\alpha = \frac{1}{50}$, and ν takes values 1, 3, 7, and ∞. Note that the variability seems to decline as ν increases and the TRF approaches a GRF. The effect of Result 1.7 can be observed by each realization being Gaussian-like, with large cross-realization variability.

The hierarchical representation of T-distributed RVs in Result 1.7 can be used to define:

Result 2 ***Hierarchical representation of TRF***
Let $\{Z(x); x \in \mathcal{D}\}$ conditoinal on the random parameters (β, ϕ^2) be:

$$\{[Z(x)|\beta, \phi^2]; x \in \mathcal{D}\} \sim G_x(g_x^T\beta, \phi^2\phi^Z_{0xx})$$

where

$$\begin{aligned} g_x &: \{g(x) = \big(g_1(x), \ldots, g_k(x)\big)^T; x \in \mathcal{D}\} \\ \phi^Z_{0xx} &: \{\phi^Z_0(x', x''); x', x'' \in \mathcal{D} \times \mathcal{D}\} \end{aligned}$$

with g_x and ϕ^Z_{0xx} known trend functions and spatial correlation functions respectively, and β a random $(k \times 1)$ vector and ϕ^2 a univariate random variance. Further, let

$$\begin{aligned} [\beta|\phi^2] &\sim N_k\left(\mu_\beta, \phi^2\Phi_0^\beta\right) \\ \phi^2 &\sim IG\left(\frac{\nu}{2}, \frac{\nu\omega^2}{2}\right) \end{aligned}$$

where μ_β and Φ_0^β are expectations and correlation matrix of appropriate dimensions, then

$$\{Z(x); x \in \mathcal{D}\} \sim T_x(\mu_x, \omega_{xx}, \nu)$$

with

$$\begin{aligned} \mu_x &: \{\mu(x) = g(x)^T\mu_\beta; x \in \mathcal{D}\} \\ \omega_{xx} &: \{\omega(x', x'') = \omega^2\omega_0(x', x'') \\ &\qquad = \omega^2[\phi^Z_0(x', x'') + g(x')^T\Phi_0^\beta g(x'')]; x', x'' \in \mathcal{D} \times \mathcal{D}\} \end{aligned}$$

∎

This hierarchical represenation of the TRF can be given a Bayesian interpretation where the model paramters (β, ϕ^2) in a GRF is assigned appropriate prior models. Note that this corresponds to the Bayesian kriging models discussed in Kitanidis (1986), Omre (1987), Le and Zidek (1991), Handcock and Stein (1993) and Hjort and Omre (1994).

The hierarchical representation in Result 2 can be used for simulation of TRFs, and a suitable algorithm is as follows

Algorithm 1 ***Simulation of TRF***
Simulate TRF $\{Z(x); \in \mathcal{D}\} \sim T_x\Big(g_x^T\mu_\beta, \omega^2[\phi^Z_{0xx} + g_x^T\Phi_0^\beta g_x], \nu\Big)$ *by*

- *generate* ϕ^2 *from* $IG(\frac{\nu}{2}, \frac{\nu\omega^2}{2})$
- *given* ϕ^2*, generate* β *from* $N_k(\mu_\beta, \phi^2\Phi_0^\beta)$
- *given* (β, ϕ^2)*, generate*

$$\{Z(x); x \in \mathcal{D}\} \sim G_x(g_x^T\beta, \phi^2\phi^Z_{0xx})$$

This algorithm can of course be very efficient since β and ϕ^2 are low-dimensional, and any fast algorithm for simulation of GRFs can be used.

3.3 Conditional T-distributed random field

Consider a TRF $\{Z(x); x \in \mathcal{D}\}$ and let $Z^d = [Z(x_1), \ldots, Z(x_n)]^T$ be a set of observations in arbitrary locations $(x_1, \ldots, x_n) \in \mathcal{D} \times \ldots \times \mathcal{D}$ with associated realization $z^d = (z_1, \ldots z_n)^T$. The conditional TRF is denoted $\{[Z(x)|z^d]; x \in \mathcal{D}\}$. From Result 1.4 the conditional TRF is given as

Result 3 ***Conditional TRF***
A TRF $\{Z(x); x \in \mathcal{D}\} \sim T_x(\mu_x, \omega_{xx}, \nu)$ *conditioned on the* $(n \times 1)$*-dimensional vetor of observations* $Z^d = z^d$ *is:*

$$\{[Z(x)|z^d]; x \in \mathcal{D}\} \sim T_x\left(\mu_{x|z^d}, \omega_{xx|z^d}, \nu + n\right)$$

where

$$\begin{aligned}
\mu_{x|z^d} &: \left\{[\mu(x)|z^d] = \mu(x) + \omega_{xd}^T\Omega_{dd}^{-1}(z^d - \mu_d); x \in \mathcal{D}\right\} \\
\omega_{xx|z^d} &: \left\{[\omega(x', x'')|z^d] = \xi(z^d)\left[\omega(x', x'') - \omega_{x'd}^T\Omega_{dd}^{-1}\omega_{x''d}\right]; (x', x'') \in \mathcal{D} \times \mathcal{D}\right\}
\end{aligned}$$

with

$$\begin{aligned}
\mu_d &= (\mu(x_1), \ldots, \mu(x_n))^T \\
\omega_{xd} &= (\omega(x, x_1), \ldots, \omega(x, x_n))^T \\
\Omega_{dd} &= \begin{bmatrix} \omega(x_1, x_1) & \cdots & \omega(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \omega(x_n, x_1) & \cdots & \omega(x_n, x_n) \end{bmatrix} \\
\xi(z^d) &= \frac{1}{1+\frac{n}{\nu}} \left[1 + \frac{1}{\nu} (z^d - \mu_d)^T \Omega_{dd}^{-1} (z^d - \mu_d) \right]
\end{aligned}$$

∎

Note that this entails that all conditional TRFs are TRFs themselves. The conditional TRF also has a hierarchical representation that follows from Result 2:

Result 4 ***Hierarchical representation of conditional TRF***
Let $\left\{ [Z(x)|z^d]; x \in \mathcal{D} \right\}$ *conditional on the random parameters* (β, ϕ^2) *be:*

$$\left\{ [Z(x)|z^d, \beta, \phi^2]; x \in \mathcal{D} \right\} \sim G_x \left(\mu_{x|z^d,\beta}, \phi^2 \phi^Z_{0xx|z^d,\beta} \right)$$

where $\mu_{x|z^d,\beta}$, ϕ^2 *and* $\phi^Z_{0xx|z^d,\beta}$ *are the conditional expectation function, the variance and the conditional correlation function of a* $G_x(g_x^T\beta, \phi^2\phi^Z_{0xx})$. *Otherwise the notation is as in Result 2. Further, let*

$$\begin{aligned}
[\beta|z^d, \phi^2] &\sim N_k \left(\mu_{\beta|z^d}, \phi^2 \Phi^\beta_{0|z^d} \right) \\
[\phi^2|z^d] &\sim IG \left(\eta_{1|z^d}, \eta_{2|z^d} \right)
\end{aligned}$$

where $\mu_{\beta|z^d}$ *and* $\Phi^\beta_{0|z^d}$ *are the conditional expectation and correlation matrix of a* $N_k \left(\mu_\beta, \phi^2 \Phi^\beta_0 \right)$, *while* $\eta_{1|z^d}$ *and* $\eta_{2|z^d}$ *are the conditional parameters of a* $IG(\eta_1, \eta_2)$ *with* $\eta_1 = \frac{\nu}{2}$ *and* $\eta_2 = \frac{\nu\omega^2}{2}$. *Then*

$$\left\{ [Z(x)|z^d]; x \in \mathcal{D} \right\} \sim T_x \left(\mu_{x|z^d}, \omega_{xx|z^d}, \nu + n \right)$$

with

$$\begin{aligned}
\mu_{x|z^d} &: \left\{ [\mu(x)|z^d] = g(x)^T \mu_\beta + \omega_{xd}^T \Omega_{dd}^{-1} [z^d - G_d^T \mu_\beta]; x \in \mathcal{D} \right\} \\
\omega_{xx|z^d} &: \left\{ [\omega(x', x'')|z^d] \right. \\
&\quad \left. = \xi(z^d) \left[\omega(x', x'') - \omega_{x'd}^T \Omega_{dd}^{-1} \omega_{x''d} \right]; (x', x'') \in \mathcal{D} \times \mathcal{D} \right\}
\end{aligned}$$

where

$$\begin{aligned} G_d &= \big(g(x_1),\ldots,g(x_n)\big) \\ \omega_{xx} &: \big\{\omega(x',x'') = \omega^2\big[\phi_0^Z(x',x'') + g(x')^T\Phi_0^\beta g(x'')\big]; (x',x'') \in \mathcal{D}\times\mathcal{D}\big\} \\ \xi(z^d) &= \frac{1}{1+\frac{n}{\nu}}\left[1+\frac{1}{\nu}(z^d - G_d^T\mu_\beta)^T\Omega_{dd}^{-1}(z^d - G_d^T\mu_\beta)\right] \end{aligned}$$

otherwise the notation is as in Result 2 and 3. The exact expressions are given in Appendix A. ∎

This hierarchical representation can of course be seen as the posterior model in a Bayesian setting. Moreover, this representation provides a suitable simulation algorithm for conditional TRFs:

Algorithm 2 ***Simulation of conditional TRF***
Simulate conditional TRF $\{[Z(x)|z^d]; x \in \mathcal{D}\} \sim T_x(\mu_{x|z^d}, \omega_{xx|z^d}, \nu+n)$ *with notation as in Result 4 by*

- *given* z^d*, generate* ϕ^2 *from* $IG\left(\xi_{1|z^d}, \xi_{2|z^d}\right)$
- *given* z^d *and* ϕ^2*, generate* β *from* $N_k\left(\mu_{\beta|z^d}, \phi^2\Phi^\beta_{0|z^d}\right)$
- *given* z^d *and* (ϕ^2,β)*, generate* $\{Z(x); x\in\mathcal{D}\} \sim G_x\left(\mu_{x|z^d,\beta}, \phi^2\phi^z_{0xx|z^d,\beta}\right)$

This algorithm can of course be very efficient, since β and ϕ^2 are low-dimensional, and any fast algorithm for simulation of conditional GRFs can be used.

4 Prediction in T-distributed Random Fields

A predictor for $Z(x_+)$ in an arbitrary location $x_+ \in \mathcal{D}$ based on the observations in $Z^d = [Z(x_1),\ldots,Z(x_n)]^T$ with $(x_1,\ldots,x_n) \in \mathcal{D}\times\ldots\mathcal{D}$ having realizations $z^d = (z_1,\ldots,z_n)^T$ can be obtained from the definition of TRF in Definition 3 and Result 1.4. It is reasonable to term this predictor the T-kriging predictor.

Definition 4 ***T-kriging predictor***
A RV $Z(x_+)$ *from a TRF* $\{Z(x); x\in\mathcal{D}\} \sim T_x(\mu_x, \omega_{xx}, \nu)$ *can be predicted from the* $(n\times 1)$*-dimensional vector of observations* $Z^d = z^d$ *by:*

$$[Z(x_+)|z^d] \sim T_1\left([\mu(x_+)|z^d], [\omega(x_+)|z^d], \nu+n\right)$$

where

$$\begin{aligned}\left[\mu(x_+)|z^d\right] &= \mu(x_+) + \omega_{x_+d}^T \Omega_{dd}^{-1}(z^d - \mu_d) \\ \left[\omega(x_+)|z^d\right] &= \left[\omega(x_+, x_+)|z^d\right] \\ &= \xi(z^d)\left[\omega(x_+, x_+) - \omega_{x_+d}^T \Omega_{dd}^{-1} \omega_{x_+d}\right]\end{aligned}$$

with notation as in Result 3. From Result 1.2 one has

$$\begin{aligned}\mathrm{E}\{Z(x_+)|z^d\} &= \left[\mu(x_+)|z^d\right] \quad \text{for} \quad \nu + n \geq 2 \\ \mathrm{Var}\{Z(x_+)|z^d\} &= \frac{\nu + n}{\nu + n - 2}\left[\omega(x_+)|z^d\right] \quad \text{for} \quad \nu + n \geq 3\end{aligned}$$

∎

Note that the conditional expectation predictor, which is optimal under squared error loss, when it exists, is linear in z^d, and coinsides with the traditional kriging predictors. The prediction variance, however, is also dependent on z^d. This is contrary to traditional kriging where the prediction variance is dependent on the location configuration $(x_1, \ldots, x_n)$ only.

Note further that when the number of conditioning observations increase, i.e. $n \to \infty$, the T-kriging predictor tends towards the traditional kriging predictor, since the degrees of freedom in the predictor pdf tends towards infinity which entails Gaussianity. Broadly spoken, the T-kriging predictor takes into account the fact that the variance of the RF is unknown and must be estimated. In traditional kriging the estimation uncertainty of the variance is ignored. These considerations have a parallel in the use of Student-T distributions in traditional statistical inference.

5 Parameter Estimation in T-distributed Random Fields

In this section a TRF $\{Z(x); x \in \mathcal{D}\} \sim T_x\left(g_x^T \mu_\beta, \omega^2[\phi_{0xx}^Z + g_x^T \Phi_0^\beta g_x], \nu\right)$ with notation as in Result 2 will be considered. Focus is on model parameter estimation. Assume that several independent outcomes of this TRF exists: $\{Z_i(x); x \in \mathcal{D}\}$; $i = 1, \ldots, m$. In each of these outcomes a number of observations are made $Z_i^d = [Z_i(x_{i1}), \ldots, Z_i(x_{in_i})]^T$; $i = 1, \ldots, m$ with realizations $z_i^d = (z_{i1}, \ldots, z_{in_i})^T$. The objective is to estimate the model parameters $(\mu_\beta, \omega^2, \phi_{0xx}^Z, \Phi_0^\beta, \nu)$ from these observations. The spatial correlation function ϕ_{0xx}^Z is particularly difficult to determine and it is assumed known in the exposition below. Procedures for estimating ϕ_{0xx}^Z may be developed based on estimation of spatial correlation functions in GRF models, see Chilès and Delfiner (1999).

It is of course possible to specify the full likelihood function based on the definition of the TRF and perform maximum likelihood estimation. It turns out that the parameter ν,

which defines the tail behaviour of the marginal pdf of the TRF, is particularly hard to determine by this approach. This is unfortunate, since this tail behaviour is of primary interest in the study. An alternative estimation procedure, termed hierarchical maximum likelihood, is recommended here, since it appears to provide more stable estimates of ν in particular.

The hierarchical maximum likelihood estimator is based on the hierarchical representation of TRF in Result 2. It draws on the property that given (β, ϕ^2) the individual outcomes of the TRF will be GRFs. Hence, based on the m outcomes, (β_i, ϕ_i^2); $i = 1, \ldots, m$ can be estimated by maximum likelihood estimators. The parameters of interest $(\mu_\beta, \omega^2, \Phi_0^\beta, \nu)$ can thereafter be estimated by maximum likelihood estimators based on $(\beta_i|\phi_i^2)$ and $\phi_i^2; i = 1, \ldots, m$ and their pdfs as specified in Result 2.

The likelihood function for (β_i, ϕ_i^2); $i = 1, \ldots, m$ is

$$L(\beta_1, \ldots, \beta_m, \phi_1^2, \ldots, \phi_m^2) = \prod_{i=1}^{m} \frac{1}{(2\pi\phi_i^2)^{n_i/2}} \left|\Phi_{0d_id_i}^Z\right|^{-1/2} \exp\left\{-\frac{1}{2\phi_i^2} M_i^d\right\}$$

with

$$M_i^d = (z_i^d - G_{d_i}^T\beta_i)^T(\Phi_{0d_id_i}^Z)^{-1}(z_i^d - G_{d_i}^T\beta_i)$$

where $G_{d_i} = \big(g(x_{i1}), \ldots, g(x_{in_i})\big)$ is a known $(k \times n_i)$ matrix and $\Phi_{0d_id_i}^Z$ is a $(n_i \times n_i)$-correlation matrix defined by ϕ_{0xx}^Z. The corresponding maximum likelihood estimates are

$$\begin{aligned}
\hat{\beta}_i &= [G_{d_i}(\Phi_{0d_id_i}^Z)^{-1}G_{d_i}^T]^{-1}G_{d_i}(\Phi_{0d_id_i}^Z)^{-1}z_i^d \;;\; i = 1, \ldots, m \\
\hat{\phi}_i^2 &= \frac{1}{n_i}\hat{M}_i^d \;;\; i = 1, \ldots, m
\end{aligned}$$

with $\hat{M}_i^d$ being M_i^d where $\hat{\beta}_i$ is substituted for β_i. The likelihood function for $(\mu_\beta, \Phi_0^\beta)$ is

$$\begin{aligned}
L(\mu_\beta, \Phi_0^\beta) &= \prod_{i=1}^{m} \frac{1}{(2\pi\hat{\phi}_i^2)^{n_i/2}} |\Phi_0^\beta|^{-1/2} \\
&\quad \cdot \exp\left\{-\frac{1}{2\hat{\phi}^2}(\hat{\beta}_i - \mu_\beta)^T(\Phi_0^\beta)^{-1}(\hat{\beta}_i - \mu_\beta)\right\}
\end{aligned}$$

By maximizing this likelihood function, one obtains the following estimators:

$$\begin{aligned}
\hat{\mu}_\beta &= \frac{1}{\sum_{i=1}^m \frac{1}{\hat{\phi}_i^2}} \cdot \sum_{i=1}^{m} \frac{\hat{\beta}_i}{\hat{\phi}_i^2} \\
\hat{\Phi}_0^\beta &= \frac{1}{m}\sum_{i=1}^{m} \frac{1}{\hat{\phi}_i^2}(\hat{\beta}_i - \hat{\mu}_\beta)(\hat{\beta}_i - \hat{\mu}_\beta)^T
\end{aligned}$$

Finally, the likelihood function for (ω^2, ν) is

$$L(\omega^2, \nu) = \prod_{i=1}^{m} \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{\nu\omega^2}{2}\right)^{\nu/2} \left(\frac{1}{\hat{\phi}_i^2}\right)^{\frac{\nu+2}{2}} \exp\left\{-\frac{\nu\omega^2}{2\hat{\phi}_i^2}\right\},$$

and maximizing provides the following expressions, defining the maximum likelihood estimators;

$$\begin{aligned} \hat{\omega}^2 &= \left[\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\hat{\phi}_i^2}\right]^{-1} \\ \frac{\Gamma'(\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})} - \ln\frac{\nu}{2} &= \ln\left[\prod_{i=1}^{m}\frac{1}{\hat{\phi}_i^2}\right]^{1/m} + \ln\hat{\omega}^2 \end{aligned}$$

where $\Gamma(x)$ and $\Gamma'(x)$ is the gamma function and its derivate, respectively. The last equation must be solved numerically in order to define the estimate of ν.

Properties of maximum likelihood estimators in RF in general are discussed in Stein (1999). For this particular hierarchical maximum likelihood aproach one has: $\hat{\beta}_i$ and $\hat{\phi}_i^2$ are consistent estimators in the sense that $n_i \to \infty$ by expanding $\mathcal{D} \to \mathbb{R}^n$, then $\hat{\mu}_\beta$, $\hat{\Phi}_0^\beta$, $\hat{\omega}^2$ and $\hat{\nu}$ are consistent estimators in the sense that $m \to \infty$. Note in particular that both $n_i \to \infty$ by expanding $\mathcal{D}$ and $m \to \infty$ are required to ensure consistent estimators for the model parameters of interest.

One interesting case appears if $\hat{\phi}_i^2; i = 1, \ldots, m$ all turn out identical to for example $\hat{\phi}^2$. Then it is easy to show that $\hat{\omega}^2 = \hat{\phi}^2$ and $\hat{\nu} = \infty$. This entails that the inferred TRF coinsides with a GRF. Hence the degrees of freedom, ν, is related to the dispersion in $\hat{\phi}_i^2; i = 1, \ldots, m$.

One other interesting case appears when only one realization is available, i.e. $m = 1$. Hence $z_1^d = (z_{11}, \ldots, z_{1n_1})^T$ is the only available data vector. With $n_1 \geq 2$ estimates of β_1 and ϕ_1^2 can be obtained, and the estimators are consistent in the sense specified above. The estimate of μ_β is obtainable while Φ_0^β is left unspecified. Assume for the moment that β is a constant vector μ_β which makes Φ_0^β irrelevant. The interesting feature, however, is that ω^2 and ν are estimated to $\hat{\phi}_1^2$ and ∞. This entails that the inferred TRF collapses into a GRF, since they coinside for $\nu = \infty$. Consequently, one cannot make inference about TRFs with observations from one realization of the RF only. This is related to the fact that TRFs do not exhibit diminishing spatial dependence, and the heavy-tailedness only appear cross-realizations and not in-realizations. Recall that all realizations appear GRF-like with parameter values varying between realizations.

6 Case Study

Consider the density log from the well in the Gullfaks field in the left display in Figure 1. Assume that the log has only been sampled every third meter, while the layer geometry has been determined from elsewhere, see Figure 10. The layers are classified with respect to the average value in each layer, which results in layer 1, 2, 3, 5, 7 and 9 being pooled, as in Section 2. The variance estimates, however, are highly unreliable when based on so few samples in each layer, and are therefore not used for classification. The composite histogram of all samples in the pooled layers is displayed in the left display in Figure 11.

Two alternative modelling approaches can be seen. In the traditional GRF modelling approach, the layers are considered to be realizations from one common GRF model. The variances in all layers are considered to be identical, and the differences in empirical variances are tributed to the sampling variance. The compiled set of observations can of course be used for inference about the parameters of the GRF model. Alternatively, a TRF modelling approach can be used, where the layeres are considered to be realizations from a unifying TRF model. This approach gives room for different variances in the individual layers. Also in this case, the compiled set of observations can be used for model parameter inference. Consequently, the difference in the two approaches is the interpretation of the varying empirical variance in the layers. Note further that the TRF model contains the GRF model as a limiting case, and hence the former can be seen as a generalization of the latter.

In this case study both the GRF and the TRF approach will be used and the results based on each of them compared. The two models are $G_x(\mu, \phi^2\phi_{0xx})$ and $T_x(\mu, \omega^2\omega_{0xx}, \nu)$ respectively, with μ a constant level. The spatial correlation function ϕ_{0xx} and the spatial dependence function ω_{0xx} will with constant μ have similar interpretation, see Result 2, and they are assumed known to be

$$\phi_{0xx} = \omega_{0xx} = \varrho(x', x'') = \exp\{-\alpha(x'' - x')\}$$

with $\alpha = 1/0.65$, which is in good agreement with the data. The maximum likelihood estimators based on the GRF model provides $\hat{\mu} = 2018$ and $\hat{\phi}^2 = 6047$. Alternatively, the TRF model provides hierarchical maximum likelihood estimates according to Section 5: $\hat{\mu} = 2018$, $\hat{\omega}^2 = 1383$ and $\hat{\nu} = 1.71$. The kernel smoothed histograms of the compiled observations and the inferred marginal pdfs from the two models are presented in the right display of Figure 11. The TRF model seems to reproduce the observations better than the GRF model.

Layers 2 and 3 are studied in more detail under the two competing models, since they constitute the extremes with respect to variability. Each of the layers are considered as independent realizations of the respective models, and they are evaluated under the two models inferred above. In Figure 12 conditional realizations in the two layers under

the inferred GRF and TRF model respectively, are presented. Under the GRF model the inferred variance is considered to be the true one, and hence not adapted to the observations in the layer under study. In layer 2 this entails that the variablity in the conditioning observations completely dominates the variability of the model. The opposite effect is observed in layer 3, since the variability in the observations is less. Under the TRF model simulation is made according to Algorithm 2. The variability of the model is adaptive to the value of the available observations and the realizations look much more trustworthy. To summarize, the adaptivity of the variance in the conditional TRF model make the realization more realistic than under the GRF model with non-adaptive variance.

In Figure 10 two non-observed locations are marked, one in each of the layers. In Figure 13 the predictive conditional pdfs in these two locations under the inferred GRF and TRF model respectively, are displayed. Under the GRF model the simple kriging predictor with inferred model parameters is used. The conditional pdf is Gaussian with expectation dependent on the value of the observations and variance dependent on the location of the observations only. Hence the prediciton variances in the two layers appear very similar. Under the TRF model the prediction is made according to the results in Section 4. The conditional pdf is T-distributed with expectations coinsiding with the expectations under the GRF model. The prediction variances, however, are adaptive to the actual values of the observations in the layer under study. Hence the prediction variance in layer 2 is much larger than in layer 3. Moreover, the degrees of freedom in the conditional T-distribution increases with the number of observations, since the estimation of the variance is accounted for. Hence the conditional pdf in layer 3 is more Gaussian-like than the one in layer 2. To summarize, the adaptive prediction variance in the TRF model makes the predictive pdfs more realistic than under the GRF model, where the variability is averaged over the layers. Moreover, the prediction is T-distributed under the TRF model, and accounts for the fact that the variance is unknown and estimated. Under the GRF model, the estimated variance is considered to be the true one, and the esimation effect is ignored.

One may argue that the empirical variance obtained from the observations in layers 2 and 3 are so different that the layers should not be pooled when making inference of the GRF and TRF models. It may be so, but in order to evaluate some of the really thin layers, some pooling must be made in order to infer a RF model. Then the effect demonstrated above will appear, although not as dramatic as seen here. If each layer is considered individually and inference of a GRF model is done, the results will be overly optimistic because the uncertainty in estimating the variance is not accounted for.

7 Characteristics of T-distributed Random Fields

The motivating example in Section 2 demonstrates that a RF with heavy-tailed marginal pdfs may appear from a composite of RFs with light-tailed marginal pdfs with similar

expectations but varying variances. The individual RFs may be GRF with identical expectations, but varying variances, and this may be modeled by a unifying TRF.

The TRF model has most of the favourable properties for parametric models listed in Section 1. The model is fully specified; it exhibits permutation invariance and probabilistic consistency; it has marginal invariance; is additivity closed and it is analytically tractable. The TRF model is a fairly general RF model since it can reproduce a large class of RFs with symmetric, unimodal marginal pdfs, heavy-tailed or not. Moreover, the GRF model is a limiting case of the TRF model, while the Cauchy RF model apperas as a special case. The generality of the TRF model has clear limitations, however, since RFs with neither skewed nor multimodal marginal pdfs can be reproduced. Model parameter inference is somewhat complicated, since consistent estimators can only be defined from sets of observations from several realizations of the TRF. If observations from only one single realization are available, the inferred TRF model collapses into a GRF model. The TRF model does not exhibit diminishing spatial dependence, and this is the reason for lack of consistency of the estimators based on observations in one realization only.

The TRF model compares favourably with the GRF model. The fact that the GRF is a special, limiting case of the TRF model emphasizes this. The TRF model can be seen as a composite of GRF models with identical expectation but varying variance. The heavy tails of the marginal pdf of TRF models are caused by cross-realization variability and not in-realization variability, since all realizations are GRF-like with varying variance parameters from one realization to the other. For GRF models the variance parameter is a fixed constant. If sets of observations from several realzations are available, assuming the more general TRF model will increase flexibility. Inference under the TRF and GRF models coinside if observations from only one realization is available. Conditional TRF models will have both expectations and variances which depend on the values of the observations, while conditional GRF models will have only expectations depending on these values. The corresponding variance is only dependent on location of these observations. The TRF model accounts for the fact that the model variance is largely unknown and must be estimated, while the GRF model considers the estimated variance to be the true value. Consequently, prediction variances will generally be larger under the TRF model than under the GRF model. To summarize, by consequently using a TRF model one will obtain the corresponding GRF model when that is appropriate and otherwise benefit from the additional flexibility of the TRF model. Larger mathematical complexity constitutes the down-side of this approach. Moreover, by assuming a GRF model, extensions are sometimes easier to make, for example when adding measurement error to the observations.

The TRF model compares favorably with ϕ^{-1}-GRF models for RFs, see Chilès and Delfiner (1999), with symmetric, unimodal marginal pdfs. Note, however, that ϕ^{-1}-GRFs can be used for a larger class of RF although the reliability of the results is hard to judge. The ϕ^{-1}-GRF model is generally not additivity closed and it lacks analytical tractability, hence most results must be obtained through simulation or approximations. Lastly, strong assumptions

about spatial stationarity of the RF are required in order to make inference about the ϕ-transformation. Consequently, the TRF model is considered as favorable to the ϕ^{-1}-GRF model when they both apply.

The TRF model compares favorably with LSRF models as defined in Gunning (2002). Both models apply for RFs with symmetric, unimodal marginal pdfs. The LSRF models lack analytical tractability and most results must be obtained through simulation. The fact that the LSRF model is only defined through its characteristic function with no closed form expression for the corresponding pdf demonstrates this lack of analytical tractability. Both the TRF and the LSRF models lack diminishing spatial dependence. This seems to have less consequences for the former than the latter, since the analytical tractability of the TRF model makes it possible to define efficient simulation algorithms even for conditional TRF models.

In practice, the most favourable case for TRF models is in evaluation of RFs where multiple, sparsely sampled realizations are available. Inference of the TRF model is made on the compiled set of observations, and evaluation is done on each realization individually, based on the inferred TRF model. In reservoir evalutaion, several similar geological layers may be penetrated by a well and a limited number of core samples may be collected in each layer. Note further that the TRF model is defined for a reference space of arbitrary dimensions, hence extending the model into three dimensions representing the entire geological layer is simple.

8 Conclusions

Histograms of observations from spatial phenomena are often found to be more heavy-tailed than Gaussian distributions. Well log data in petroleum applications constitute one example. The heavy-tailedness of well log histograms may be caused by the well penetrating several layers with similar well log averages, but varying variances. A parameteric T-distributed random field (RF) model able to capture this heavy-tailedness in the marginal pdf is defined. The model appears as a generalization of the familiar Student-T distribution, and it may be given a Bayesian interpretation. The large variability appears as cross-realization variability, contrary to in-realization variability, since all realizations are Gaussian-like with varying variances between the realizations. This T-distributed RF model exhibits many favourable features: It is a fully defined, valid random field model; it defines a class of RF models closed for linear operators; it is analytically tractable; for RFs with symmetric, unimodal marginal pdfs it is fairly general, and contains the Gaussian RF model as a limiting case; and reliable parameter estimators are defined based on observations from multiple realizations.

An alternative RF model with heavy-tailed marginal pdfs could be imagined. A Gaussian

RF model with the variance being a RF would also exhibit heavy-tailed marginal pdfs. However, we have so far not been able to define a parametric RF model of this type which is analytically tractable.

The T-distributed RF model compares favourably with the Gaussian, ϕ-transformed and Levy-Stable RF models. The model appears as a generalization of Gaussian RF models, and captures heavy-tailed marginal pdfs. It is analytically tractable, which is not the case for neither the ϕ-transform nor the Levy-Stable RF models. Hence, for RFs with symmetric, unimodal marginal pdfs, the T-distributed RF model provides a recommendable alternative.

References

[1] Chilès, J-P. and Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons

[2] Cornish, E.A. (1954), The multivariate t-distribution associated with a set of normal sample eluminates, *Aust. Jour.Phys*, 7, 531-542.

[3] Gunning, J. (2002), On the Use of Multivariate Lévy-Stable Random Field Models for Geological Heterogeneity, *Mathematical Geology*, **34**, No.1

[4] Handcock, M.S. and Stein M.L. (1993), A Bayesian Analysis of Kriging, *Technometrics*, Vol. 35, No.4, November.

[5] Hjort, N.L. and Omre, H. (1994), Topics in spatial statistics, *Scandinavian Journal of Statistics*, Vol 21, 289-357.

[6] Johnson, N.L. and Kotz, S. (1972) *Distributions in Statistics: Continuous Multivariate distributions*, Wiley, New York.

[7] Kitanidis, P.K. (1986), Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resources Research*, **22**, 499-507

[8] Le, N.D. and Zidek, J.V. (1991) *Interpolation With Uncertaion Spatial Covariances: A Bayesian Alternative To Kriging*, Technical report no.113, Department of Statistics, University of British Colombia.

[9] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis*, Academic Press Inc. (London) Ltd.

[10] Omre, H. (1987), Bayesian kriging — merging observations and qualified guesses in kriging, *Math. Geol.*, **19**, 25-39

[11] Stein, M.L. (1999) *Interpolation of Spatial Data - Some Theory for Kriging*, Springer-Verlag New York Inc.

[12] Walpole, R.E and Myers, R.H. (1993) *Probability and Statistics for Engineers and Scientists*, Prentice-Hall

[13] Welsh, A.H. (1996) *Aspects of Statistical Inference*, Wiley, New York

[14] Yaglom, A.M. (1962) *An Introductino to the Theory of Stationary Random Functions*, Dove Publications Inc., New York

Appendix

A Parameters in Hierarchical Conditional TRF

The exact relations defined in Result 4 are defined here:

$$\begin{aligned}
\mu_{x|z^d,\beta} &: \left\{[\mu(x)|z^d,\beta] = g(x)^T\beta + (\phi^Z_{0xd})^T(\Phi^Z_{0dd})^{-1}(z^d - G_d^T\beta); x \in \mathcal{D}\right\} \\
\phi^Z_{0xx|z^d,\beta} &: \left\{[\phi^Z_0(x',x'')|z^d,\beta] = \phi^Z_0(x',x'') - (\phi^Z_{0x'd})^T(\Phi^Z_{0dd})^{-1}\phi^Z_{0x''d}; x',x'' \in \mathcal{D}\times\mathcal{D}\right\}
\end{aligned}$$

with

$$\begin{aligned}
\phi^Z_{0xd} &= \left(\phi^Z_0(x,x_1),\ldots,\phi^Z_0(x,x_n)\right)^T \\
\Phi^Z_{0dd} &= \begin{bmatrix} \phi^Z_0(x_1,x_1) & \cdots & \phi^Z_0(x_1,x_n) \\ \vdots & \ddots & \vdots \\ \phi^Z_0(x_n,x_1) & \cdots & \phi^Z_0(x_n,x_n) \end{bmatrix} \\
G_d &= \left(g(x_1),\ldots g(x_n)\right) \\
\mu_{\beta|z^d} &= \mu_\beta + \Phi^\beta_0 G_d\left[\Phi^Z_{0dd} + G_d^T\Phi^\beta_0 G_d\right]^{-1}(z^d - G^T\mu_\beta) \\
\Phi^\beta_{0|z^d} &= \Phi^\beta_0 - \Phi^\beta_0 G_d\left[\Phi^Z_{0dd} + G_d^T\Phi^\beta_0 G_d\right]^{-1} G_d^T\Phi^\beta_0 \\
\eta_{1|z^d} &= \eta_1 + \frac{n}{2} \\
\eta_{2|z^d} &= \xi(z^d)\left[\eta_2 + \frac{\omega^2 n}{2}\right]
\end{aligned}$$

Table 1: Averages and empirical variances in the layers of Figure 1

layer	average	empirical variance
1	1962	351
2	1986	56402
3	2038	744
4	2331	54659
5	2014	147
6	2361	67417
7	1997	448
8	2554	41636
9	2053	5969

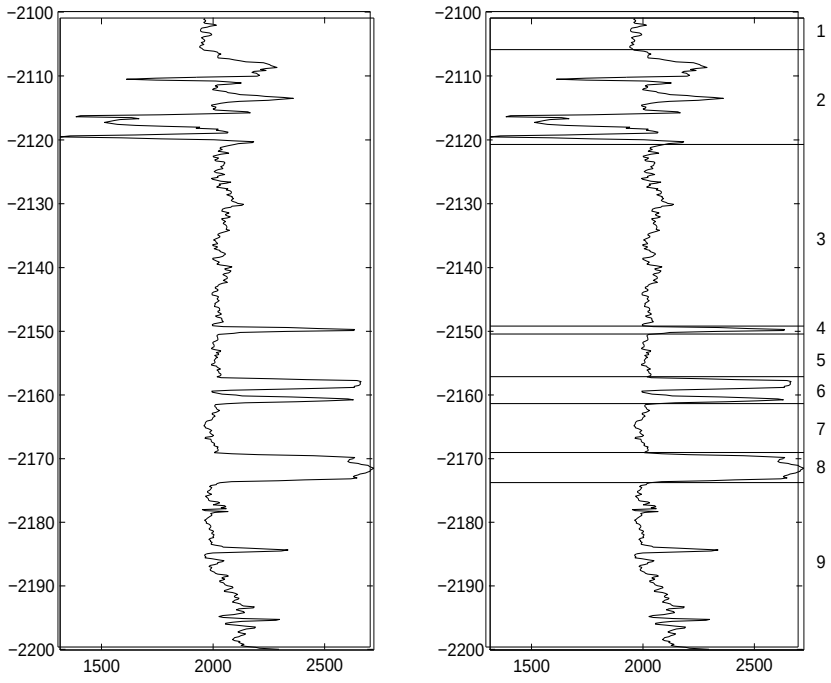

Figure 1: *Left: Bulk density log from well C33 of the Gullfaks field in the North sea. Right: Layer interpretation of the log.*

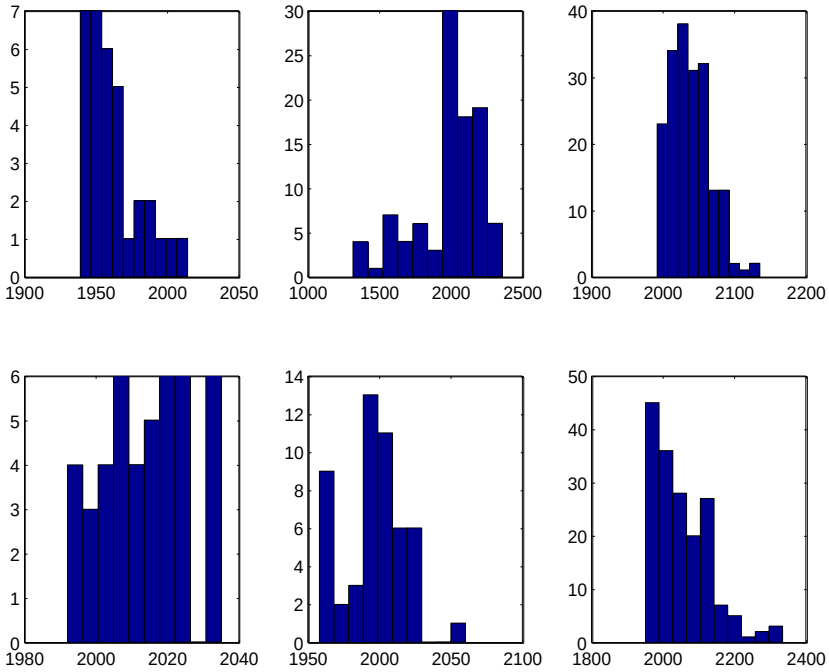

Figure 2: *Histogram of density observations in each of the pooled layers. Upper: Layer 1,2,3. Lower: Layer 5,7,9.*

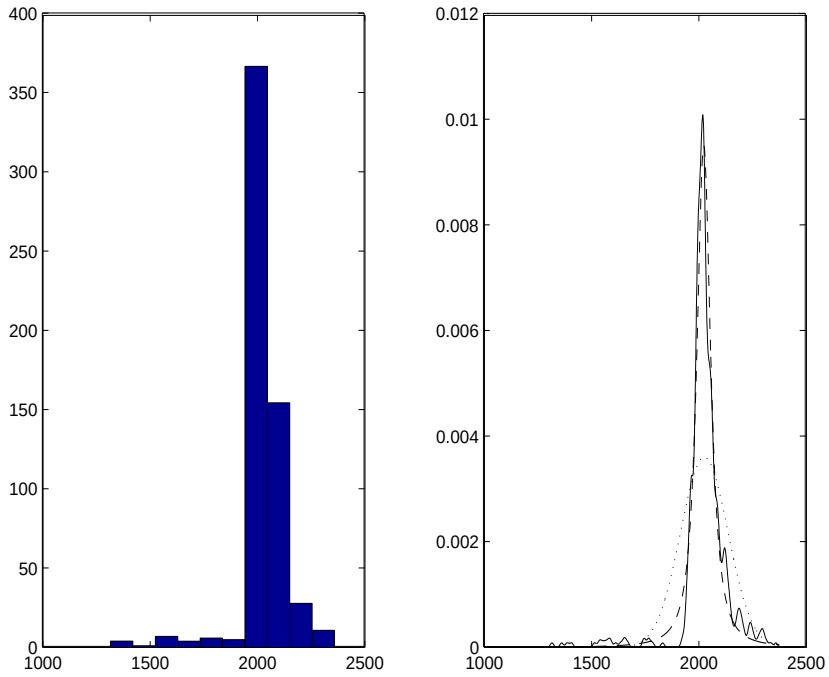

Figure 3: *Left: Histogram of density observations in pooled layers. Right: Kernel estimate of pdf (—); Gaussian estimate of pdf (· · ·); and Student-T estimate of pdf (– – –).*

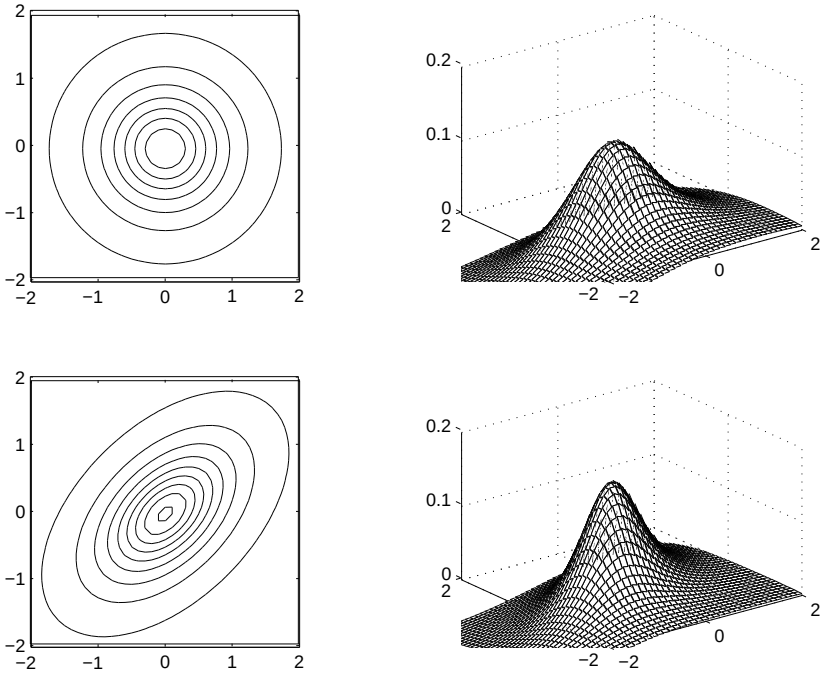

Figure 4: *Upper:* $T_2(0, \Omega, 1)$*-distribution with unit variance and no correlation. Lower:* $T_2(0, \Omega, 1)$*-distribution with unit variance and correlation 0.5.*

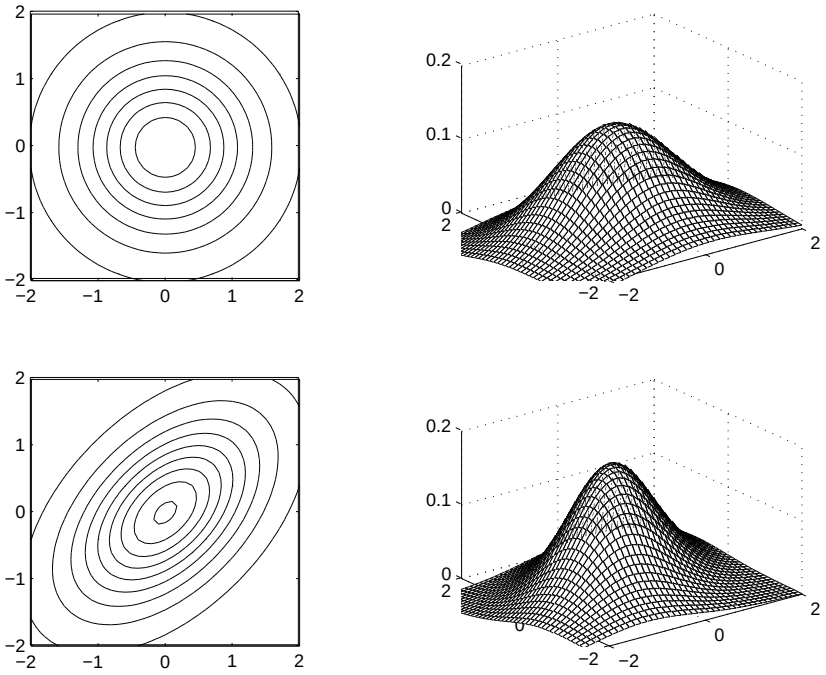

Figure 5: *Upper:* $T_2(0, \Omega, 7)$*-distribution with unit variance and no correlation. Lower:* $T_2(0, \Omega, 7)$*-distribution with unit variance and correlation 0.5.*

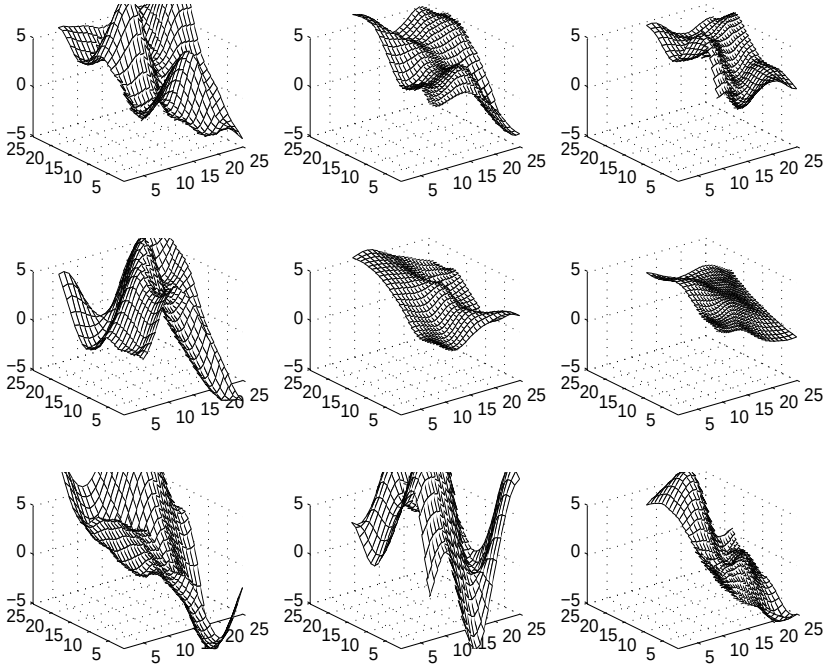

Figure 6: *Realizations of T-distributed RFs with $\nu = 1$.*

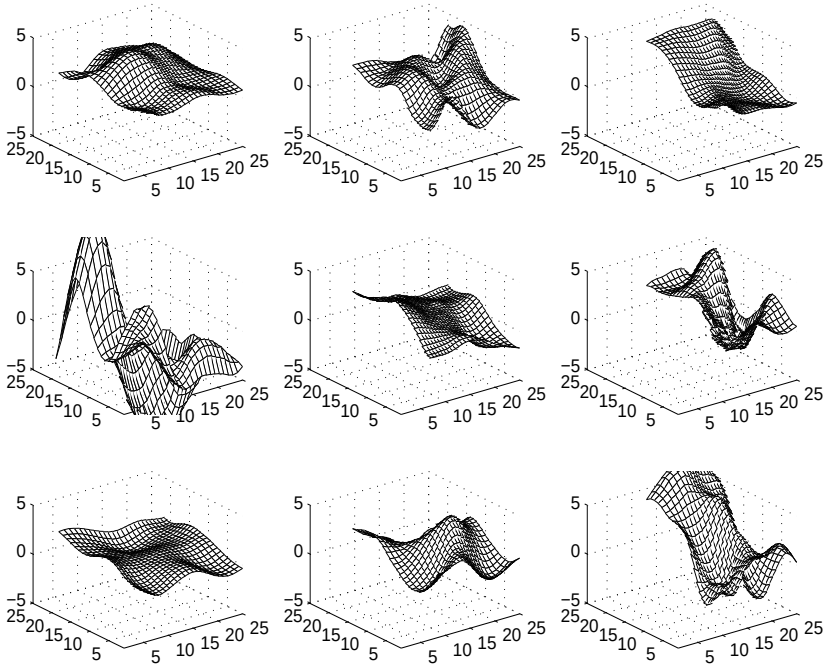

Figure 7: *Realizations of T-distributed RFs with $\nu = 3$.*

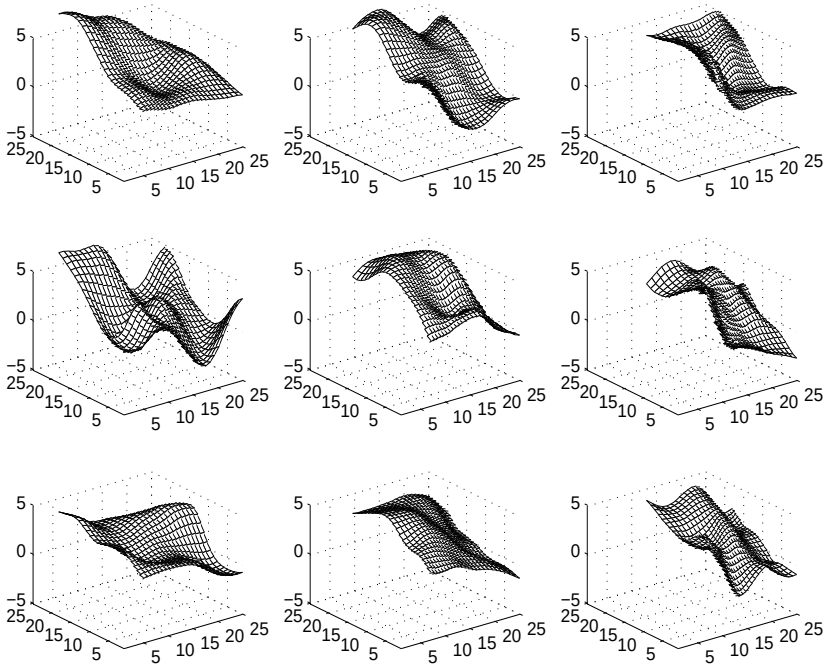

Figure 8: *Realizations of T-distributed RFs with $\nu = 5$.*

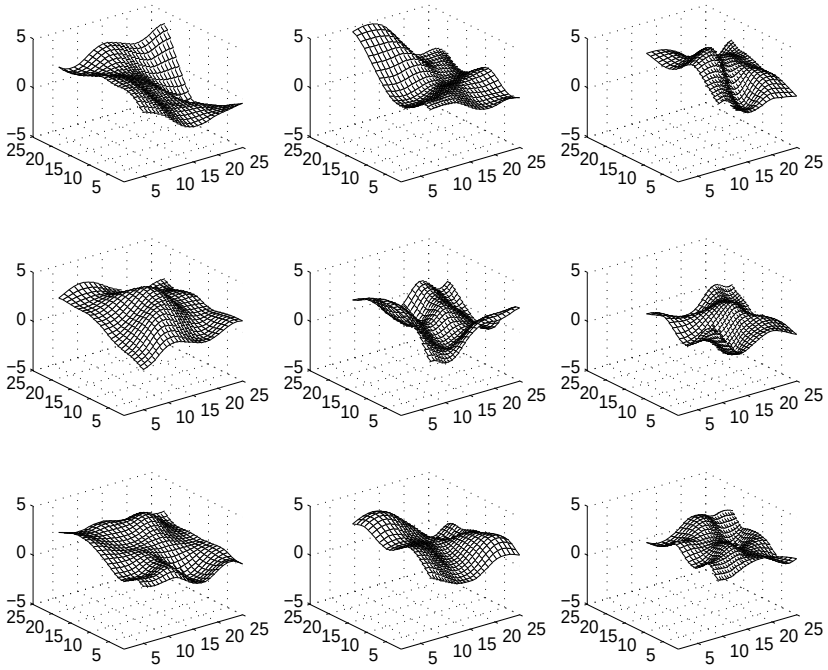

Figure 9: *Realizations of T-distributed RFs with $\nu = \infty$, i.e. Gaussian RFs.*

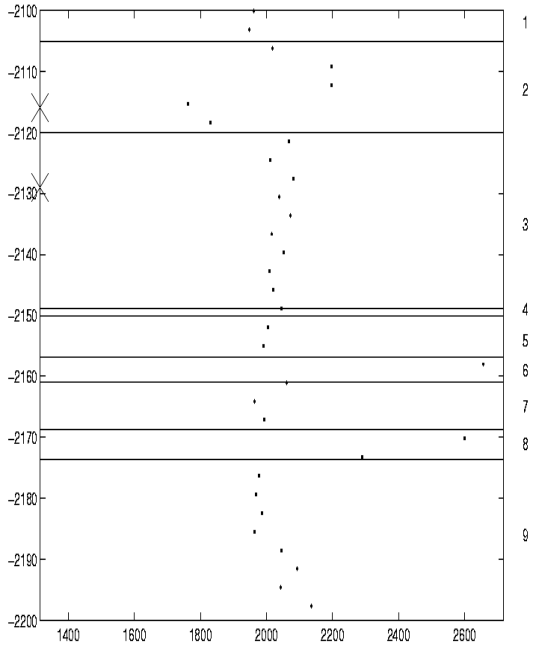

Figure 10: *Bulk density log observations from well C33 from the Gullfaks field in the North sea with layer interpretation. The two locations to be predicted is marked (×).*

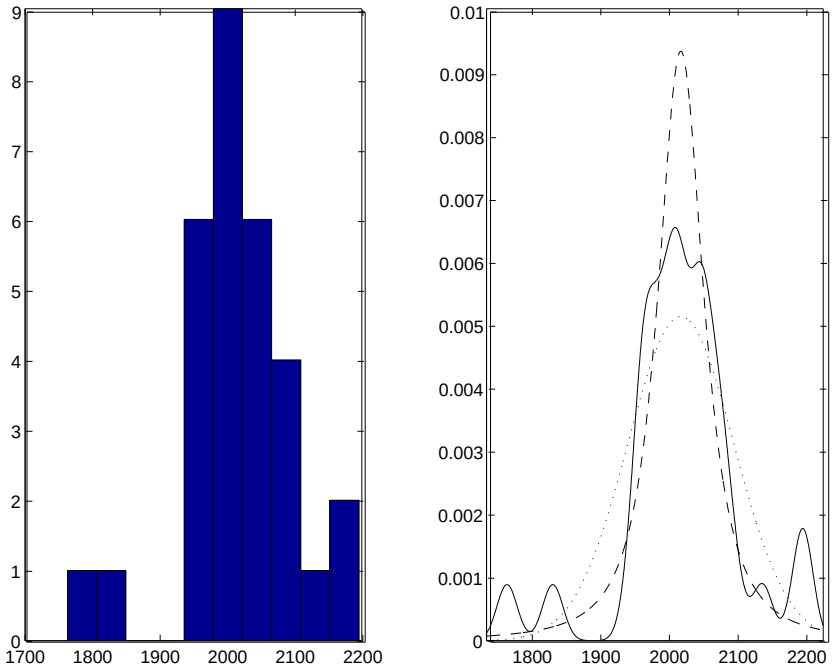

Figure 11: *Left: Histogram of density observations in pooled layers. Right: Kernel estimate of pdf (—); Gaussian estimate of pdf (· · ·); and Student-T estimate of pdf (− − −).*

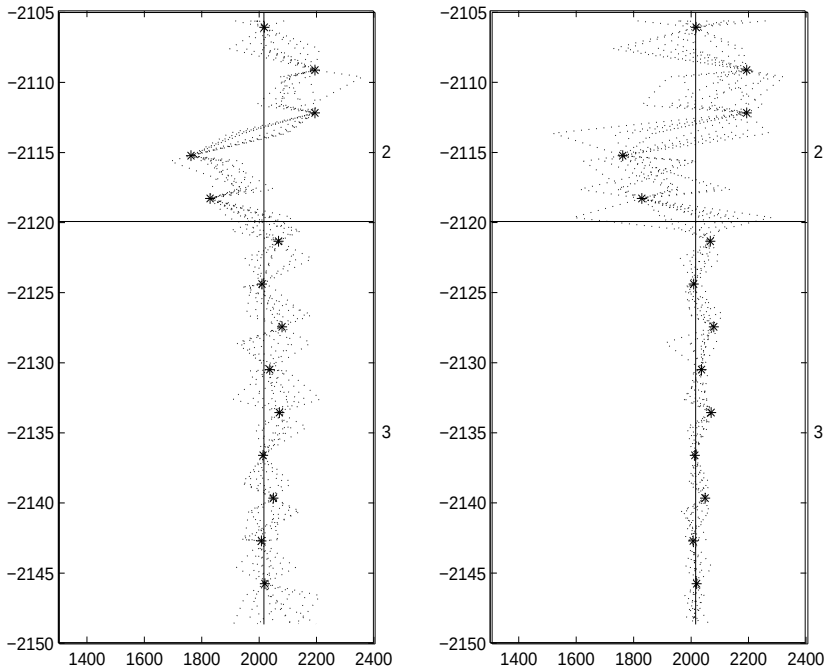

Figure 12: *Left: Ten conditional realizations from the inferred Gaussian RF model. Right: Ten conditional realizations from theand inferred TRF model.*

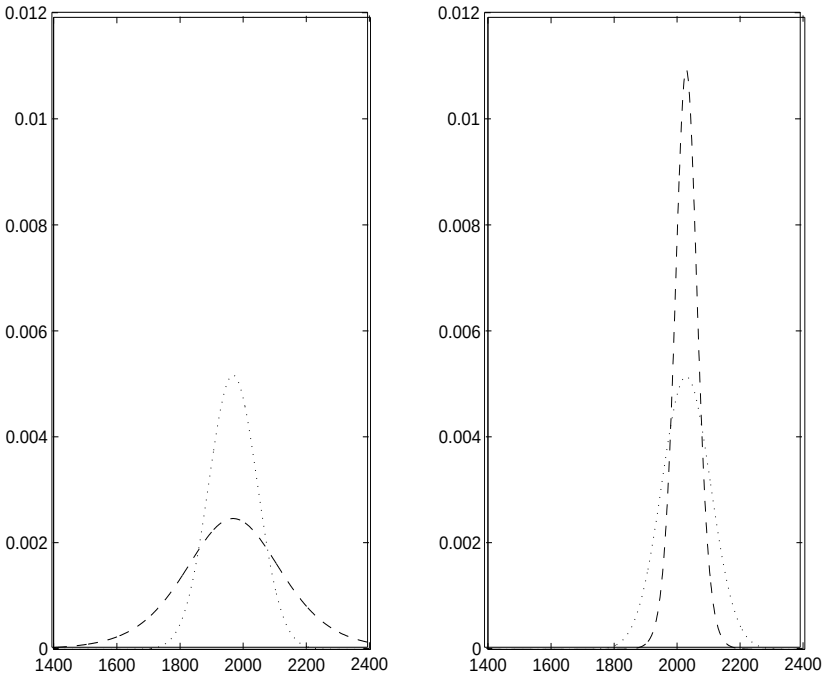

Figure 13: *Prediction pdf based on the inferred Gaussian RF model (· · ·) and the inferred T-distributed RF model (− − −) in layer 2 (left) and layer 3 (right).*

Continuous Near-Well Bayesian Linearized AVO Inversion

Jo Røislien
Department of Petroleum Engineering and Applied Geophysics
Norwegian University of Science and Technology

Henning Omre
Department of Mathematical Sciences
Norwegian University of Science and Technology

Abstract

Most existing geophysical inversion methods within the Bayesian framework are based on gridding of the prior distribution, and then performing the inversion on this grid. This paper introduces a grid free approach to Bayesian geophysical inversion, making it easy to combine both seismic observations and well log data in the same model, while accounting for their exactly correct locations. By assuming that the joint prior distribution for the elastic parameters P-wave velocity, S-wave velocity and density is a log-Gaussian random field, and performing a linearization of the well log data, the posterior distribution for the elastic parameters given both seismic observations and well log data is analytically available, and approximately a Gaussian random field. As the solution is given on a continuous form, the grid design does not have to be decided prior to the inversion. The continuous representation is useful in estimation of depth conversion and wavelet parameters, in fluid flow simulation with grid refinements, and in real-time drilling management. The inversion is demonstrated on real data from the Sleipner field in the North Sea.

URN:NBN:no-3476

1 Introduction

The objective of geophysical inversion is to describe the properties of the underground, given some geophysical measurements like seismics and well logs. Geophysical inversion is inherently ill-posed, meaning that no one, stable solution is obtainable. In order to obtain a stable solution on the material properties, one needs to provide, directly or indirectly, information on the preferred solution. As stated by Backus (1988), any geophysical inverse problem has both an existence half, and a uniqueness half. From a statisticial point of view, this is essential, as one does not only strive for the most probable solution, but also seeks to be able to quantify the uncertainty of this estimate.

The Bayesian framework is well suited for this task. Here the constraints on the solution is introduced through a prior distribution on the geophysical parameters of interest. This prior is often assumed to be a stationary Gaussian random field, due to the simplicity of the solutions when solving linear inverse problems, see Tarantola (1987). The prior distribution is then coupled with the likelihood function; the link between the physical measurements and the geophysical parameters. The complete solution in a probabilistic framework is provided by the posterior distribution, which formally is proportional to the product of the prior distribution and the likelihood.

One of the measurements at hand is seismic data, for example amplitude versus offset (AVO) data. If these data are pre-processed, Bayesian linearized AVO inversion can be performed, see Buland and Omre (2003a) for a discussion on the matter. Here a Gaussian random field is assumed for the prior, resulting in an analytical solution for the posterior. In Buland et al. (2003d) this method is expanded to a full 3D model, with full spatial correlation, again with an analytical solution. The inversion is very fast, using the fast Fourier transform. In this model, also the wells are taken into consideration, and inversion results are conditioned on the well information. This is achieved by forcing the well locations to coinside with the seismic grid. However, due to the large area of study, the effects of doing so are negligible.

In the near vicinity of the well, the geometry is of utter importance. We seek a grid free representation to be able to incorporate all information in its correct location close to a well, and invert to find P-wave velocity, S-wave velocity and density. In the near-well area, the uncertainty in the well measurements are of importance, too, so these need to be stochastically modelled. By assuming that the elastic variables are a priori log-Gaussian random field, a continuous analytical solution is obtained for the posterior distribution, conditioned on both seismic data and well log measurements. Thus, one can freely discretize the posterior random field itself. The problem of discretizing a continuous random field using any set of basis functions is addressed in Røislien et al. (2004). The inversion is demonstrated on real data from the Sleipner field.

This article is organized as follows; in Section 2 the link between seismic data, well log data

and the elastic parameters is described mathematically; in Section 3 a stochastic model for the problem is introduced, and a Bayesian inversion is performed; in Section 4 the inversion approach is demonstrated on real data from the Sleipner field in the North Sea; finally, Section 5 contains a summary.

2 Linearized Models for the Observations

Consider a volume surrounding a deviated well, and assume that within this volume there are both surface seismic observations on some grid and well observations along the well-path. See Figure 1 for a schematic overview. Here d_s denotes the seismic observations and d_w the well observations. Seismic AVO data is well suited for for extracting information about the elastic subsurface parameters, as the inversion problem may be linearized if the data has been appropriately prosessed, see Buland and Omre (2003a). The well data is information gathered from well logs. Examples of logging techniques are accoustic logging for measuring P- and S-wave velocities, and gamma ray for measuring the bulk density. Though the velocity data is linked to the measuring of travel times of the P- and S-waves, these raw time data are rarely available. The same holds for the density data. The well data can thus be viewed as direct measurements of the elastic parameters.

An isotropic, elastic medium is fully described by the vector $[\alpha, \beta, \rho]$ of elastic variables, with α representing the P-wave velocity, β the S-wave velocity and ρ the density. This vector is continuous in both the lateral and temporal direction, and can be written as a continuous spatial random vector field; $\{\alpha(\mathbf{x},t), \beta(\mathbf{x},t), \rho(\mathbf{x},t); \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^2, t \in \mathcal{T} \subset \mathbb{R}_+\}$. For more on random fields, see Adler (1981), Christakos (1992) and Abrahamsen (1997). We apply the notation $\mathcal{R} = \mathcal{D} \times \mathcal{T}$ for the space spanned by $\mathcal{D}$ and $\mathcal{T}$ and the short-hand notation $m(\mathbf{x},t) = [\alpha(\mathbf{x},t), \beta(\mathbf{x},t), \rho(\mathbf{x},t)]$ for the elastic parameter vector. The interest lies in describing the random field $\{m(\mathbf{x},t); (\mathbf{x},t) \in \mathcal{R}\}$ statistically within $\mathcal{R}$, conditioned on both seismic data and well observerations. The seismic data and the elastic variables are related through the Zoeppritz equation according to the right hand side of Figure 2, whereas the well observations and the elastic variables are related through a model at left hand side of Figure 2. In the following subsections, these two relations will be defined mathematically.

2.1 Linearized Model for Seismic Observations

In general, a seismic observation in location $(\mathbf{x},t)$ for some angle θ is a convolution between a wavelet $\psi(\mathbf{x},t)$ and some measurable function $h_s^\theta\big(m(\mathbf{x},t)\big)$ of the elastic variables;

$$d_s^\theta(\mathbf{x},t) = \int_{\mathcal{L}^s} \psi(\mathbf{x}, t-v) h_s^\theta(m(\mathbf{x},v))\mathrm{d}v + \epsilon_s^\theta(\mathbf{x},t), \tag{1}$$

with $\epsilon_s^\theta(\mathbf{x},t)$ an error term. Allowing the wavelet to be angle dependent is straightforward, see Buland and Omre (2003a,b,c). Seismic traces are assumed to be vertical, implying that the integration $\mathcal{L}^s$ is a line integral over $\mathcal{T}$ only.

A weak contrast approximation to the PP reflection coefficient extended to a time continuous reflectivity function reads

$$\begin{aligned} c_{PP}(\mathbf{x},t,\theta) &= a_\alpha(\mathbf{x},t,\theta)\frac{\partial}{\partial t}\ln\alpha(\mathbf{x},t) \\ &\quad + a_\beta(\mathbf{x},t,\theta)\frac{\partial}{\partial t}\ln\beta(\mathbf{x},t) + a_\rho(\mathbf{x},t,\theta)\frac{\partial}{\partial t}\ln\rho(t) \end{aligned} \tag{2}$$

with

$$a_\alpha(\mathbf{x},t,\theta) = \frac{1}{2}\left(1+\tan^2\theta\right) \tag{3}$$

$$a_\beta(\mathbf{x},t,\theta) = -4\frac{\beta^2(\mathbf{x},t)}{\alpha^2(\mathbf{x},t)}\sin^2\theta \tag{4}$$

$$a_\rho(\mathbf{x},t,\theta) = \frac{1}{2}\left(1-4\frac{\beta^2(\mathbf{x},t)}{\alpha^2(\mathbf{x},t)}\sin^2\theta\right), \tag{5}$$

see Buland et al. (2003d). For simplification, we introduce the vector notation $a(\mathbf{x},t,\theta) = [a_\alpha(\mathbf{x},t,\theta), a_\beta(\mathbf{x},t,\theta), a_\rho(\mathbf{x},t,\theta]^T$ for these functions. Assuming a background model with a constant β/α-ratio, so that $a(\mathbf{x},t,\theta) = a(\theta)$, expression (2) can be rewritten in compact form as

$$h_s^\theta(m(\mathbf{x},t)) = c_{PP}(\mathbf{x},t,\theta) = a(\theta)^T\frac{\partial}{\partial t}\ln m(\mathbf{x},t)\ . \tag{6}$$

Using the notation $m_L(\mathbf{x},t) = \ln m(\mathbf{x},t) = [\ln\alpha(\mathbf{x},t), \ln\beta(\mathbf{x},t), \ln\rho(\mathbf{x},t)]$, expression (1) may now be rewritten as

$$d_s^\theta(\mathbf{x},t) = \int_{\mathcal{L}^s}\psi(\mathbf{x},t-v)a(\theta)^T\frac{\partial m_L(\mathbf{x},v)}{\partial v}\mathrm{d}v + \epsilon_s^\theta(\mathbf{x},t), \tag{7}$$

which is linear with respect to the the time derivative of $m_L(\mathbf{x},t)$.

Seismic data are available as a discrete set in given locations. Assume that the observed seismic gathers are available in locations $(\mathbf{x},t)_i; i=1,\ldots,n_p$ and for angles $\theta_j; j=1,\ldots,n_\theta$ in each gather. This results in the seismic observation vector

$$d_s = \begin{bmatrix} d_s^1 \\ \vdots \\ d_s^{n_s} \end{bmatrix} \tag{8}$$

with $n_s = n_p \times n_\theta$, where each of the seismic observations $d_s^i; i=1,\ldots,n_s$ are given from expression (7).

2.2 Linearized Model for Well Observations

In general, a well observation in location $(\mathbf{x}, t)$ is a convolution between a wavelet $\xi(\mathbf{x}, t)$ and some measurable function $h_w(m(\mathbf{x}, t))$ of the elastic variables;

$$d_w(\mathbf{x}, t) = \int_{\mathcal{L}^w} \xi(\mathbf{x} - \mathbf{u}, t - v) h_w(m(\mathbf{u}, v)) \mathrm{d}\mathbf{u}\mathrm{d}v + \epsilon_w(\mathbf{x}, t), \tag{9}$$

with $\epsilon_w(\mathbf{x}, t)$ an error term. We allow non-vertical well-paths, so that the integral $\mathcal{L}^w$ is a line integral along the well trace in $\mathcal{R}$.

A general linear model is assumed for the well observations,

$$h_w(m(\mathbf{x}, t)) = Bm(\mathbf{x}, t) \tag{10}$$

with B a 3×3 matrix. Due to the data aquisition process, B will often the identity matrix. However, we will use a general B throughout this paper for demonstrational purposes. Expression (9) may now be rewritten as

$$d_w(\mathbf{x}, t) \;\; = \;\; \int_{\mathcal{L}^w} \xi(\mathbf{x} - \mathbf{u}, t - v) B m(\mathbf{u}, v) \mathrm{d}\mathbf{u}\mathrm{d}v + \epsilon_w(\mathbf{x}, t) \tag{11}$$

which is linear with respect to $m(\mathbf{x}, t)$.

Well data are available as a discrete set in given locations. Assume that the well observations are available in locations $(\mathbf{x}, t)_i; i = 1, \ldots, n_w$. This results in the well observation vector

$$d_w = \begin{bmatrix} d_w^1 \\ \vdots \\ d_w^{n_w} \end{bmatrix} \tag{12}$$

where each of the well observations $d_w^i; i = 1, \ldots, n_w$ are given from expression (11).

3 Stochastic Model

We cast the near-well inversion problem in a Bayesian framework. The mathematical relations between the elastic parameters and the data were outlined in Section 2. Here, additional stochastic properties are discussed, and the posterior distribution of the randon field $\{\alpha(\mathbf{x}, t), \beta(\mathbf{x}, t), \rho(\mathbf{x}, t); (\mathbf{x}, t) \in \mathcal{R}\}$ conditioned on both the seismic and well data is developed.

3.1 Likelihood model

We model the seismic errors as zero mean Gaussian residuals for each of the seismic observations;

$$\epsilon_s = [\epsilon_s^1, \ldots \epsilon_s^{n_s}]^T \sim N_{n_s}(0, \Gamma_{\epsilon_s}) \tag{13}$$

with Γ_{ϵ_s} being a $(n_s \times n_s)$-dimensional covariance matrix defined by

$$\gamma_{\epsilon_s}^{\mathbf{x}t\theta} = \mathrm{Cov}\left\{\epsilon_s^{\theta_1}(\mathbf{x}_1, t_1), \epsilon_s^{\theta_2}(\mathbf{x}_2, t_2)\right\} \tag{14}$$

Thus, from Section 2.1 we have that

$$\left[d_s | m_L(\mathbf{x}, t)\right] \sim N_{n_s}\left(\mu_{d_s|m_L}, \Gamma_{d_s|m_L}\right) \tag{15}$$

with $\mu_{d_s|m_L}$ a $(n_s \times 1)$-dimensional expectation vector defined by

$$\begin{aligned} \mu_{d_s|m_L}^{\mathbf{x}t\theta} &= \mathrm{E}\left\{d_s^{\theta}(\mathbf{x}, t) | m_L(\mathbf{x}, t)\right\} & (16)\\ &= \int_{\mathcal{L}^s} \psi(\mathbf{x}, t-v) a(\theta)^T \frac{\partial m_L(\mathbf{x}, v)}{\partial v} \mathrm{d}v & (17) \end{aligned}$$

and $\Gamma_{d_s|m_L}$ being a $(n_s \times n_s)$-dimensional covariance matrix identical to Γ_{ϵ_s}.

We model the well errors as zero mean Gaussian residuals for each of the well observations;

$$\epsilon_w = [\epsilon_w^1, \ldots \epsilon_w^{n_w}]^T \sim N_{n_w}(0, \Gamma_{\epsilon_w}) \tag{18}$$

with Γ_{ϵ_w} being a $(n_w \times n_w)$-dimensional covariance matrix defined by

$$\gamma_{\epsilon_w}^{\mathbf{x}t} = \mathrm{Cov}\left\{\epsilon_w(\mathbf{x}_1, t_1), \epsilon_w(\mathbf{x}_2, t_2)\right\} \tag{19}$$

Thus, from Section 2.2 we have that

$$\left[d_w | m_L(\mathbf{x}, t)\right] \sim N_{n_w}\left(\mu_{d_w|m_L}, \Gamma_{d_w|m_L}\right) \tag{20}$$

with $\mu_{d_w|m_L}$ a $(n_w \times 1)$-dimensional expectation vector defined by

$$\begin{aligned} \mu_{d_w|m_L}^{\mathbf{x}t} &= \mathrm{E}\left\{d_w(\mathbf{x}, t) | m_L(\mathbf{x}, t)\right\} & (21)\\ &= \int_{\mathcal{L}^w} \xi(\mathbf{x}-\mathbf{u}, t-v) B \exp\{m_L(\mathbf{u}, v)\} \mathrm{d}\mathbf{u}\mathrm{d}v & (22) \end{aligned}$$

and $\Gamma_{d_w|m_L}$ being a $(n_w \times n_w)$-dimensional covariance matrix identical to Γ_{ϵ_w}. Note that the well likelihood is slightly changed from expression (11), since it is now made conditional on $m_L(\mathbf{x}, t)$ rather than $m(\mathbf{x}, t)$. An appropriate transformation is made in expression (22)

For the complete observation vector $d = (d_s^T, d_w^T)^T$, we now have

$$\left[d | m_L(\mathbf{x}, t)\right] = \left[(d_s^T, d_w^T)^T | m_L(\mathbf{x}, t)\right] \sim N_{n_d}(\mu_{d|m_L}, \Gamma_{d|m_L}) \tag{23}$$

with $(n_d \times 1)$-dimensional expectation vector

$$\mu_{d|m_L} = \begin{bmatrix} \mu_{d_s|m_L} \\ \mu_{d_w|m_L} \end{bmatrix} \tag{24}$$

and $(n_d \times n_d)$-dimensional covariance matrix

$$\Gamma_{d|m_L} = \begin{bmatrix} \Gamma_{d_s|m_L} & 0 \\ 0 & \Gamma_{d_w|m_L} \end{bmatrix}, \tag{25}$$

and $n_d = n_s + n_w$. Assuming that the seismic errors ϵ_s and well errors ϵ_w are independent, results in the off-diagonal blocks to be all zeros.

3.2 Prior model

We assume that $\{m(\mathbf{x}, t); (\mathbf{x}, t) \in \mathcal{R}\}$ is a continuous log-Gaussian random field. This implies that $\{m_L(\mathbf{x}, t); (\mathbf{x}, t) \in \mathcal{R}\}$ is a continuous Gaussian random field in $\mathbf{x}$ and t,

$$m_L(\mathbf{x}, t) \sim G_{\mathbf{x}t}\left(\mu_{m_L}^{\mathbf{x}t}, \gamma_{m_L}^{\mathbf{x}t}\right) \tag{26}$$

with expectation function

$$\begin{aligned} \mu_{m_L}^{\mathbf{x}t} &= \mathrm{E}\{m_L(\mathbf{x}, t)\} && (27)\\ &= \left[\mathrm{E}\{\ln \alpha(\mathbf{x}, t)\}, \mathrm{E}\{\ln \beta(\mathbf{x}, t)\}, \mathrm{E}\{\ln \rho(\mathbf{x}, t)\}\right]^T && (28)\\ &= [\mu_\alpha(\mathbf{x}, t), \mu_\beta(\mathbf{x}, t), \mu_\rho(\mathbf{x}, t)]^T && (29) \end{aligned}$$

for all $\mathbf{x} \in \mathcal{D}$ and $t \in \mathcal{T}$, and covariance function

$$\begin{aligned} \gamma_{m_L}^{\mathbf{x}t} &= \mathrm{Cov}\{m_L(\mathbf{x}_1, t_1), m_L(\mathbf{x}_2, t_2)\} && (30)\\ &= \Gamma_0 \varrho(\xi, \tau) && (31) \end{aligned}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and $t_1, t_2 \in \mathcal{T}$. The last equality is the result of assuming stationarity. Here Γ_0 is a (3×3)-dimensional location invariant covariance matrix between $\ln \alpha(\mathbf{x}, t)$, $\ln \beta(\mathbf{x}, t)$ and $\ln \rho(\mathbf{x}, t)$, and $\varrho(\xi, \tau)$ is a spatial correlation function where $\xi = |\mathbf{x}_2 - \mathbf{x}_1|$ denotes lateral distance and $\tau = |t_2 - t_1|$ temporal distance. That is, the covariance function may be divided into a part consisting of the covariances between $(\ln \alpha, \ln \beta, \ln \rho)$, and a spatial correlation function. This correlation function must be positive definite. For more on such functions, see Abrahamsen (1997).

3.3 Linearized stochastic models

The seismic observation in expression (7) is a linear operator with respect to $\partial_t m_L(\mathbf{x}, t)$, with ∂_t denoting the t-derivative, plus random noise. As differentiation is a linear operator,

the differentiation with respect to t results in a new vector field in $\mathbf{x}$ and t which will also be Gaussian,

$$\partial_t m_L(\mathbf{x},t) \sim G_{\mathbf{xt}}\Big(\partial_t \mu_{m_L}(\mathbf{x},t), \partial_{t_1}\partial_{t_2}\gamma^{\mathbf{xt}}_{m_L}\Big) \tag{32}$$

see Christakos (1992), Abrahamsen (1997) and Buland and Omre (2003a). In the case of stationarity,

$$\begin{aligned}\partial_{t_1}\partial_{t_2}\gamma^{\mathbf{xt}}_{m_L} = \gamma^{\mathbf{xt}}_{m'_L} &= \mathrm{Cov}\left\{\partial_{t_1} m_L(\mathbf{x}_1,t_1), \partial_{t_2} m_L(\mathbf{x}_2,t_2)\right\} &(33)\\ &= -\Gamma_0 \partial^2_\tau \varrho(\xi,\tau) &(34)\end{aligned}$$

One will also need the cross-covariance between $m_L(\mathbf{x},t)$ and $\partial_t m_L(\mathbf{x},t)$, which is

$$\begin{aligned}\partial_{t_1}\gamma^{\mathbf{xt}}_{m_L} = \gamma^{\mathbf{xtxt}}_{m'_L m_L} &= \mathrm{Cov}\left\{\partial_{t_1} m_L(\mathbf{x}_1,t_1), m_L(\mathbf{x}_2,t_2)\right\} &(35)\\ &= \mathrm{sign}(t_1-t_2)\Gamma_0 \partial_\tau \varrho(\xi,\tau). &(36)\end{aligned}$$

where the last equality is the result of assuming stationarity.

The well observation (11) is a linear operator with respect to $m(\mathbf{x},t)$ plus random noise. Linearizing the log-Gaussian random field $m(\mathbf{x},t)$, it can be shown that

$$m(\mathbf{x},t) = \exp\{m_L(\mathbf{x},t)\} \quad \tilde{\rightarrow} \quad G_{\mathbf{xt}}\Big(\mu^{\mathbf{xt}}_m, \mu^{\mathbf{x}_1 t_1}_m \gamma^{\mathbf{xt}}_{m_L} \mu^{\mathbf{x}_2 t_2}_m\Big) \tag{37}$$

with $\tilde{\rightarrow}$ denoting approximation in distribution, see Appendix B.1 for proof of the result in the discrete case. Expression (37) holds whenever the location parameter becomes much larger than the scale parameter. The expectation function is

$$\mu^{\mathbf{xt}}_m = \exp\{\mu^{\mathbf{xt}}_{m_L}\} = \exp\Big\{\mathrm{E}\{m_L(\mathbf{x},t)\}\Big\}. \tag{38}$$

One will also need the crosscovariances between $m_L(\mathbf{x},t)$ and $\exp\{m_L(\mathbf{x},t)\}$, and $\partial_t m_L(\mathbf{x},t)$ and $\exp\{m_L(\mathbf{x},t)\}$, which are

$$\begin{aligned}\gamma^{\mathbf{xtxt}}_{m_L m} &= \mathrm{Cov}\Big\{m_L(\mathbf{x}_1,t_1), \exp\{m_L(\mathbf{x}_2,t_2)\}\Big\} &(39)\\ &= \gamma^{\mathbf{xt}}_{m_L} \exp\left\{\mu^{\mathbf{x}_2 t_2}_{m_L} + \gamma^{\mathbf{x}_2 t_2}_{m_L}\right\} &(40)\\ \gamma^{\mathbf{xtxt}}_{m'_L m} &= \mathrm{Cov}\Big\{\partial_{t_1} m_L(\mathbf{x}_1,t_1), \exp\{m_L(\mathbf{x}_2,t_2)\}\Big\} &(41)\\ &= \partial_{t_1}\gamma^{\mathbf{xt}}_{m_L} \exp\left\{\mu^{\mathbf{x}_2 t_2}_{m_L} + \gamma^{\mathbf{x}_2 t_2}_{m_L}\right\}. &(42)\end{aligned}$$

see Appendix B.2 for a proof in the discrete case.

3.4 Posterior model

With the likelihood being Gaussian and linear in the conditioning variable, and the prior distribution being Gaussian, the posterior distribution will also be Gaussian. Thus, with the linearization in expression (37) for the well information, the distribution of the elastic parameter vector conditioned on both seismic and well data, d, is approximately a continuous Gaussian random field in $\mathbf{x}$ and t;

$$\big[m_L(\mathbf{x},t)|d\big] \sim G_{\mathbf{x}t}\Big(\mu^{\mathbf{x}t}_{m_L|d}\ ,\ \gamma^{\mathbf{x}t}_{m_L|d}\Big) \tag{43}$$

with expectation function

$$\mu^{\mathbf{x}t}_{m_L|d} \ = \ \mu^{\mathbf{x}t}_{m_L} + \gamma_{m^{\mathbf{x}t}_L d}\Gamma_d^{-1}(d-\mu_d) \tag{44}$$

for $\mathbf{x} \in \mathcal{D}$ and $t \in \mathcal{T}$, and covariance function

$$\gamma^{\mathbf{x}t}_{m_L|d} \ = \ \gamma^{\mathbf{x}t}_{m_L} - \gamma_{m^{\mathbf{x}_1 t_1}_L d}\Gamma_d^{-1}\gamma_{d m^{\mathbf{x}_2 t_2}_L} \tag{45}$$

for $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and $t_1, t_2 \in \mathcal{T}$, with

$$\gamma_{m^{\mathbf{x}t}_L d} \ = \ \Big[\mathrm{Cov}\{m_L(\mathbf{x},t), d^1\}, \ldots, \mathrm{Cov}\{m_L(\mathbf{x},t), d^n\}\Big]^T \tag{46}$$

having elements defined by

$$\mathrm{Cov}\left\{m_L(\mathbf{x}_1,t_1), d^\theta_s(\mathbf{x}_2,t_2)\right\} \ = \ \int_{\mathcal{L}^s} \psi(\mathbf{x}_2, t_2 - v)\gamma^{\mathbf{x}_1 t_1 \mathbf{x}_2 v}_{m_L m'_L} a(\theta)\mathrm{d}v \tag{47}$$

$$\mathrm{Cov}\left\{m(\mathbf{x}_1,t_1), d_w(\mathbf{x}_2,t_2)\right\} \ = \ \int_{\mathcal{L}^w} \xi(\mathbf{x}_2 - \mathbf{u}, t_2 - v)\gamma^{\mathbf{x}_1 t_1 \mathbf{u} v}_{m_L m} B^T \mathrm{d}\mathbf{u}\mathrm{d}v \tag{48}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, $t_1, t_2 \in \mathcal{T}$ and $\theta \in [0, 2\pi)$. Further

$$\Gamma_d = \begin{bmatrix} \Gamma_{d_s} & \Gamma_{d_s d_w} \\ \Gamma_{d_w d_s} & \Gamma_{d_w} \end{bmatrix}, \tag{49}$$

and

$$\mu_d = \begin{bmatrix} \mu_{d_s} \\ \mu_{d_w} \end{bmatrix} \tag{50}$$

see Appendix A for a complete definition of expressions (49) and (50). Expressions (47) and (48) are the crosscovariance between the elastic variables and a seismic observation, and the crosscovariance between the elastic variables and a well observation, respectively.

3.5 Discretization of the posterior distribution

Having the solution (43) on a continous form allows for discretizing of the solution, rather than having to discretize before the actual inversion, having to carry that choice through. In Røislien et al. (2004) it is shown how to discretize a continous random field $\{Z(y); y \in \mathcal{R}\}$ with respect to any set of basis functions, using the expansion

$$Z(y) = \lim_{n\to\infty} \sum_{i=1}^{n} A_i f_i(y) \tag{51}$$

with $\{f_i(y)\}_{i=1}^{n}$ a set of known basis functions, an $A = [A_1, \ldots, A_n]^T$ a vector of unknown stochastic parameters. If $n < \infty$ expression (51) is an approximation. As an example; using orthogonal box basis functions, each of the n stochastic parameters can be described as

$$A_i = \frac{1}{|V_i|} \int_{V_i} Z(u)\mathrm{d}u \quad ; \; i = 1, \ldots, n \tag{52}$$

with $V_i \subset \mathcal{R}$, and $|V_i|$ denoting the volume of integration defined by the box basis functions. The expectation and covariance for the random variables is

$$\mathrm{E}(A_i) = \frac{1}{|V_i|} \int_{V_i} \mu(u)\mathrm{d}u \quad ; \; i = 1, \ldots, n \tag{53}$$

$$\mathrm{Cov}(A_i, A_j) = \frac{1}{|V_i||V_j|} \int_{V_i} \int_{V_j} C(u, v)\mathrm{d}u\mathrm{d}v \quad ; \; i = 1, \ldots, n \tag{54}$$

with $\mu(y)$ and $C(y_1, y_2)$ the expectation and covariance function of $Z(y)$, respectively. In the special case of $\{Z(y); y \in \mathcal{D}\}$ being a Gaussian random field, A will be a multivariate Gaussian random vector.

3.6 Computational aspects

The posterior distribution in expression (43) is an update of the prior distribution (26) with respect to the data d. In order to calculate the posterior expectation function $\mu_{m_L}^{\mathbf{xt}}|d$ and posterior covariance function $\gamma_{m_L}^{\mathbf{xt}}|d$ for $[m_L(\mathbf{x}, t)|d]$ in some location, given in expressions (44) and (45) respectively, we need to update the prior expectation function $\mu_{m_L}^{\mathbf{xt}}$ and prior covariance function $\gamma_{m_L}^{\mathbf{xt}}$. This is done by calculating $\gamma_{m_L^{\mathbf{xt}}d}$, Γ_d and μ_d, defined by line integrals given in Section 3 and Appendix A. Each of these integrals generally need to be solved numerically and separately. For each new location in the posterior discretization, preprogrammed subroutines calculating these integrals are then called. Thus, adding new locations in the posterior grid is straightforward once the numerical subroutines have been implemented.

If the number of posterior grid nodes becomes large, it can be computationally faster to discretize the problem aforehand, turning the problem into a matrix inversion problem, see Appendix C for details. However, after having performed this matrix inversion for a given grid, grid refinements can then be consistengly added where desired, using the continuous solution.

In Buland et al. (2003d) Bayesian AVO inversion is performed using the FFT, resulitng in a very fast inversion, as well as low CPU demand. The FFT approach can, however, not be applied here, due to the optional gridding design and the discretizing of the posterior rather than the prior. The cost of the added flexibility in this continuous approach is increased computing time and complexity.

4 Inversion of Sleipner Data

The Sleipner Øst Field is located in the South Viking Graben in the Norwegian Block 15/9. For more detail about the dat set and the seismic processing, see Buland and Omre (2003a).

In the inversion, a rectangular piece of the seismic survey limited by inlines 1625-1633 and crosslines 1290-1295 is used. Within this small area most seismic traces are fairly similar, see Figure 3 for a complete angle gather for the 9 available angles $[5^o, 9^o, 13^o, \ldots, 37^o]$. We have, however, restricted the inversion to the four angles $[5^o, 9^o, 13^o, 17^o]$ in order to speed up the process. As for the temporal direction, we have chosen depth 2280ms to 2350ms. Within this volume, a non-vertical well is embedded. The complete well logs for P-wave velocity, S-wave velocity and density are shown in Figure 4. As the well log data are given as a function of depth in metres, the well log data are converted to depth in time.

4.1 Model parameter estimation

The most important piece of information for estimating parameters is the well logs. For the assumed constant background model we find $\hat{\mu}_\alpha = 7.951$, $\hat{\mu}_\beta = 7.319$ and $\hat{\mu}_\rho = 7.701$ using standard estimators, corresponding to P-wave velocity $2838m/s$, S-wave velocity $1508m/s$ and density $2211kg/m^3$. The elements of the stationary covariance matrix

$$\Gamma_0 = \begin{bmatrix} \sigma^2_{\ln\alpha} & \sigma_{\ln\alpha}\sigma_{\ln\beta}c_{\alpha\beta} & \sigma_{\ln\alpha}\sigma_{\ln\rho}c_{\alpha\rho} \\ \sigma_{\ln\beta}\sigma_{\ln\alpha}c_{\beta\alpha} & \sigma^2_{\ln\beta} & \sigma_{\ln\beta}\sigma_{\ln\rho}c_{\beta\rho} \\ \sigma_{\ln\rho}\sigma_{\ln\alpha}c_{\rho\alpha} & \sigma_{\ln\rho}\sigma_{\ln\beta}c_{\rho\beta} & \sigma^2_{\ln\rho} \end{bmatrix} \tag{55}$$

from expression (31) are also estimated using standard estimators. We find $\hat{\sigma}^2_{\ln\alpha} = 0.0117$, $\hat{\sigma}^2_{\ln\beta} = 0.0350$ and $\hat{\sigma}^2_{\ln\rho} = 0.00163$ for the variances of m_L, and $\hat{c}_{\alpha\beta} = 0.8857$, $\hat{c}_{\alpha\rho} = -0.1651$,

$\hat{c}_{\beta\rho} = -0.4266$ for the correlations between $(\ln\alpha, \ln\beta, \ln\rho)$. As the expectations are of magnitude 1 and the the variances of magnitude 10^{-2} the use of the approximation (37) in Section 3.3 is justified.

The spatial coupling is introduced through a temporal and a lateral correlation function, with the total correlation function is the product of these. The empirical correlation function in the temporal direction is estimated for several time lags in the well, and fitted by eye by an ordinary Bessel correlation function,

$$\varrho_J(\tau; \nu_1, \nu_2) = \Gamma(\nu_2 + 1) 2^{\nu_2} (\nu_1 \tau)^{-\nu_2} J_{\nu_2}(\nu_1 \tau) \tag{56}$$

with $\Gamma(x)$ here denoting the gamma function, see Figure 5. The parameters used are $\nu_1 = 0.33$ and $\nu_2 = 0.5$. The fit is satisfactory. The correlation in the lateral direction can be found from the seismics, and is fitted by a first order exponential function,

$$\rho_\xi(\xi; R_\xi) = \exp\left\{ \frac{\sqrt{\xi_1^2 + \xi_2^2}}{R_\xi} \right\}, \tag{57}$$

with range $R_\xi = 250$m, see Buland and Omre (2003a). In the simulation, however, we have used $R_\xi = 25$ in order to demonstrate the behaviour around the well with less CPU demand.

In Buland and Omre (2003a) the seismic wavelets for this set of data was estimated from the well log data separately for each angle. However, as all wavelets are fairly similar for the four chosen angles, we apply the same Ricker wavelet for all angles, using peak frequency 25Hz. The zero-phase Ricker wavelet is given as

$$\psi^\theta(t; \phi) = \psi(t; \phi) = c(1 - 2\pi^2\phi^2 t^2) \exp(-\pi^2 \phi^2 t^2) \tag{58}$$

with ϕ being the peak frequency and t the travel times, and c a normalizing constant. Introducing different wavelets for different angles is straightforward. Estimation of wavelets from the data is discussed in Buland and Omre (2003b,c). The well measurements are subject to an averaging having the size of the tool. Thus, for the well data, a uniform averaging function with width $7 \times 15.24cm$, corresponding to the whole length of the tool, has been used for both the velocity and density data. In Figure 6 the seismic wavelet and the averaging function for the well data are compared. The difference in scale is dramatical, but posesses no real problem in the inversion, as the upscaling is implicitely taken care of.

The seismic noise is assumed to be a mixture of white and coloured noise,

$$\epsilon_s^\theta(\mathbf{x}, t) = e_1 + S e_2(\theta) \tag{59}$$

The first term represents white noise, $e_1 \sim N(0, \sigma_1^2)$. Letting $e_1 = e_1(\mathbf{x}, t)$ is straightforward. The other term is an angle dependent error. For each angle θ_i, the error term $e_2(\theta_i)$ is white noise, but correlated between different angles by

$$\rho_\theta = \exp\left\{ -\frac{|\theta_i - \theta_j|}{R_\theta} \right\} \tag{60}$$

Using standard estimators, the parameters were estimated to be $\hat{\sigma}_1^2 = 0.0114$, $\hat{\sigma}_2^2 = 0.0001$ and $\hat{R}_\theta = 10^o$, see Buland and Omre (2003a). For other ways of modelling the error, see for example Buland et al. (2003d). All well errors are assumed to be independent and Gaussian,

$$\epsilon_w(\mathbf{x}, t) = \epsilon_w \sim N(0, \sigma_w^2) \tag{61}$$

with $\hat{\sigma}_w^2 = 1000$, corresponding to an accuracy of 95% for the well measurements. For more on density log errors, see for example Frykman and Deutsch (2002).

4.2 Simulation results

We invert the 3D-volume defined by inlines 1625-1633, crosslines 1290-1295 and depth $2280ms$ to $2355ms$. This corresponds to 5 inlines and 6 crosslines, a total of 30 seismic gathers, covering an 100m×62.5m area. The well penetrates this volume, see Figure 7 for a schematic overview in the $\mathbf{x}$-plane.

We calculate the results of the inversion along a 2D cut between crosslines 1292 and 1293, for a regular grid with locations every 4m in the lateral direction and every 1.5ms in the temporal direction. The well runs through this 2D cut in a single point.

The conditional mean after the inversion along the 2D cut conditioned on the seismic observations only, is shown in the left display of Figure 8. The result is laterally very smooth, and varies little throughout the domain of study. It does, however, reveal the main vertical behaviour of the elastic parameters. As was mentioned in Buland and Omre (2003a) the AVO inversion reveals little about the bulk density, so this is by far the smoothest display. For a more detailed view, see Figure 9. Here an inversion along the well path, conditioned on the seismic observations only, is shown. Again, we observe how the seismic AVO inversion captures the main trends in the elastic parameters.

The conditional mean after the inversion along the 2D cut conditioned on the well log data only, is shown in the right display of Figure 8. The result reveals a lot of information about the elastic parameters close to the well. Further away from the well, however, the assumed constant background model is the only piece of information. For a more detailed view, see Figure 10. Here an inversion along the well path, conditioned on the well log data only, is shown. Observe that since a uniform averaging function was assumed for the well log data, the results after the inversion are even less smooth than the actual well logs.

The conditional mean and the conditional variance after the inversion conditioned on both seismic and well data, is shown in Figure 11; the left display shows the mean, whereas the rigth display shows the variance. The mean is detailed close to the well, whereas further away from it, it is the information from the seismics that dominates. The variance is smaller close to the well, and increases with increasing distance. Samples from this

posterior distribution are shown in Figure 12. As the variance is lower close to the well, the variation between samples is lower in this area. Further away from the well, the difference between the samples increases.

5 Conclusion

A continuous linear Bayesian AVO inversion method is developed, combining both seismic AVO data and well log data. The difference in scale for the two different types data is implicitely taken care of in the model. The result of the Bayesian inversion is an analytically available posterior for the elastic parameters P-wave velocity, S-wave velocity and density. This posterior is approximately a continuous Gaussian random field. The inversion is tested on real data from the Sleipner field. It is demonstrated how the inversion conditioned on both seismics and well data reveals detailed information about the elastic parameters close to the well, whereas further away from the well, the less informative seismic data is the main contributor in the inversion.

One may benefit from this continuous representation of the posterior model in depth conversion and wavelet estimation along the lines of Buland and Omre (2003b,c) since all information is correctly located. Moreover, grid refinement in fluid flow simulation and real-time drilling management can be done consistently under the model defined here.

Acknowledgements

We thank Statiol for permission to publish this paper.

References

[1] Abrahamsen, P. (1997) A review of Gaussian random fields and correlation functions, Technical report 878, Norwegian Computing Centre, Norway.

[2] Adler, R.J. (1981) *The Geometry of Random Fields*, John Wiley & Sons, New York

[3] Buland, A and Omre, H. (2003a) Bayesian Linearized AVO inversion, *Geophysics*, **68** 185-198

[4] Buland, A and Omre, H. (2003b) Bayesian Wavelet Estimation from Seismic and Well Data, *Geophysics*, **68**, 2000-2009

[5] Buland, A and Omre, H. (2003c) Joint AVO Inversion, Wavelet and Noise Level Estimation, *Geophysical Prospecting*, **6**, 531-550

[6] Buland, A., Kolbjørnsen, O. and Omre, H. (2003d) Rapid Spatially Coupled AVO inversion in the Fourier Domain, *Geophysics*, **68**, 824-836

[7] Christakos, G. (1992) *Random field models in earth sciences*, Academic Press Inc., San Diego.

[8] Eberhart-Phillips, D., Han, D-H. and Zoback M. D. (1989) Empirical relationships among seismic velocity, effective pressure, porosity, and clay content in sandstone, *Geophysics*, **54**, 82-89

[9] Frykman, P. and Deutsch, C.V. (2002) Practical Application of Geostatistical Scaling Laws for Data Integration, *Petrophysics*, **43**(3), p.153-171

[10] Han, D-H., Nur, A. and Morgan D. (1986) Effects of porosity and clay content on wave velocities in sandstones, *Geophysics*, **51**, 2093-2107

[11] Johnson, Wichern (1992) *Applied Multivariate Analysis*, Prentice-Hall, Inc.

[12] Wyllie, M.R.J., Gregory, A.R. and Gardner, L.W. (1956) Elastic wave velocities in heterogenous and porous media, *Geophysics*, **21**, 41-70

Appendix

A Marginal Distribution for Observations

From expressions (13) and (32) and the properties of the Gaussian distribution, the seismic observations d_s will be multivariate Gaussian distributed,

$$d_s \sim N_{n_s}\left(\mu_{d_s}, \Gamma_{d_s}\right) \tag{62}$$

with μ_{d_s} a $(n_s \times 1)$-dimensional expectation vector defined by

$$\begin{aligned} \mu_{d_s}^{\mathbf{x}t\theta} &= \mathrm{E}\left\{d_s^{\theta}(\mathbf{x},t)\right\} && (63)\\ &= \int_{\mathcal{L}^s} \psi(\mathbf{x}, t-v) a(\theta)^T \mathrm{E}\left\{\frac{\partial m_L(\mathbf{x},v)}{\partial v}\right\} \mathrm{d}v && (64) \end{aligned}$$

and Γ_{d_s} a $(n_s \times n_s)$-dimensional covariance matrix defined by

$$\begin{aligned} \gamma_{d_s}^{\mathbf{x}t\theta} &= \mathrm{Cov}\left\{d_s^{\theta_1}(\mathbf{x}_1,t_1), d_s^{\theta_2}(\mathbf{x}_2,t_2)\right\} && (65)\\ &= \int_{\mathcal{L}^s}\int_{\mathcal{L}^s} \psi(\mathbf{x}_1, t_1-v_1) a(\theta_1)^T \gamma_{m_L'}^{\mathbf{x}v} a(\theta_2) \psi(\mathbf{x}_2, t_2-v_2) \mathrm{d}v_1 \mathrm{d}v_2 + \gamma_{\epsilon_w}^{\mathbf{x}t\theta} \end{aligned}$$

From expressions (18) and (37) and the properties of the Gaussian distribution, the well observations will be approximately multivariate Gaussian distributed,

$$d_w \sim N_{n_w}\big(\mu_{d_w}, \Gamma_{d_w}\big) \tag{66}$$

with μ_{d_w} a $(n_w \times 1)$-dimensional expectation vector defined by

$$\begin{aligned} \mu_{d_w}^{\mathbf{x}t} &= \mathrm{E}\left\{d_w(\mathbf{x}, t)\right\} && (67)\\ &= \int_{\mathcal{L}^w} \xi(\mathbf{x}-\mathbf{u}, t-v) B \mathrm{E}\big\{\exp\{m_L(\mathbf{u}, v)\}\big\} \mathrm{d}\mathbf{u}\mathrm{d}v && (68) \end{aligned}$$

and Γ_{d_w} a $(n_w \times n_w)$-dimensional covariance matrix defined by

$$\begin{aligned} \gamma_{d_w}^{\mathbf{x}t} &= \mathrm{Cov}\left\{d_w(\mathbf{x}_1, t_1), d_w(\mathbf{x}_2, t_2)\right\} && (69)\\ &= \int_{\mathcal{L}^w}\int_{\mathcal{L}^w} \xi(\mathbf{x}_1-\mathbf{u}_1, t_1-v_1) B \mu_m^{\mathbf{u}_1 v_1} \gamma_{m_L}^{\mathbf{u}v} \mu_m^{\mathbf{u}_2 v_2} \\ &\qquad \cdot B^T \xi(\mathbf{x}_2-\mathbf{u}_2, t_2-v_2) \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{u}_2 \mathrm{d}v_1 \mathrm{d}v_2 + \gamma_{\epsilon_w}^{\mathbf{x}t} \end{aligned}$$

For the joint distribution of the seismic and well observations we have

$$d = [d_s^T, d_w^T]^T \sim N_{n_d}(\mu_d, \Gamma_d) \tag{70}$$

with μ_d the $(n_d \times 1)$-dimensional expectation vector

$$\mu_d = \begin{bmatrix} \mu_{d_s} \\ \mu_{d_w} \end{bmatrix} \tag{71}$$

and Γ_d the $(n_d \times n_d)$-dimensional covariance matrix

$$\Gamma_d = \begin{bmatrix} \Gamma_{d_s} & \Gamma_{d_s d_w} \\ \Gamma_{d_w d_s} & \Gamma_{d_w} \end{bmatrix}. \tag{72}$$

Here $\Gamma_{d_s d_w} = \Gamma_{d_w d_s}^T$ is a $(n_s \times n_w)$-dimensional cross-covariance matrix defined by

$$\begin{aligned} \gamma_{d_s d_w}^{\mathbf{x}t\theta\mathbf{x}t} &= \mathrm{Cov}\left\{d_s^\theta(\mathbf{x}_1, t_1), d_w(\mathbf{x}_2, t_2)\right\} && (73)\\ &= \int_{\mathcal{L}^s}\int_{\mathcal{L}^w} \psi(\mathbf{x}_1, t_1-v_1) a(\theta)^T \gamma_{m'_L m}^{\mathbf{x}_1 v_1 \mathbf{u}_2 v_2} \\ &\qquad \cdot B^T \xi(\mathbf{x}_2-\mathbf{u}_2, t_2-v_2) \mathrm{d}\mathbf{u}_2 \mathrm{d}v_1 \mathrm{d}v_2 \end{aligned}$$

The notation is as in Section 3.

B Results

Result B.1 Convergence of Multivariate Lognormal distribution

Assume that $Y \sim N_n(\mu, C)$ *and* $U \sim N_n(0, I)$.
If Prob $\left\{(\text{diag}(\mu))^{-1} C^{\frac{1}{2}} U \ll 1_n\right\} \to 1_n$, *then*

$$e^Y \xrightarrow{\sim} N_n\big(e^\mu, \, diag(e^\mu)\; C \; diag(e^\mu)\big) \tag{74}$$

with $\xrightarrow{\sim}$ *denoting convergence in distribution and* 1_n *being a vector of* n *ones.* ∎

Proof

If $Y \sim N_n(\mu, C)$ then $Y = \mu + C^{\frac{1}{2}} U$ with $U \sim N_n(0, I_n)$. Thus the ith component of Y may be expressed as $Y_i = \mu_i + \sum_{j=1}^n a_{ij} U_j$ for $i = 1, \ldots, n$ with $\{a_{ij}\}_{i,j=1}^n$ being the elements of $C^{\frac{1}{2}}$. With $e^Y = \left[e^{Y_1}, \ldots, e^{Y_n}\right]^T$ one finds

$$\begin{aligned} e^{Y_i} &= \exp\left\{\mu_i \left(1 + \frac{1}{\mu_i} \sum_{j=1}^n a_{ij} U_j\right)\right\} && (75) \\ &\approx 1 + \sum_{k=1}^\infty \frac{\mu_i^k}{k!} \left(1 + \frac{k}{\mu_i} \sum_{j=1}^n a_{ij} U_j\right) && (76) \\ &= e^{\mu_i} + e^{\mu_i} \sum_{j=1}^n a_{ij} U_j && (77) \end{aligned}$$

for $i = 1, \ldots, n$ where the approximation is valid under the assumtion of the theorem, as

$$\frac{1}{\mu_i} \sum_{j=1}^n a_{ij} U_j \sim N\left(0, \sum_{j=1}^n (\frac{a_{ij}}{\mu_i})^2\right), \tag{78}$$

which approaches a Dirac Delta function, as $\frac{a_{ij}}{\mu_i} \to 0$ whenever $n < \infty$. Returning to vector notation the result follows. ∎

Result B.2 Covariance between multivariate Gaussian, differentiated Gaussian and Log-normally distributed random variables

If the random variable $Y = Y(t)$ *is n-variate Gaussian distributed,* $Y \sim N_n(\mu, C)$, *then* $\exp(Y) \sim LogN_n(\mu, C)$ *and* $\partial_t Y \sim N_n(\partial_t \mu, \partial_t \partial_s C)$. *Further, if* $C = C(t, s)$ *is symmetric,*

1. $\text{Cov}\{Y, \exp(Y)\} = C D_{\mu, C}$

2. $\mathrm{Cov}\{\exp(Y), Y\} = D_{\mu,C}C$

3. $\mathrm{Cov}\{\partial_t Y, \exp(Y)\} = \partial_t C D_{\mu,C}$

4. $\mathrm{Cov}\{\exp(Y), \partial_t Y\} = D_{\mu,C}\partial_t C$

with $D_{\mu,C} = diag(\exp\{\mu + diag(C)\})$. ∎

Proof

We have

$$\begin{aligned}
\mathbb{E} &= \mathrm{E}\{\exp(Y)\} && (79)\\
&= c_\mu \cdot c \int_{\mathbb{R}^n} \begin{bmatrix} e^{y_1} \\ \vdots \\ e^{y_n} \end{bmatrix} \exp\left\{-\frac{1}{2}\left(y^T A y - y^T A \mu - \mu^T A y\right)\right\} \mathrm{d}y && (80)\\
&= c_\mu \cdot c \int_{\mathbb{R}^n} \mathcal{I}(y) \mathrm{d}y && (81)
\end{aligned}$$

with $c = \frac{1}{(2\pi)^{n/2}} \cdot |A|^{1/2}$, $c_\mu = \exp\left\{-\frac{1}{2}\left(\mu^T A \mu\right)\right\}$, $A = C^{-1}$, and $\mathcal{I}(y)$ defined from expression (80). The kth element of $\mathcal{I}(y)$ is

$$\begin{aligned}
\mathcal{I}_k(y) &= \exp\left\{-\frac{1}{2} y^T A y + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n y_i a_{ij} \mu_j + \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \mu_i a_{ij} y_j + y_k\right\}\\
&= \exp\left\{-\frac{1}{2}\left(y^T A y - \sum_{i=1}^n y_i \sum_{j=1}^n \mu_j (a_{ij} + a_{ji}) - 2y_k\right)\right\} && (82)\\
&= \exp\left\{-\frac{1}{2}(y - b^k)^T A (y - b^k)\right\} && (83)
\end{aligned}$$

for $k = 1, \ldots, n$. By completion of the square one finds

$$b^k = \mu + C_k$$

for $k = 1, \ldots, n$ whenever C is symmetric. Here C_k is the kth column vector of C. Now

$$\mathbb{E}_k = c_\mu c \int_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}(y - b^k)^T C^{-1} (y - b^k)\right\} dx \cdot \exp\left\{\frac{1}{2}(b^k)^T C^{-1} b^k\right\}$$

for $k = 1, \ldots, n$. As the integrals are equal to c^{-1}, we have

$$\begin{aligned}
\mathrm{E}\{\exp(Y)\} &= \exp\left\{-\frac{1}{2}\mu^T C^{-1}\mu\right\} \begin{bmatrix} \exp\left\{\frac{1}{2}(b^1)^T C^{-1} b^1\right\} \\ \vdots \\ \exp\left\{\frac{1}{2}(b^n)^T C^{-1} b^n\right\} \end{bmatrix} \\
&= \exp\left\{\begin{bmatrix} C_1^T C^{-1}(\mu + \frac{1}{2}C_1) \\ \vdots \\ C_n^T C^{-1}(\mu + \frac{1}{2}C_n) \end{bmatrix}\right\} \\
&= \exp\left\{\begin{bmatrix} C_1^T \\ \vdots \\ C_n^T \end{bmatrix} C^{-1}\mu + \frac{1}{2}\mathrm{diag}\begin{bmatrix} C_1^T C^{-1} C_1 & \ldots & C_1^T C^{-1} C_n \\ \vdots & \ddots & \vdots \\ C_n^T C^{-1} C_n & \ldots & C_n^T C^{-1} C_n \end{bmatrix}\right\} \\
&= \exp\left\{\mu + \frac{1}{2}\mathrm{diag}\left(\frac{1}{n}\begin{bmatrix} C^T \\ \vdots \\ C^T \end{bmatrix}^T \begin{bmatrix} C & & 0 \\ & \ddots & \\ 0 & & C \end{bmatrix}^{-1} \begin{bmatrix} C \\ \vdots \\ C \end{bmatrix}\right)\right\} \\
&= \exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}
\end{aligned}$$

Proof of 1.
We have

$$\mathrm{Cov}\{Y, \exp(Y)\} = \mathrm{E}\{Y \exp(Y)^T\} - \mu \mathrm{E}\{\exp(Y)^T\}.$$

with

$$\mathrm{E}\{Y \exp(Y)^T\} = c \int_{\mathbb{R}^n} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} [e^{y_1} \ldots e^{y_n}] \exp\left\{-\frac{1}{2}(y-\mu)^T A (y-\mu)\right\} dy \qquad (84)$$

which, by comparison with the previous calculations, is reduced to n^2 factors of the kind

$$\begin{aligned}
\mathbb{E}_{kl} &= \int_{\mathbb{R}^n} c\, y_l \exp\left\{-\frac{1}{2}(y-b^k)^T C^{-1}(y-b^k)\right\} dy \\
&\qquad \cdot c_\mu \exp\left\{\frac{1}{2}(b^k)^T C^{-1} b^k\right\} \\
&= \mathrm{E}\{Y_l\}|_{N(b^k, C)}\ \mathrm{E}\{\exp(Y_k)\} \\
&= b_l^k \mathrm{E}\{\exp(Y_k)\}
\end{aligned}$$

for $k, l = 1 \ldots n$. Thus, with

$$B = \begin{bmatrix} b_1^1 & \ldots & b_1^n \\ \vdots & \ddots & \vdots \\ b_n^1 & \ldots & b_n^n \end{bmatrix} = \mu^T 1_{n\times n} + C$$

and $1_{n\times n}$ being an $n \times n$-matrix of all ones,

$$\begin{aligned}
\mathrm{Cov}\{Y, \exp(Y)\} &= B\mathrm{diag}\left(\mathrm{E}[\exp(Y)]\right) - \mu\mathrm{E}[\exp(Y)^T] \\
&= (\mu^T 1_{n\times n} + C)\ \mathrm{diag}\left(\exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}\right) \\
&\quad -\mu\left[\exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}\right]^T \\
&= C\ \mathrm{diag}\left(\exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}\right)
\end{aligned}$$

Proof of 3.
We have

$$\mathrm{Cov}\{\partial_t Y, \exp Y\} = \mathrm{E}\{\partial_t Y(\exp Y)^T\} - \partial_t \mu c_\mu D_b \tag{85}$$

with

$$\mathrm{E}\{\partial_t Y(\exp Y)^T\} = c\int_{\mathbb{R}^n} \partial_t Y(\exp Y)^T \exp\left\{-\frac{1}{2}(y-\mu)^T A(y-\mu)\right\} dy$$

which, by comparison with the previous calculations, is reduced to n^2 factors of the kind

$$\begin{aligned}
\mathbb{E}_{kl} &= \int_{\mathbb{R}^n} c\,[\partial_t y]_l \exp\left\{-\frac{1}{2}(y-b^k)^T C^{-1}(y-b^k)\right\} dy \cdot c_\mu \exp\left\{\frac{1}{2}(b^k)^T C^{-1} b^k\right\} \\
&= \mathrm{E}\{\partial_t Y_l\}|_{N(b^k,C)}\,\mathrm{E}\{\exp(Y_k)\} \\
&= \partial_t b_l^k \mathrm{E}\{\exp(Y_k)\}
\end{aligned}$$

for $k, l = 1 \ldots n$. Thus

$$\begin{aligned}
\mathrm{Cov}\{\partial_t Y, \exp(Y)\} &= \partial_t B\mathrm{diag}\left(\mathrm{E}\{\exp(Y)\}\right) - \partial_t \mu \mathrm{E}\{\exp(Y)^T\} \\
&= \partial_t(\mu^T 1_{n\times n} + C)\ \mathrm{diag}\left(\exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}\right) \\
&\quad -\partial_t \mu\left[\exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}\right] T \\
&= \partial_t C\ \mathrm{diag}\left(\exp\left\{\mu + \frac{1}{2}\mathrm{diag}(C)\right\}\right)
\end{aligned}$$

Proof of 2 and 4.
Since

$$\left(\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}[e^{y_1} \ldots e^{y_n}]\right)^T = \begin{bmatrix} e_1^y \\ \vdots \\ e^{y_n} \end{bmatrix}[y_1 \ldots y_n]$$

the result follows. ∎

C Discretization process

A discretized version of the convolution (7) for the seismic data in the jth trace is

$$d_{s_j} = S_j(A_j m'_{L,s_j}) + e_{s_j} \tag{86}$$

for all observed traces in an angle gather. Here S_j is a matrix version of the source wavelet, A_j a matrix of constants dependent on the seismic angle, m_{L,s_j} the logarithm of the elastic parameters in the trace, e_{s_j} the associated error term, and the prime denoting partial differentiation with respect to time t. For details on the individual elements of these matrices, see Buland & Omre (2003a). In Section 3.1 the seismic error term was modelled as Gaussian,

$$e_{s_j} \sim N\left(0, C_{e_{s_j}}\right), \tag{87}$$

and in Section 3.2 the elastic parameters m were assumed to a priori Log-Gaussian, resulting in m'_L being Gaussian

$$m'_{L,s_j} \sim N\left(\mu_{m'_L,s_j}, C_{m'_L,s_j}\right), \tag{88}$$

In (87) and (88) $C_{e_{s_j}}$, $\mu_{m'_L}$ and $C_{m'_L}$ are the discrete counterparts of the corresponding continuous functions, with appropriate dimension. Having n_d seismic observations in a trace and n_θ angle gathers, we find

$$d_{s_j} \sim N_{n_d \cdot n_\theta}\left(S_j A_j \mu_{m'_{L,s_j}}, S_j(A_j C_{m'_{L,s_j}} A_j^T) S_j^T + C_{e_{s_j}}\right) \tag{89}$$

for $j = 1 \ldots n_t$, where n_t is the number of seismic traces. For the collection of all seismic gathers $d_s = [d_{s_1}^T, \ldots, d_{s_{n_t}}^T]^T$ within the 3D volume, expression (86) generalizes to

$$d_s = S A m'_{L,S} + e_S \tag{90}$$

with $S = \text{diag}[S_1, \ldots, S_{n_t}]$, $A = \text{diag}[A_1, \ldots, A_{n_t}]$, $e_S = [e_{S_1}^T, \ldots, e_{S_{n_t}}^T]^T$ and $m'_{L,s} = [m'^T_{L,s_1}, \ldots, m'^T_{L,s_{n_t}}]^T$. Thus the n_t seismic gathers are jointly Gaussian distributed

$$d_s \sim N_{n_d \cdot n_\theta \cdot n_t}\left(S A \mu_{m'_{L,s}}, S A C_{m'_{L,s}} A^T S^T + C_{e_S}\right) \tag{91}$$

with $\mu_{m'_{L,s}} = [\mu^T_{m'_{L,s_1}}, \ldots, \mu^T_{m'_{L,s_{n_t}}}]^T$ and

$$C_{m'_{L,s}} = \begin{bmatrix} C_{m'_{L,s_1}} & \cdots & C_{m'_{L,s_1},m'_{L,s_{n_t}}} \\ \vdots & \ddots & \vdots \\ C_{m'_{L,s_{n_t}},m'_{L,s_1}} & \cdots & C_{m'_{L,s_{n_t}}} \end{bmatrix}$$

$$C_{e_s} = \begin{bmatrix} C_{e_{s_1}} & \cdots & C_{e_{s_1},e_{s_{n_t}}} \\ \vdots & \ddots & \vdots \\ C_{e_{s_{n_t}},e_{s_1}} & \cdots & C_{e_{s_{n_t}}} \end{bmatrix}$$

under the assumption that $\text{Cov}(m'_{s_j}, e_{s_k}) = 0$ for $j, k = 1, \ldots n_t$. If one assumes that the errors associated with each seismic trace is individual, that is, $\text{Cov}(e_{s_j}, e_{s_k}) = 0$ for $j, k = 1, \ldots n_t$, $j \neq k$, then C_{e_s} will be block diagonal.

A discretized version of the convolution (11) for the well data is

$$d_w \;=\; TB_w m_w + e_T \tag{92}$$

Here T is a matrix version of the averaging function and B_w an expanded matrix version of B, and m_w the elastic parameters along the well, and e_T the associated error term. In Section (3.1) we modelled the well error term as Gaussian,

$$e_T \sim N\left(0, C_{e_T}\right), \tag{93}$$

and in Section 3.2 the elastic parameters m were assumed to be a priori Log-Gaussian, resulting in m_w being approximately Gaussian through result B.1

$$m_w \overset{\sim}{\rightarrow} \left(\mu_{m_w}, C_{m_w}\right). \tag{94}$$

In (93) and (94) C_{e_T}, μ_{m_w} and C_{m_w} are the discrete counterparts of the corresponding continuous functions, with appropriate dimension. We find

$$d_w \;\overset{\sim}{\rightarrow}\; N_{n_w}\left(TK\mu_{m_w}\ ,\ TB_w C_{m_w} B_w^T T^T + C_{e_T}\right) \tag{95}$$

with n_w the sum of observations of α, β and ρ in the well.

In total, for the posterior distribution (43), the model parameters μ_{d_s} and C_{d_s} are thus found from expression (91), and μ_{d_w} and C_{d_w} from expression (95), whereas

$$\begin{aligned}
C_{m_L,d_s} &= \text{Cov}(m_L, SAm'_{L,s} + e_S) \\
&= C_{m_L,m'_{L,s}} A^T S^T, \\
C_{m_L,d_w} &= \text{Cov}(m_L, TB_w m_w + e_T) \\
&= C_{m_L,m_w} B_w^T T^T, \\
C_{d_s,d_w} &= \text{Cov}(SAm'_{L,s} + e_S\ ,\ TB_w m_w + e_T) \\
&= SAC_{m'_{L,s},m_w} B_w^T T^T.
\end{aligned}$$

and so on, with $C_{m_L,m'_{s,L}}$, C_{m_L,m_w} and $C_{m'_{s,L},m_w}$ being the discrete counterparts of the corresponding continous functions defined in Section 3 and Appendix A.

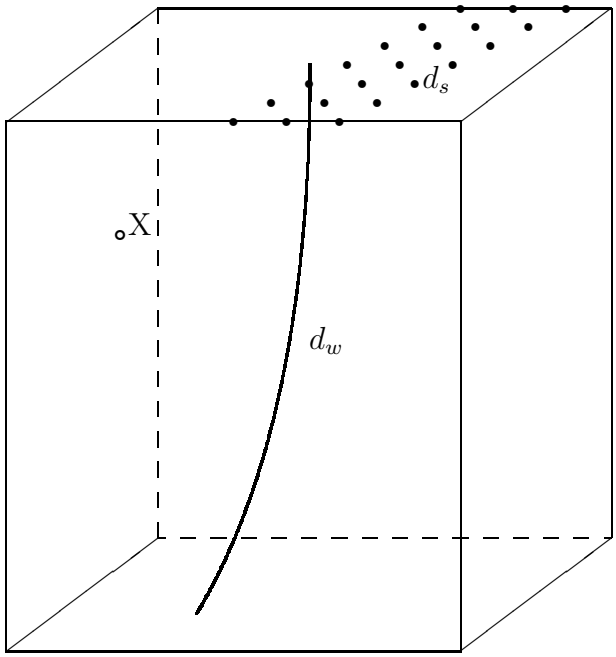

Figure 1: *Schematical overview of seismic and well observations within a 3D volume.*

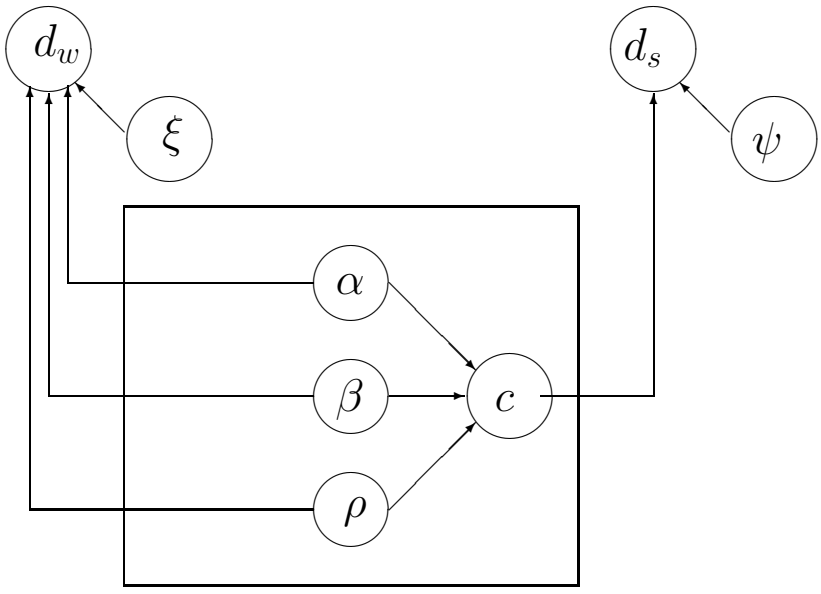

Figure 2: *Graph of connections between the elastic parameters and the observations d_w and d_s. Seismic wavelet function ψ and well averaging function ξ.*

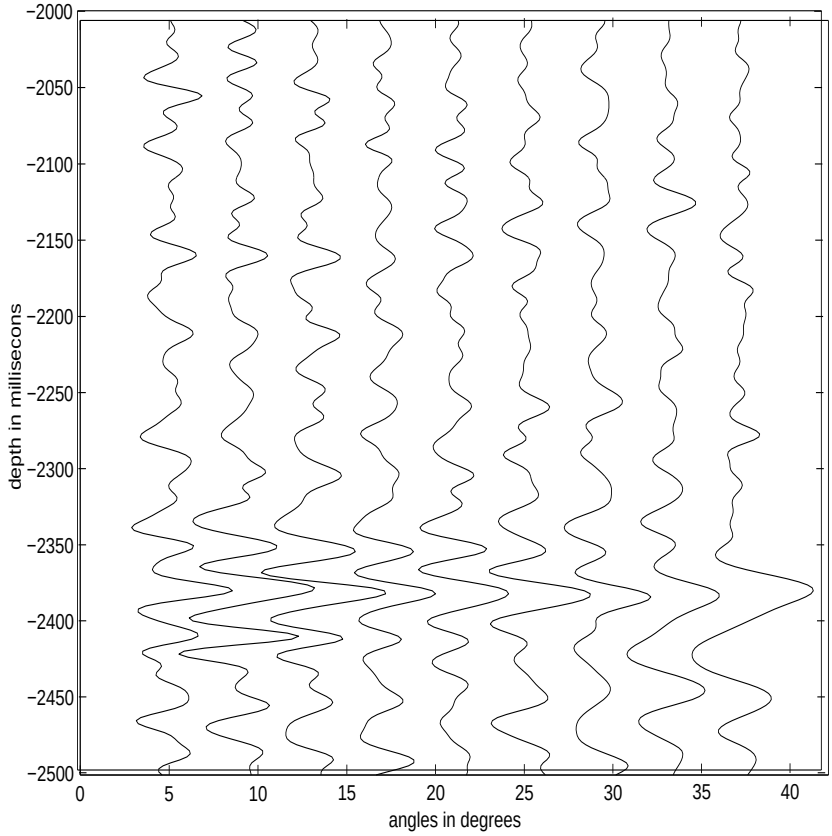

Figure 3: *A typical seismic gather in the area of study.*

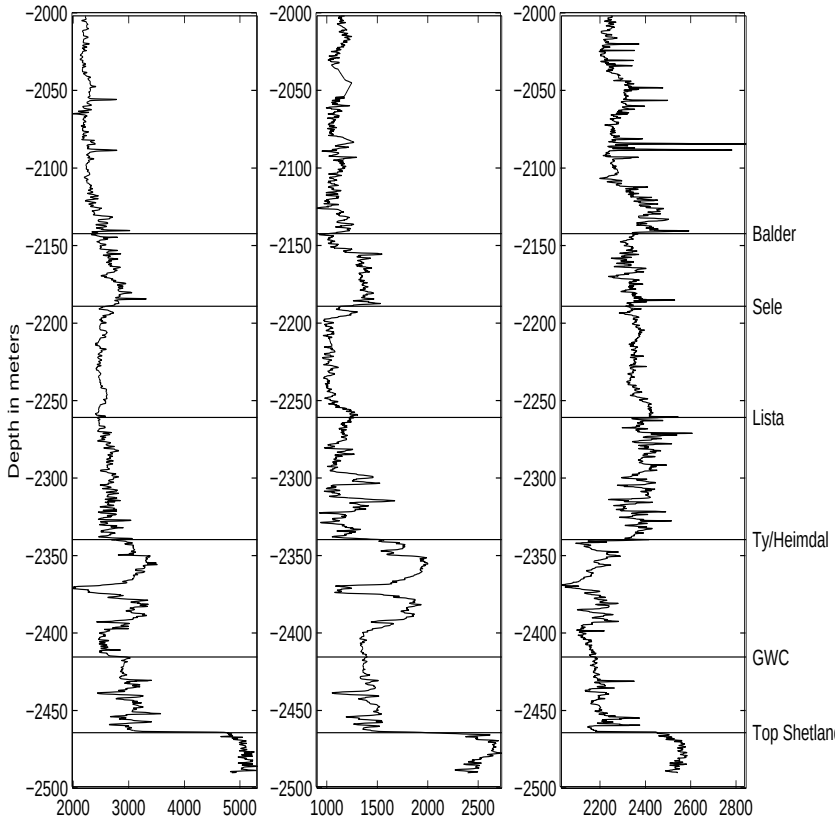

Figure 4: *Well logs from the Sleipner field: P-wave velocity (left), S-wave velocity (middle) and bulk density (right).*

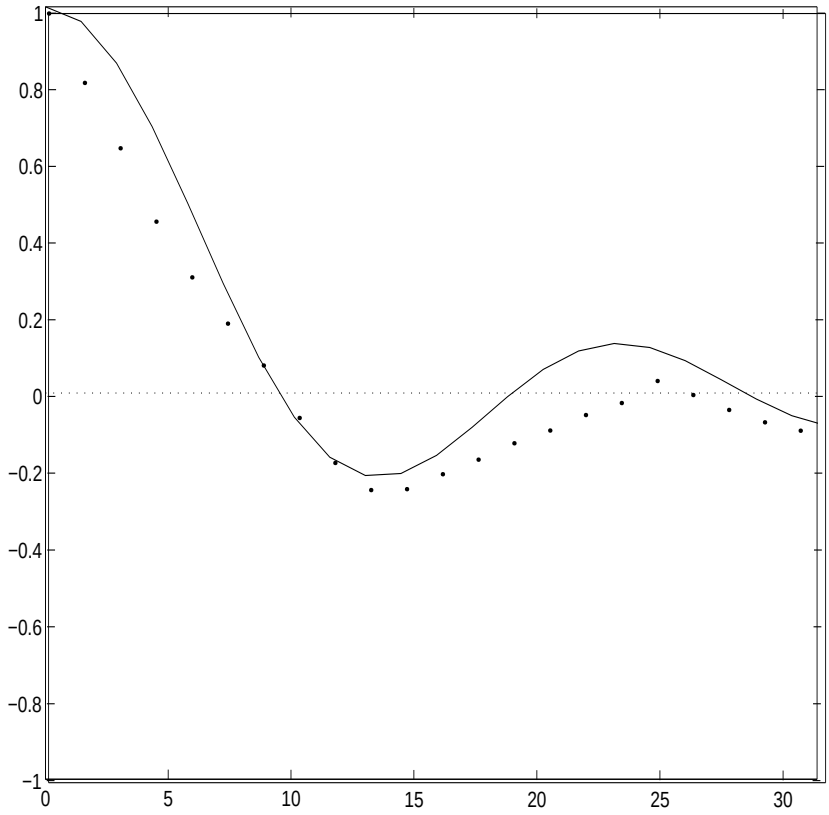

Figure 5: *Empirical correlation function* $(\cdots)$ *with the inferred ordinary Bessel correlation function* $(—)$.

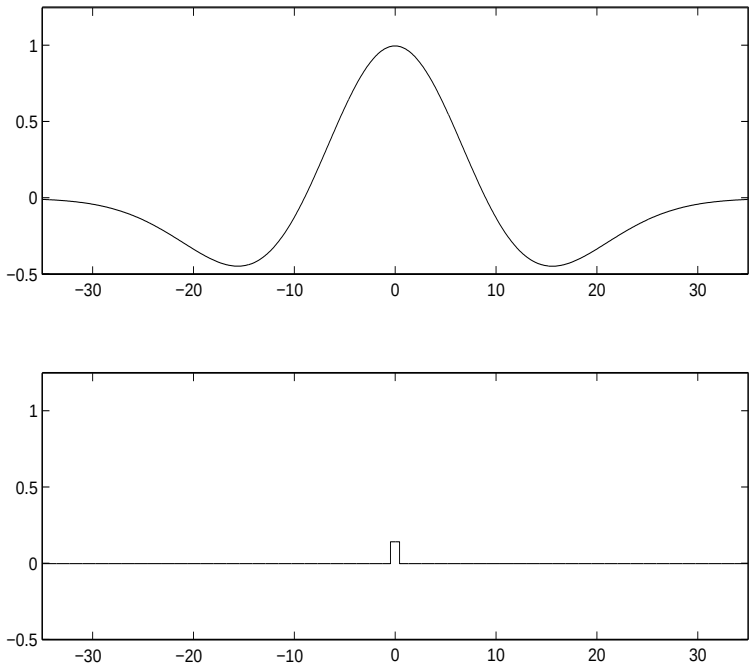

Figure 6: *Seismic zero phase Ricker wavelet ψ (upper) and uniform well averaging function ξ (lower).*

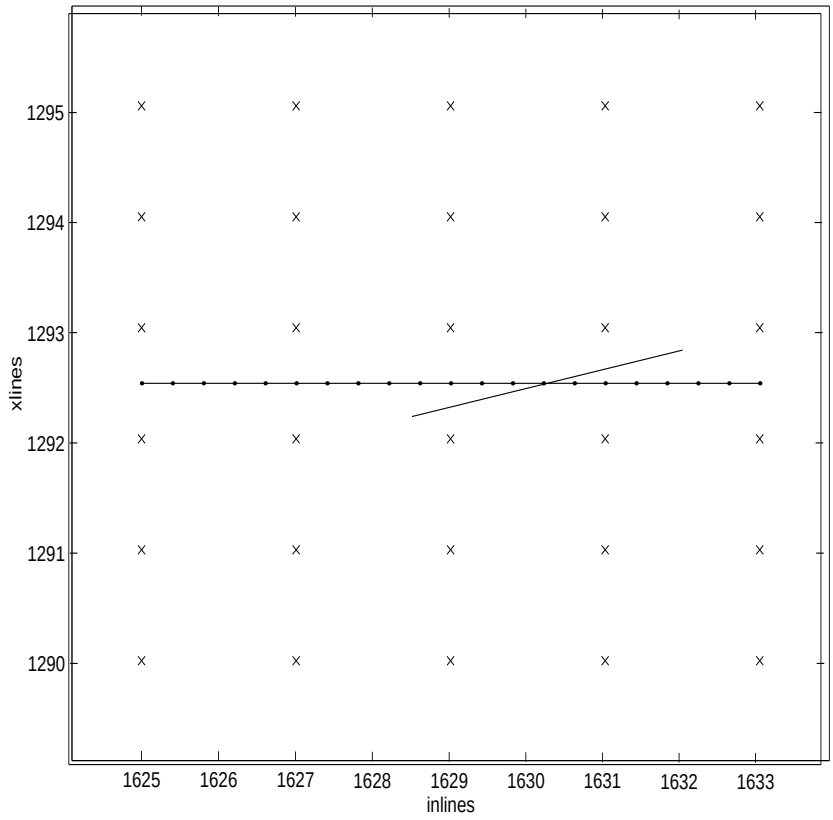

Figure 7: *Location of seismic gathers* ($\times$)*, 2D cut of inversion traces* ($- \cdot - \cdot -$) *and well from depth 2280ms to 2355ms* (—)*. Entry point for the well is at the upper right, exit at the lower left.*

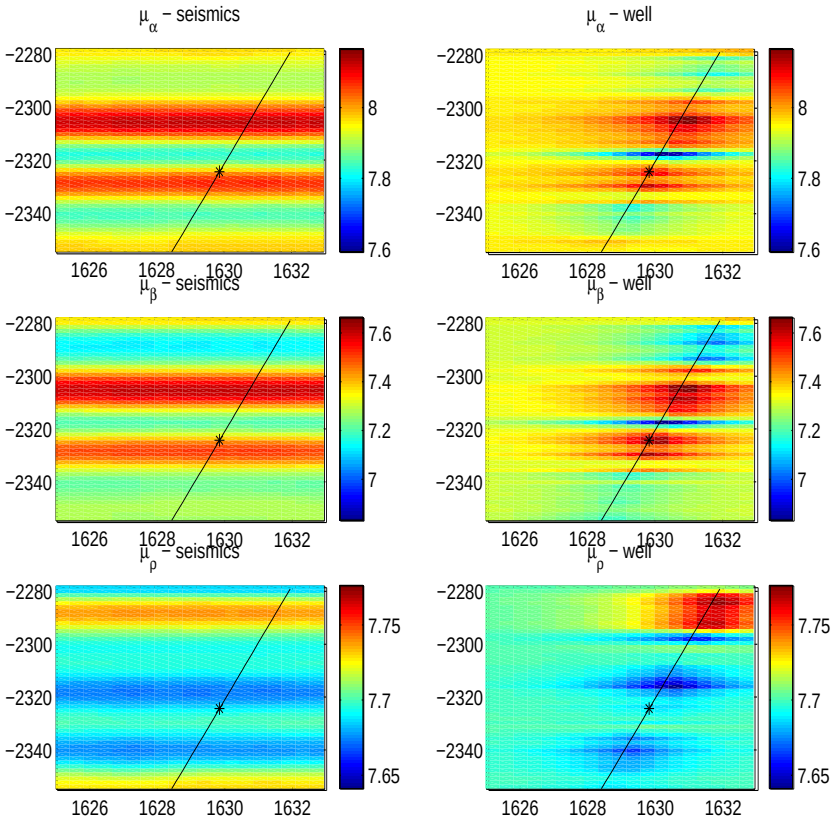

Figure 8: *Conditional mean after inversion conditioned on seismic observations only (left) and well logs only (right). The star marks where the well penetrates the 2D cut. Above the star the well lies "behind" the display, below the star the well lies "in front of" the display.*

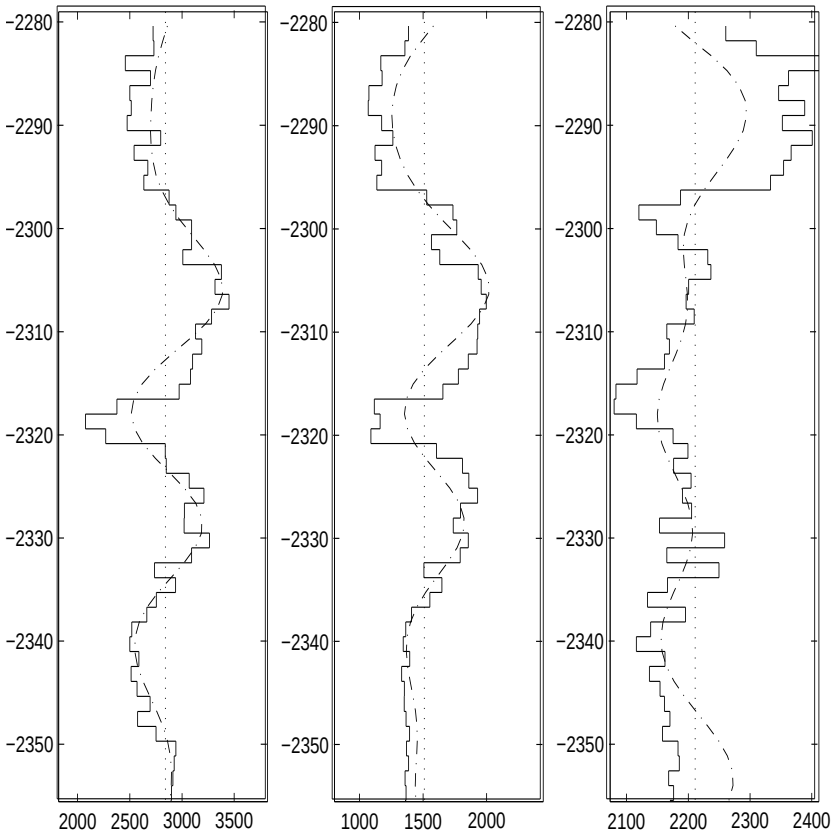

Figure 9: *Conditional mean* $(- \cdot -)$ *along the well path after inversion conditioned on seismic observations only.Well log data* $(—)$*.*

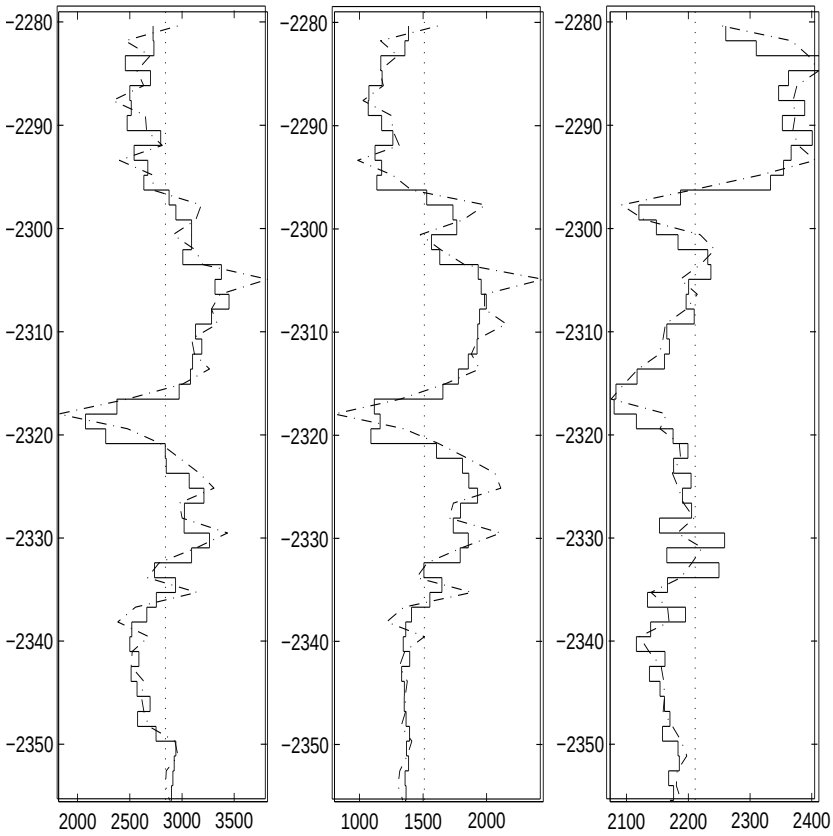

Figure 10: *Conditional mean* $(-\cdot-)$ *along the well path after inversion conditioned on well logs only. Well log data* $(—)$*.*

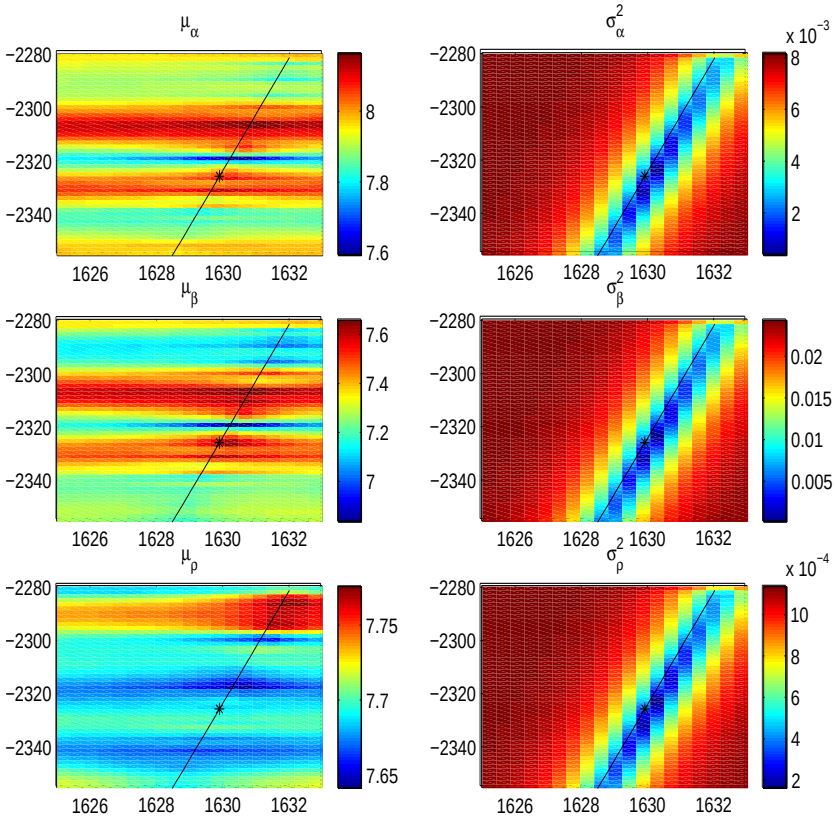

Figure 11: *Inversion conditoned on both seismic and well information: Conditional mean (left) and conditional variances (right). The star marks where the well penetrates the 2D cut. Above the star the well lies "behind" the display, below the star the well lies "in front of" the display.*

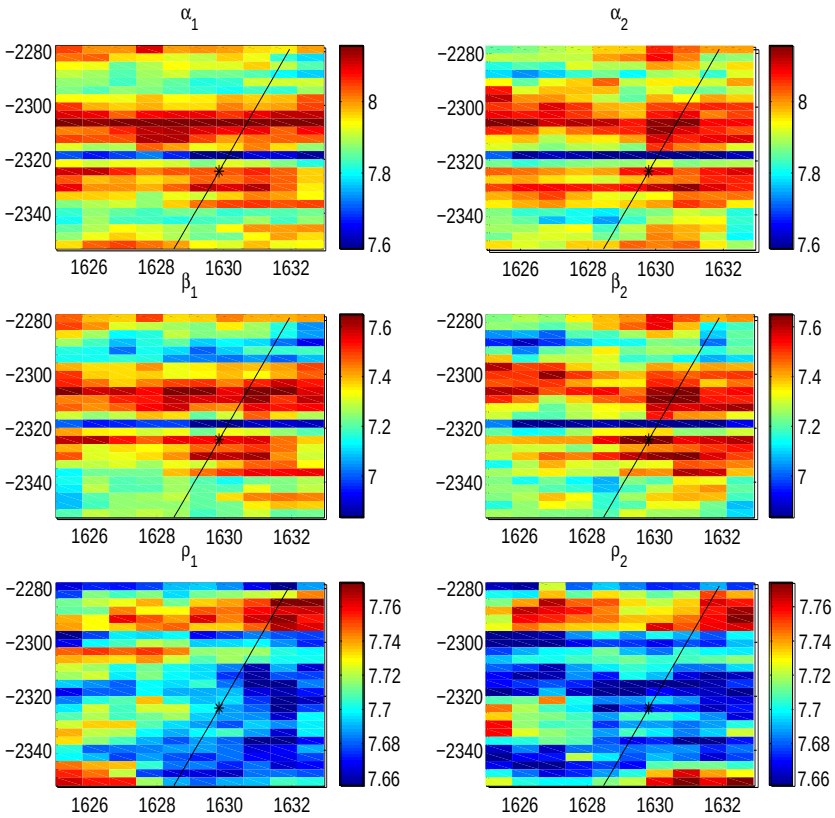

Figure 12: *Two samples (left and right) from the posterior distribution. The star marks where the well penetrates the 2D cut. Above the star the well lies "behind" the display, below the star the well lies "in front of" the display.*

Discretized Representations of Continuous Random Fields

Jo Røislien
Department of Petroleum Engineering and Applied Geophysics
Norwegian University of Science and Technology

Alberto Malinverno, Michael Prange, Nick Bennett
Schlumberger-Doll Research
Ridgefield, CT, USA

Abstract

In geophysical inverse problems, continuous fields of material properties must be approximated by a discrete basis function representation, which in its simplest form entails gridding. The inverse problem is then to determine the uncertainty in the coefficients of these basis functions given a set of noisy measurements. This work explores the relationship between the statistical properties of a continuous random field and the corresponding statistical properties of the basis function coefficients. The mean and covariance of the coefficients of a set of basis functions for a given mean and spatial covariance function of a continuous random field is determined. The established connection between a discrete representation of a property field and the underlying continuous field makes it possible to probabilistically compare different discrete representations. Moreover, consistency between different discrete representations can be ensured. The use of uniform averaging, linear interpolation and wavelets as discretization procedures are discussed.

URN:NBN:no-3476

1 Introduction

Geophysical inverse problems deal with the issue of making inference about subsurfacae properties, based on some geophysical measurements. These problems are generally ill-posed and non-unique. In a statistical setting such inverse-problems are naturally stated in a Bayesian setting, leading to solutions represented by a posterior probability density function (pdf) on the subsurface property space. This pdf is usually represented on a grid, which constitutes a discretization of the solution, see for example Buland and Omre (2003).

Another example is the choice of the prior distribution in Bayesian inference. The prior distribution has the role of regularization, limiting the space of possible solutions, see for example Malinverno (2002). The choice of prior becomes increasingly important when the available data is sparse. Then the finitie dimensional representation of the infinite-dimensional prior should be handled with care.

Even though the subsurface properties are continuous, and most methodologies are based on discrete vector representations of these, little attention has been paid to the discretization itself. This work establishes a connection between a continuous random field and the discrete representation of this field using some set of basis functions. Not surprisingly, the results simplify for Gaussian random fields, for which some general litterature exist, see for example Lifshits (1995) and Jansson (1997).

The article is organized as follows; in Section 2 the discretization procedure is outlined and the main results are established; Section 3 contains a discussion on how to evaluate and compare different choices of basis fuctions, whereas Section 4 contains some examples using different sets of basis functions; finally, Section 5 contains a summary.

2 Inversion to Basis Function Parameters

Assume a continuous random field $\{m(x); x \in \mathbb{R}^q\}$. We would like to represent this random field through a given set of basis functions $\Phi_n = \{f_i(x)\}_{i=1}^n$ as an expansion of the form

$$m(x) = \lim_{n\to\infty} \sum_{j=1}^{n} \theta_j f_j(x) \tag{1}$$

with $\theta = [\theta_1, \ldots, \theta_n]^T$ a vector of random variables. If $n < \infty$ expression (1) is an approximation of $m(x)$ onto a discrete, finite set of parameters. Setting $f(x) = [f_1(x), \ldots, f_n(x)]^T$ the approximation (1) can be written in compact notation as

$$m_{\Phi_n}(x) \quad = \quad f(x)^T\theta. \tag{2}$$

A random field on $\mathbb{R}^q$ is a function whose values are random variables for any $x \in \mathbb{R}^q$, see Abrahamsen (1994) or Lifshits (1995) for more formal definitions. Through the reparametrization (2) we seek to split the properties of the random field $m(x)$ into a deterministic spatial part $f(x)$ and a stochastic part θ.

Assume that the distribution of the random field $\{m(x); x \in \mathbb{R}^q\}$ is known, with expectation $\{\mu(x); x \in \mathbb{R}^q\}$ and spatial covariance function $\{C(x, y); x, y \in \mathbb{R}^q\}$, and that the set of basis functions Φ_n is given. We are interested in how the stochastic properties of $m(x)$ are inherited by θ. We concentrate on L^2 spaces, with inner product

$$\langle r, s\rangle = \int_{\mathcal{R}^q} r(x)s(x)\mathrm{d}x \tag{3}$$

for functions $r(x)$ and $s(x)$. From the basis function expansion (1) we then have

$$\langle m, f_i\rangle \;\; = \;\; \sum_{j=1}^{n} \theta_j \langle f_j, f_i\rangle \tag{4}$$

for $i = 1 \ldots n \leq \infty$. Introducing the notations

$$\hat{m}_i = \langle m, f_i\rangle \tag{5}$$

$$\hat{f}_{ij} = \langle f_i, f_j\rangle, \tag{6}$$

expression (4) results in the matrix equation

$$\begin{bmatrix} \hat{m}_1 \\ \vdots \\ \hat{m}_n \end{bmatrix} = \begin{bmatrix} \hat{f}_{11} & \ldots & \hat{f}_{n1} \\ \vdots & \ddots & \vdots \\ \hat{f}_{1n} & \ldots & \hat{f}_{nn} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \tag{7}$$

written in compact notation as

$$\hat{m} = F\theta \tag{8}$$

with $\hat{m}$ and $F = F^T$ defined from expression (7). We then have

$$\theta = H\hat{m}, \tag{9}$$

with $H = F^{-1}$ having elements $\{h_{ij}\}_{i,j=1}^{n}$. As long as the f_i's are linearly independent, F will have full rank. For each of the components of θ we have

$$\theta_i = \sum_{j=1}^{n} h_{ij}\hat{m}_j \tag{10}$$

with the h_{ij} being scalars. From expression (5) we observe that $\hat{m}_i$ is the smoothed random field of $m(x)$ with respect to the weighting function $f_i(x)$, see Appendix A. Thus, expression

(10) states that θ_i is a linear combination of n smoothed random fields. The case where Φ_n is an orthogonal set of basis functions is just a special case for the transformation $H = F^{-1}$, with $H = \text{diag}\left(\hat{f}_{11}^{-1}, \ldots, \hat{f}_{nn}^{-1}\right)$. Expression (10) then reduces to

$$\theta_i = \frac{\hat{m}_i}{\hat{f}_{ii}} \,. \tag{11}$$

Under orthonormality $\hat{f}_{ii} = 1$ and $F = H = I$.

If $m(x)$ is a Gaussian random field, the expectation and variance of $\hat{m}_i$ is

$$\begin{aligned} \text{E}\{\hat{m}_i\} &= \int f_i(x)\mu(x)\text{d}x & (12) \\ \text{Cov}\{\hat{m}_i, \hat{m}_j\} &= \int\int f_i(x)C(x,y)f_j(y)\text{d}x\text{d}y, & (13) \end{aligned}$$

see Appendix A.2. The distribution of $\hat{m}$ is

$$\hat{m} \sim N\left(\mu_{\hat{m}}, C_{\hat{m}}\right) \tag{14}$$

with $\mu_{\hat{m}}$ and $C_{\hat{m}}$ defined from expressions (12) and (13) respectively. Thus

$$\theta = H\hat{m} \sim N\left(H\mu_{\hat{m}}, HC_{\hat{m}}H^T\right). \tag{15}$$

Further, the reparametrization $m_{\Phi_n}(x)$, see expression (2), will be a continuous Gaussian random field with expectation and covariance function

$$\begin{aligned} \text{E}\{m_{\Phi_n}(x)\} &= f(x)^T H\mu_{\hat{m}} & (16) \\ \text{Cov}\{m_{\Phi_n}(x), m_{\Phi_n}(y)\} &= f(x)^T HC_{\hat{m}}H^T f(y). & (17) \end{aligned}$$

3 Evaluating the Discretization

So far we have been concentrating on the properties of the parameter vector θ under the reparametrization (2) given the sets $\mathcal{M} = \{\mu(x), C(x,y)\}$and $\Phi_n = \{f_i(x)\}_{i=1}^n$. Clearly, the choice of basis functions Φ_n with respect to a given $\mathcal{M}$ is not indifferent, as some choices are better than others. Figure 1 shows two different fits to a given $\mu(x)$ resulting from two different sets of basis functions. The approximation error at a location x for the reparametrization, using the notation (2) for $m_{\Phi_n}(x)$, is

$$\begin{aligned} m_\Delta(x) &= m(x) - m_{\Phi_n}(x) & (18) \\ &= m(x) - f(x)^T H\int f(x)m(x)\text{d}x & (19) \\ &= K(m(x)) & (20) \end{aligned}$$

with $K(m(x))$ denoting the linear transformation defined from expression (19). If $m(x)$ is a Gaussian random field, the approximation error $m_\Delta(x)$ is also a Gaussian random field when conditioning on the choice of set of basis functions Φ_n. This new Gaussian random field has expectation and variance

$$\begin{aligned}
\mu_\Delta(x) &= \mu(x) - f(x)^T H \int \mu(x) f(x) \mathrm{d}x && (21)\\
C_\Delta(x,y) &= C(x,y) && (22)\\
&\quad - \int C(x,v) f(v)^T \mathrm{d}v H^T f(y) && (23)\\
&\quad - f(x)^T H \int C(y,u) f(u) \mathrm{d}u && (24)\\
&\quad + f(x)^T H \int \int f(u) C(u,v) f(v)^T \mathrm{d}u \mathrm{d}v H^T f(y) && (25)
\end{aligned}$$

respectively, see Appendix B for calculations. Assuming that Φ_n is an orthonormal set of basis functions, the approximation variance in a location x is

$$\begin{aligned}
\sigma^2_\Delta(x) &= C_\Delta(x,x) && (26)\\
&= C(x,x) - 2\sum_{i=1}^{n} \int f_i(x) C(x,v) f_i(v) \mathrm{d}v && (27)\\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} f_i(x) \int \int f_i(u) C(u,v) f_j(v) \mathrm{d}u \mathrm{d}v f_j(x) && (28)
\end{aligned}$$

with $C(x,x) = \sigma^2(x)$ the local variance of $m(x)$. The results will thus depend on the correlation function.

A global measure of misfit on the domain $\mathcal{D}$ of interest given a specific choice of Φ_n and $\mathcal{M}$, is the Mean Integrated Squared Error (MISE), see Boyd and Steele (1978). It is defined as

$$\begin{aligned}
\mathrm{MISE}(\Phi_n) &= \mathrm{E} \int_{\mathcal{D}} m^2_\Delta(x) \mathrm{d}x && (29)\\
&= \int_{\mathcal{D}} \mathrm{E}\{m_{\Phi_n}(x) - m(x)\}^2 \mathrm{d}x + \int_{\mathcal{D}} \mathrm{Var}\{m_{\Phi_n}(x)\} \mathrm{d}x && (30)
\end{aligned}$$

which is the sum of the square integrated bias and the integrated variance. To compute expression (30), we generally need to rely on numerical methods. The MISE also allows for comparison of different choices of sets of basis functions. We can calculate $\mathrm{MISE}(\Phi_n)$ for sets Φ_n given n and $\mathcal{M}$, and then use

$$\hat{\Phi}_n = \mathrm{argmin}_{\Phi_n} \{\mathrm{MISE}(\Phi_n)\} \quad (31)$$

as a criterion for choosing a certain set Φ_n of basis functions.

4 Examples of Basis Functions

We will here show examples of different sets of basis functions. The orthogonal basis used in uniform averaging is discussed in subsection 4.1, and the non-orthogonal linear interpolation basis is presented in subsection 4.2. Finally, wavelets are discussed in subsection 4.3.

4.1 Layered model – Uniform averaging

Assume a continuous Gaussian random field $\{m(x); x \in \mathcal{L} \subset \mathbb{R}^1\}$ with mean $\mu(x)$ and covariance function $C(x, y)$. We wish to approximate this continuous Gaussian random field with a layered model having n homogenous layers, see Figure 2. That is, we seek to project a continuous random field onto a discrete set of n random parameters $\theta = [\theta_1, \ldots, \theta_n]^T$ using orthogonal basis functions $f_j(x), j = 1, \ldots, n$, see expression (2). Each function $f_j(x)$ is chosen to average $m(x)$ over an interval, leading to the set of functions

$$f_j(x) = w_j \mathcal{I}(x_{j-1} < x < x_j) \text{ for } j = 1, \ldots, n \text{ and } x \in \mathcal{L}, \tag{32}$$

with $\mathcal{I}(A)$ denoting the indicator function for the event A, and the w_j's being scalars, see Figure 3. The functions (32) form an orthogonal basis as a straightforward calculation leads to

$$\langle f_i, f_j \rangle = \begin{cases} 0 & , \ i \neq j \\ w_j^2 \Delta x_j & , \ i = j \end{cases}, \tag{33}$$

with $\Delta x_j = x_j - x_{j-1}$. With the piecewise uniform weighting function described by expression (32), each of the n intervals are effectively averaged into a single value given by

$$\hat{m}_j = \int_{\mathcal{L}} w_j \mathcal{I}(x_{j-1} < x < x_j) m(x) \mathrm{d}x \tag{34}$$

$$= w_j \int_{\mathcal{L}_j} m(x) \mathrm{d}x \tag{35}$$

$$\tag{36}$$

with $\mathcal{L}_j = (x_{j-1}, x_j)$. The expectation and covariance are

$$\mathrm{E}\{\hat{m}_j\} = w_j \int_{\mathcal{L}_j} \mu(y) \mathrm{d}y \tag{37}$$

$$\mathrm{Cov}\{\hat{m}_i, \hat{m}_j\} = w_i w_j \int_{\mathcal{L}_i} \int_{\mathcal{L}_j} C(u, v) \mathrm{d}u \mathrm{d}v. \tag{38}$$

The parameter vector θ is distributed according to expression (15) with H being diagonal and

$$h_{jj} = \frac{1}{w_j^2 \Delta x_j}. \tag{39}$$

Componentwise we thus have

$$\theta_j = \frac{1}{w_j \Delta x_j} \int_{\mathcal{L}_j} m(x) \mathrm{d}x \tag{40}$$

from expression (11), with

$$\begin{aligned} \mathrm{E}\{\theta_j\} &= \frac{1}{w_j \Delta x_j} \int_{\mathcal{L}_j} \mu(x) \mathrm{d}x & (41) \\ \mathrm{Cov}\,\{\theta_i, \theta_j\} &= \frac{1}{w_i w_j \Delta x_i \Delta x_j} \int_{\mathcal{L}_i} \int_{\mathcal{L}_j} C(u, v) \mathrm{d}u \mathrm{d}v. & (42) \end{aligned}$$

The calculations of these integrals depend on the explicit form of the mean $\mu(x)$ and the covariance function $C(x, y)$. For calculations using the the exponential correlation function $\rho(|x - y|) = \exp(-\frac{|x-y|}{\lambda})$, see Appendix C. For more on correlation functions, see Abrahamsen (1994). As $m(x)$ is a Gaussian random field, θ will be multivariate Gaussian distributed, see expression (15).

As for choosing the weights w_j, a natural choice is to choose them such that the θ_j's contain the layer averages of $m(x)$, yielding

$$w_j = 1. \tag{43}$$

However, this is not a necessary criterion; for example, setting

$$w_j = \sqrt{\frac{1}{\Delta x_j}} \tag{44}$$

makes the basis (32) orthonormal, but lacks an intuitive physical interpretation.

4.2 Linear interpolation basis

The linear interpolation basis

$$f_i(x) = \begin{cases} 0 & , \quad 0 \leq x < x_{i-1} \\ w_i \frac{x - x_{i-1}}{x_i - x_{i-1}} & , \quad x_{i-1} \leq x < x_i \\ w_i \frac{x - x_{i+1}}{x_i - x_{i+1}} & , \quad x_i \leq x < x_{i+1} \\ 0 & , \quad x_{i+1} \leq x \end{cases}, \tag{45}$$

for $i = 1, \ldots, n$ shown in Figure 4, is a non-orthogonal basis. Straightforward calculations yield

$$\begin{aligned} \langle f_i, f_i \rangle &= \frac{w_i^2}{3} \left[(x_{i+1} - x_i) + (x_i - x_{i-1}) \right] = \frac{2}{3} w^2 \Delta x & (46) \\ \langle f_i, f_{i+1} \rangle &= \frac{w_i w_{i+1}}{6} (x_{i+1} - x_i) = \frac{1}{6} w^2 \Delta x & (47) \end{aligned}$$

where the rightmost expressions are the results of assuming equidistant spacing, $x_i - x_{i-1} = \Delta x$ and $w_i = w$ for $i = 1, \ldots, n$. Note that we will have a special case of expression (46) at the boundaries,

$$\langle f_1, f_1 \rangle = \langle f_n, f_n \rangle = \frac{1}{3}\, w^2 \Delta x. \tag{48}$$

We observe that the matrix F in expression (8) now will be tridiagonal, for example

$$F = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} & \frac{1}{3} \end{bmatrix} \tag{49}$$

for $n = 4$. As a result, $H = F^{-1}$ will in general be a full matrix and the calculation of

$$\theta = F^{-1} \langle m, f \rangle \tag{50}$$

can be rather resource demanding for large n. From expression (10) we see that each θ_j thus includes contributions from possibly distant elements of $\hat{m} = \langle m, f \rangle$. However, the results from Section 2 still apply, and expression (50) defines the distribution of the vector θ as a linear transformation of the distribution of $\hat{m}$.

4.3 Wavelets

In this section we will use wavelet bases to perform the reparametrization (1). The starting point is discretization of the continuous wavelet transform through multiresolution analysis. For general references on wavelets, see Daubechies (1992) and Vidakovic (1999).

4.3.1 Wavelet bases

Assume some scaling function $\phi(x)$ for $x \in \mathbb{R}^1$. The functions

$$\phi_{ij}(x) = 2^{j/2} \phi(2^j x - i), \tag{51}$$

for $j = 0 \ldots n$ and $i = 0 \ldots 2^j - 1$ are orthonormal, and the vector space

$$V^j = \text{span}\{\phi_{ij}(x)\}_{i=0}^{2^j - 1}. \tag{52}$$

will span the space of any function, through a basis expansion of the form (2) as $j \to \infty$. From the scaling function $\phi(x)$ one can derive a unique wavelet function $\psi(x)$, often called the mother wavelet, see Vidakovic (1999). From $\psi(x)$ we define the wavelet functions

$$\psi_{ij} = 2^{j/2} \psi(2^j x - i) \tag{53}$$

for $j = 0 \ldots n$ and $i = 0 \ldots 2^j - 1$. The wavelet functions (53) are orthonormal and we define the vector space

$$W^j = \text{span}\{\psi_{ij}(x)\}_{i=0}^{2^j-1} \tag{54}$$

for $j = 1, \ldots, n$. For wavelets, the property

$$V^{j+1} = V^j \oplus W^j \tag{55}$$

holds, with $\oplus$ being the outer product. Thus

$$V^n = V^0 \oplus W^0 \oplus W^1 \oplus \ldots \oplus W^{n-1}, \tag{56}$$

allowing for controlled and increased resolution when expanding a function using the basis formed by the wavelet functions. This can be written as the basis expansion

$$m_\psi(x) = \theta_{00}^\phi \phi_{00}(x) + \sum_{j=0}^{n} \sum_{i=0}^{2^j-1} \theta_{ji} \psi_{ij}(x) \tag{57}$$

with θ_{ji} playing the same role as θ_i in expression (2), but with double subscript for differentiating between location and scale, and the function $\phi_{00}(x)$ from expression (51) defining the overall average of the decomposed function around which the wavelet functions (53) add increasing levels of detail. Performing a similar analysis as in Section 2 on the expansion (57) yields

$$\theta_{pq} = \langle m, \psi_{pq} \rangle \tag{58}$$

$$\theta_{00}^\phi = \langle m, \phi_{00} \rangle \tag{59}$$

for $p = 0, \ldots, n$ and $q = 0, \ldots, 2^p - 1$ due to the orthonormality of the wavelet functions and the scaling function, see expression (11) for reference. Expressions (58) and (59) state that the wavelet coefficients are smoothed versions of the random field $m(x)$ with respect to the wavelet functions $\psi_{ij}(x)$ and the scaling function $\phi_{00}(x)$.

Example: The Haar wavelet basis
The Haar scaling function is the box-car function

$$\phi(x) = \begin{cases} 1 & , \quad x \in [0, 1) \\ 0 & , \quad x \notin [0, 1) \end{cases} , \tag{60}$$

see Figure 3, with the corresponding Haar wavelet function being

$$\psi(x) = \begin{cases} 1 & , \quad x \in [0, \frac{1}{2}) \\ -1 & , \quad x \in [\frac{1}{2}, 1) \\ 0 & , \quad x \notin [0, 1) \end{cases} . \tag{61}$$

Expansion through the Haar wavelet functions is equivalent to an approximation by step functions whose values are the averages of the function over appropriate dyadic intervals. Figure 5 shows the Haar approximation of the function

$$g(x) = \exp(-x^2) + 0.7x \cdot \mathcal{I}(0.5 \leq x \leq 1) \ , \ x \in [0,1] \tag{62}$$

with resolution $n = 2$. The function $\mathcal{I}(A)$ is the indicator function for the event A. ∎

Returning to the general wavelet formulation, one has from expressions (12) and (13) that

$$\mathrm{E}\{\theta_{ij}\} = \int \mu(x)\psi_{ij}(x)\mathrm{d}x \tag{63}$$

$$\mathrm{Cov}\{\theta_{ij}, \theta_{pq}\} = \int\int \psi_{ij}(x)C(x,y)\psi_{pq}(y)\mathrm{d}x\mathrm{d}y \tag{64}$$

and similarly for θ_{00}^{ϕ}. If $m(x)$ is a Gaussian random field, θ will be multivariate Gaussian distributed, see expression (15).

4.3.2 Biorthogonal wavelets

The Haar wavelet is the only known wavelet that is compactly supported, orthogonal and symmetric. In order to construct more families of compactly supported, symmetric wavelets, biorthogonal wavelets have been introduced. For a more detailed outline on biorthogonal wavelets, see Cohen *et al.* (1992), Vidakovic (1999) and Walter and Shen (2001).

Assume that we have a wavelet expansion (57), but that the basis $\{\psi_{jk}\}$ is not orthogonal, and the spaces V^j and W^j are not orthogonal. We introduce another set of basis functions $\{\tilde{\psi}_{jk}\}$ called the dual of $\{\psi_{jk}\}$, having the properties

$$\langle\tilde{\phi}_{ij}, \phi_{ik}\rangle = \delta_{jk} \tag{65}$$

$$\langle\tilde{\psi}_{ij}, \psi_{i'k}\rangle = \delta_{ii'}\delta_{jk} \tag{66}$$

$$\langle\tilde{\psi}_{ij}, \phi_{ik}\rangle = 0 \tag{67}$$

$$\langle\tilde{\phi}_{ij}, \psi_{ik}\rangle = 0 \tag{68}$$

instead of the orthogonality conditions of subsection 4.3.1. Here δ_{pq} is the Kronecker delta function. We assume that the function space and its dual space are the same, thus any function can be represented in a biorthogonal basis as

$$m_{\psi}(x) = \sum_{ij}\langle m, \tilde{\psi}_{ij}\rangle\psi_{ij}(x) = \sum_{ij}\langle m, \psi_{ij}\rangle\tilde{\psi}_{ij}(x) \tag{69}$$

For details on construction of biorthogonal wavelets, see Daubechies (1992). Following the outline of subsection 4.3.1 with the basis $\{\psi_{ij}\}$ and the wavelet expansion (57), we now

have $\theta_{ij} = \langle m, \tilde{\psi}_{ij} \rangle$ and

$$\mathrm{E}\{\theta_{ij}\} = \int \mu(x)\tilde{\psi}_{ij}(x)\mathrm{d}x \tag{70}$$

$$\mathrm{Cov}\{\theta_{ij}, \theta_{pq}\} = \int\int \tilde{\psi}_{ij}(x)C(x,y)\tilde{\psi}_{pq}(y)\mathrm{d}x\mathrm{d}y. \tag{71}$$

If $m(x)$ is a Gaussian random field, then

$$\theta \sim N(\mu_\theta, C_\theta) \tag{72}$$

with the mean and variance μ_θ and C_θ to be calculated from expressions (70) and (71) respectively.

5 Summary

We have shown how to discretize a continuous random field onto a discrete set of stochastic parameters given a set of basis functions. This is of importance in for example seismic inversion and the definition of prior distributions, as well as discretization of continuously available posterior distributions. The discretization is valid for any set of basis functions, as well as any statistical distribution. The Mean Integrated Square Error (MISE) can be used as a criterion for evaluting different choices of sets of basis fuctions.

References

[1] Abrahamsen, P. (1994) *A Review of Gaussian Random Fields and Correlation Fuctions*, Report 878, Norsk Regnesentral

[2] Adler, R.J. (1981) *The Geometry of Random Fields*, John Wiley & Sons, New York

[3] Boyd, D.W. and Steele, J.M. (1978) Lower Bounds for Nonparametric Density Estimation Rates, *The Annals of Statistics*, Vol.6,No.4,932-934

[4] Buland, A. and Omre, H. (2003) Bayesian linearized AVO inversion, *Geophysics*, **68**, 185-198

[5] Christakos, G. (1992) *Random Field Models in Earth Sciences*, Academic Press Inc., San Diego

[6] Cohen, A., Daubechies, I., and Feauveau, J. C. (1992) Biorthogonal Bases of Compactly Supported Wavelets, *Communications on Pure and Applied Mathematics*, **XLIV**, pp. 485–560.

[7] Cramer, H. and Leadbetter, R. (1967) *Stationary and Related Stochastic Processes*, Wiley, N.Y

[8] Daubechies, I. (1992) *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia

[9] Gradshteyn, I. S., and Ryzhik, I. M. (1994) *Table of Integrals, Series, and Products*, Fifth edition, Alan Jeffrey (ed.), Academic Press.

[10] Janson, S. (1997), *Gaussian Hilbert Spaces*, Cambridge University Press.

[11] Kotz, Balakrishnan, Johnson (2000), *Continuous Multivariate Distributions*, Volume 1, John Wiley & Sons, Inc.

[12] Lifshits, M.A. (1995), *Gaussian Random Functions*, Kluwer Academic Press.

[13] Malinverno, A. (2002) Parsimonious Bayesian Markov Chain Monte Carlo Inversion in a Nonlinear Geophysical Problem, *Geophys. J. Int* **151**, 675-688

[14] Vidakovic, B. (1999) *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc.

[15] Walter, G.G. and Shen, X. (2001) *Wavelets and Other Orthogonal Systems*, Chapman & Hall/CRC

Appendix

A Stochastic integration

A.1 General results for $\mathbb{R}^1$

Assume a zero mean random variable $m(x)$ with covariance function $C(x, y)$. Further, assume the random variable defined by the Riemann integral

$$J = \int_a^b g(x)m(x)\mathrm{d}x \tag{73}$$

with $I = [a, b]$ some finite or infinite interval, and $g(x)$ a non-random function. If $C(x, y)$ is continuous in I^2, and $g(x)$ is such that

$$Q = \int_a^b \int_a^b g(x)g(y)C(x, y)\mathrm{d}x\mathrm{d}y < \infty \tag{74}$$

then the integral J exists as the quadratic mean limit of the associated Riemann sum, and

$$\begin{aligned} \mathrm{E}\{J\} &= 0 \quad (75) \\ \mathrm{E}\left\{|J|^2\right\} &= Q. \quad (76) \end{aligned}$$

Stochastic integrals of the form (73) possess ordinary formal properties of integrals. For more on stochastic integration, see Cramér and Leadbetter (1967), Adler (1981) and Christakos (1992).

A.2 Gaussian

Assume that $\{m(x); x \in \mathbb{R}^q\}$ is a continuous Gaussian random field with mean $\mu(x)$ and covariance function $C(x, y)$. That is, the random field might be non-stationary both in mean and variance. Generally an averaging Gaussian random field is defined by the Riemann integral

$$Z(x) = \int_V w(x, y)m(y)\mathrm{d}y \quad (77)$$

with $V \subset \mathbb{R}^q$, and $w(x, y)$ piecewise continuous and bounded on $\mathbb{R}^q \otimes \mathbb{R}^q$, and is again a Gaussian random field. Further

$$\begin{aligned} \mathrm{E}\{Z(x)\} &= \int_V w(x, y)\mu(y)\mathrm{d}y \quad (78) \\ \mathrm{Cov}\{Z(x), Z(y)\} &= \int_V \int_V w(x, u)C(u, v)w(y, v)\mathrm{d}u\mathrm{d}v \quad (79) \end{aligned}$$

The weighting, or smoothing, function will typically be a function of some distance measure, $w(x, y) = w(x - y)$. For reference, see Abrahamsen (1994). The fact that an averaged Gaussian random field is again a Gaussian random field can be stated more formally, see for example Janson (1997).

B Distribution of a linear transformation of a continuous Gaussian random field

Assume a Gaussian random field $m(x)$ with mean $\mu(x)$ and covariance function $C(x, y)$. We seek the distribution of

$$m_\Delta(x) = m(x) - f(x)^T H \int m(u)f(u)\mathrm{d}u \quad (80)$$

with $f(x)$ and H defined in Section 2. As $m_\Delta(x)$ is only a linear transformation of the GRF $m(x)$, $m_\Delta(x)$ will be a GRF as well, with mean

$$\mu_\Delta(x) = \mu(x) - f(x)^T H \int \mu(u) f(u) \mathrm{d}u \tag{81}$$

and covariance function

$$\begin{aligned} C_\Delta(x,y) &= \mathrm{E}\left\{[m_\Delta(x) - \mu_\Delta(x)][m_\Delta(y) - \mu_\Delta(y)]^T\right\} && (82)\\ &= \mathrm{E}\Bigg\{[m(x) - \mu(x) - f(x)^T H \int (m(u) - \mu(u)) f(u)\mathrm{d}u] && (83)\\ &\quad \cdot [m(y) - \mu(y) - f(y)^T H \int (m(v) - \mu(v)) f(v)\mathrm{d}v]^T\Bigg\} && (84) \end{aligned}$$

For the quadratic terms we have

$$\mathrm{E}\left\{[m(x) - \mu(x)][m(y) - \mu(y)]^T\right\} = C(x,y) \tag{85}$$

and

$$\begin{aligned} &\mathrm{E}\left\{\left[f(x)^T H \int (m(u) - \mu(u)) f(u)\mathrm{d}u\right]\left[f(y)^T H \int (m(v) - \mu(v)) f(v)\mathrm{d}v\right]^T\right\} && (86)\\ = \; & f(x)^T H \int\int f(u) C(u,v) f(v)^T \mathrm{d}u\mathrm{d}v H^T f(y) && (87) \end{aligned}$$

respectively. The crossterms are

$$\begin{aligned} &\mathrm{E}\left\{[m(x) - \mu(x)]\left[f(y)^T H \int (m(v) - \mu(v)) f(v)\mathrm{d}v\right]^T\right\} && (88)\\ = \; & \mathrm{E}\left\{\int (m(x) - \mu(x))(m(v) - \mu(v)) f(v)^T \mathrm{d}v H^T f(y)\right\} && (89)\\ = \; & \int C(x,v) f(v)^T \mathrm{d}v H^T f(y) && (90) \end{aligned}$$

and

$$\begin{aligned} &\mathrm{E}\left\{\left[f(x)^T H \int (m(u) - \mu(u)) f(u)\mathrm{d}u\right][m(y) - \mu(y)]^T\right\} && (91)\\ = \; & \mathrm{E}\left\{f(x)^T H \int (m(y) - \mu(y))(m(u) - \mu(u)) f(u)\mathrm{d}u\right\} && (92)\\ = \; & f(x)^T H \int C(y,u) f(u)\mathrm{d}u && (93) \end{aligned}$$

respectively. Thus $m_\Delta(x)$ is a Gaussian random field having mean (81) and covariance function

$$\begin{aligned} C_\Delta(x,y) &= C(x,y) && (94)\\ &\quad - \int C(x,v) f(v)^T \mathrm{d}v H^T f(y) && (95)\\ &\quad - f(x)^T H \int C(y,u) f(u) \mathrm{d}u && (96)\\ &\quad + f(x)^T H \int\int f(u) C(u,v) f(v)^T \mathrm{d}u \mathrm{d}v H^T f(y) && (97) \end{aligned}$$

C Exponential covariance function in a layered medium

Assume a Gaussian random field described as in Section 4.1.

C.1 Integration of covariance function

Restating the results for θ, using the weights (43), $w_j = 1$, we have

$$\theta_j = \frac{1}{\Delta x_j} \int_{x_{j-1}}^{x_j} m(x) \mathrm{d}x \tag{98}$$

with

$$\begin{aligned} \mathrm{E}\{\theta_j\} &= \frac{1}{\Delta x_j} \int_{x_{j-1}}^{x_j} \mu(x) \mathrm{d}x && (99)\\ \mathrm{Cov}\,\{\theta_i, \theta_j\} &= \frac{1}{\Delta x_i \Delta x_j} \int_{x_{i-1}}^{x_i} \int_{x_{j-1}}^{x_j} C(u,v) \mathrm{d}u \mathrm{d}v. && (100) \end{aligned}$$

for $i, j = 1, \ldots, n$. Assume a stationary covariance function, for example the exponential,

$$C(u,v) = C(|u-v|) = \sigma^2 \exp\left(-\frac{|u-v|}{\lambda}\right) \tag{101}$$

where λ is a correlation length. For more on correlation functions, see Abrahamsen (1994). Figures 6 and 7 show two different possibilites when performing the integrations.

In Figure 6 we have $u > v$, that is $i > j$ and $x_{i-1} \geq x_j$, thus being the integration for the covariances. Figure 7 shows what happens when $i = j$, that is, the variance. Performing

the integrations yields

$$\begin{aligned}
\mathrm{Cov}\left\{\theta_i, \theta_j\right\} &= \frac{\sigma^2}{\Delta x_i \Delta x_j} \int_{x_{j-1}}^{x_j} \int_{x_{i-1}}^{x_i} \exp\left(-\frac{|u-v|}{\lambda}\right) \mathrm{d}u\mathrm{d}v && (102)\\
&= \frac{(\sigma\lambda)^2}{\Delta x_i \Delta x_j} \left[\exp\left(-\frac{x_i - x_{j-1}}{\lambda}\right) - \exp\left(-\frac{x_i - x_j}{\lambda}\right)\right. && (103)\\
&\quad \left. + \exp\left(-\frac{x_{i-1} - x_j}{\lambda}\right) - \exp\left(-\frac{x_{i-1} - x_{j-1}}{\lambda}\right)\right] && (104)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left\{\theta_i\right\} &= \frac{2\sigma^2}{(\Delta x_i)^2} \int_{x_{i-1}}^{x_i} \int_{v}^{x_i} \exp\left(-\frac{|u-v|}{\lambda}\right) \mathrm{d}u\mathrm{d}v && (105)\\
&= \frac{2(\sigma\lambda)^2}{(\Delta x_i)^2} \left[\exp\left(-\frac{x_i - x_{i-1}}{\lambda}\right) - 1 + \frac{x_i - x_{i-1}}{\lambda}\right] && (106)
\end{aligned}$$

C.2 Simplifications and limits

Assume that all layers have the same thickness $x_i - x_{i-1} = x_j - x_{j-1} = \Delta x$ for $i, j = 1, \ldots, n$. Starting from $x_0 = 0$ we then have $x_j = j\Delta x$. Further, set $\gamma = \frac{\Delta x}{\lambda}$. We obtain, due to symmetry,

$$\mathrm{Cov}\left\{\theta_i, \theta_j\right\} = \sigma^2 \cdot \frac{1}{\gamma^2} \frac{[\exp(-\gamma) - 1]^2}{\exp(-\gamma)} \exp\left(-\gamma|i-j|\right). \quad (107)$$

That is, the correlation function for the parameters is exponential as well, but with an adjusted scaling parameter depending on γ. For the variance we have

$$\mathrm{Var}\left\{\theta_i\right\} = 2\sigma^2 \frac{1}{\gamma^2} \left[\exp\left(-\gamma\right) + \gamma - 1\right] \quad (108)$$

We observe that the variance is not found directly from setting $i = j$ in (107).

Assume $\Delta x \gg \lambda$. That is, we approach a white noise phenomena, as the correlation length is too short to have any impact. With γ large we have

$$\mathrm{Cov}\left\{\theta_i, \theta_j\right\} \approx 0 \quad (109)$$

directly from (107). For the variance we find from (108)

$$\begin{aligned}
\mathrm{Var}\left\{\theta_i\right\} &\approx 2\sigma^2 \frac{1}{\gamma} && (110)\\
&= \sigma^2 \frac{2\lambda}{\Delta x} && (111)
\end{aligned}$$

As Δx becomes larger, $\mathrm{Var}\{\theta_i\}$ becomes smaller because we are averaging over greater length.

Assume then $\lambda \gg \Delta x$. The high correlation length will now make the individual layer thicknesses un-noticeable. Then $\gamma \ll 1$, and

$$\lim_{\gamma \to 0} \mathrm{Cov}\{\theta_i, \theta_j\} = \sigma^2 \exp\left(-\gamma |i-j|\right) \tag{112}$$

as

$$\lim_{\gamma \to 0} \left[\frac{1}{\gamma^2} \cdot \frac{\left(1-\exp(-\gamma)\right)^2}{\exp(-\gamma)} \right] = 1, \tag{113}$$

see Gradshteyn and Ryzhik (1994), and

$$\lim_{\gamma \to 0} \mathrm{Var}\{\theta_i\} = \sigma^2 \tag{114}$$

as

$$\lim_{\gamma \to 0} \left[\frac{2}{\gamma^2} [\exp(-\gamma) + \gamma - 1] \right] = 1, \tag{115}$$

see Gradshteyn and Ryzhik (1994). So, for the parameter θ we have a situation similar to that for the original field $m(x)$: The variance is σ^2, and the correlation function is exponential with the same parameter as the continuous field.

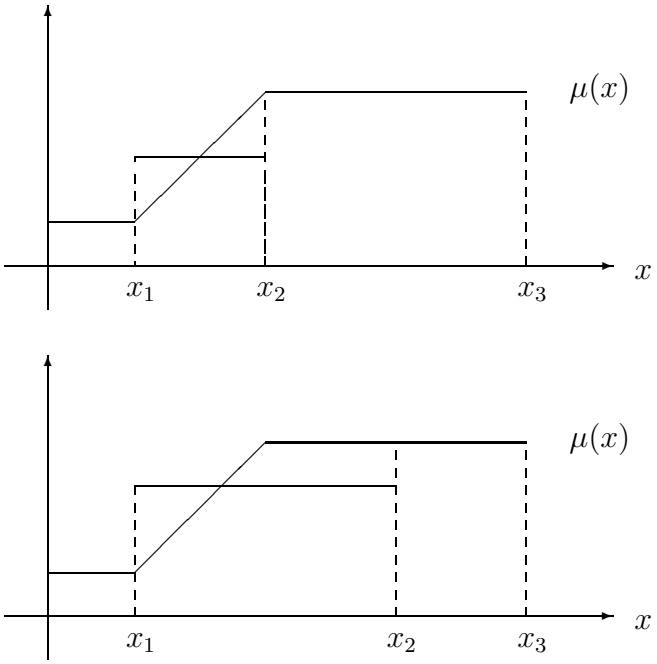

Figure 1: *Two different layered models with $n = 3$ layers for fitting a prior random field $m(x)$ with mean $\mu(x)$.*

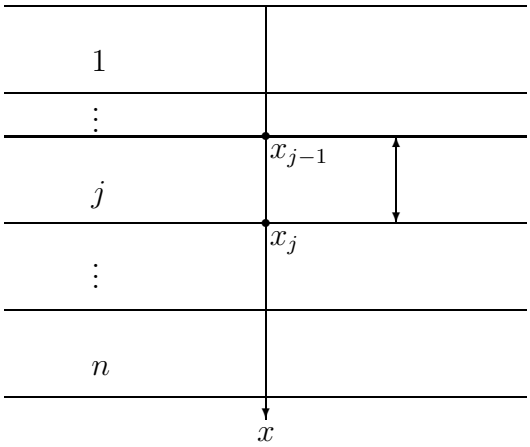

Figure 2: *Layered earth model with n homogenous layers.*

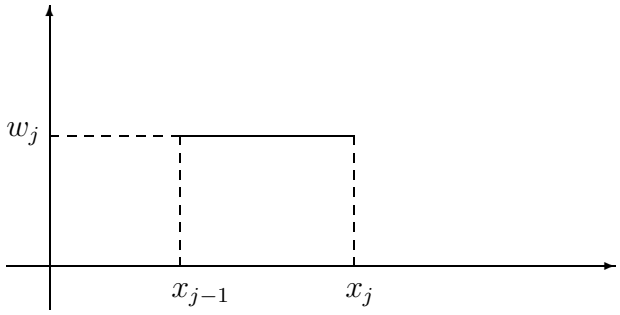

Figure 3: *Boxcar basis function.*

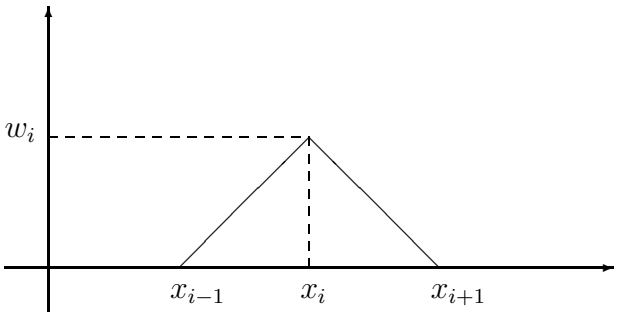

Figure 4: *Linear interpolation basis function.*

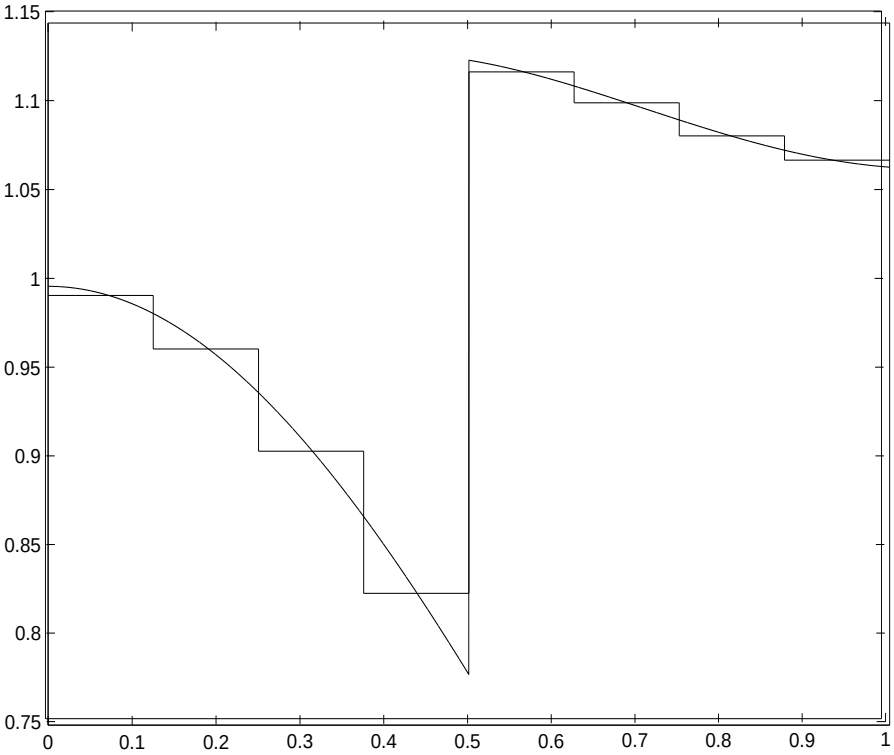

Figure 5: *Haar approximation.*

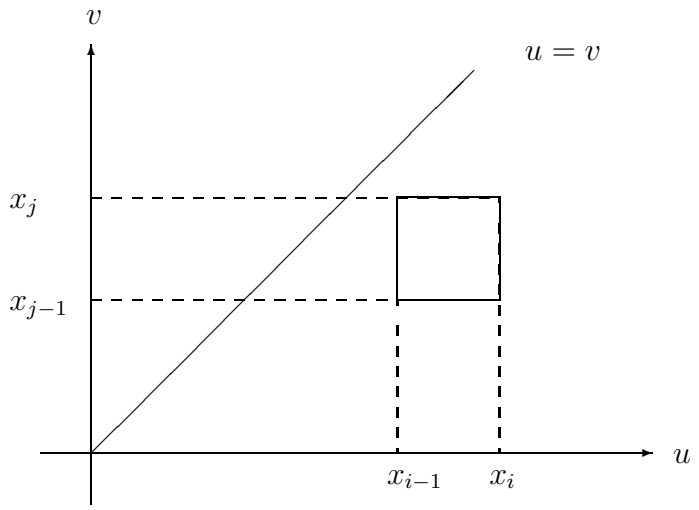

Figure 6: *The second quadrant, and $u > v$, implying $i > j$.*

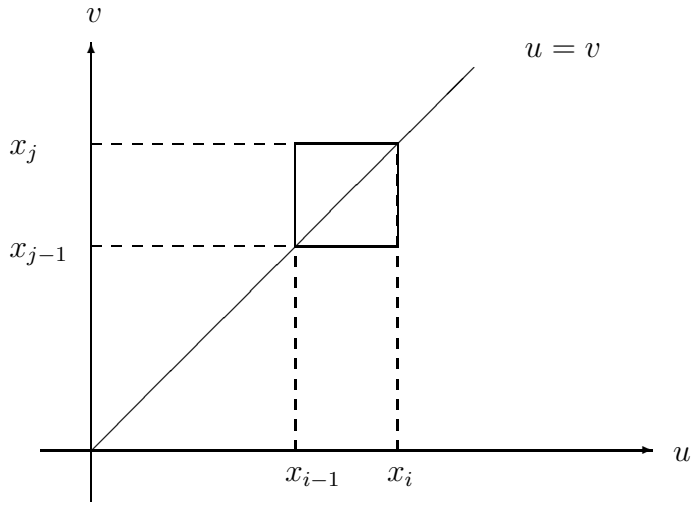

Figure 7: *The second quadrant, and $i = j$.*

Prediction of oil production conditioned on the breakthrough time

Jo Røislien
Department of Petroleum Engineering and Applied Geophysics
Norwegian University of Science and Technology

Peter King
Department of Earth Science and Engineering
Imperial College, London

Sergey V. Buldyrev, Shlomo Havlin
Eduardo Lopez, H.Eugene Stanley
Center for Polymer Studies and Department of Physics
Boston University

Abstract

The prediction of future oil production is very difficult, as hydrocarbon reservoirs are extremely complex structures. Modern reservoir models typically take hours of cpu time to run. In this work it is shown how knowledge of the breakthrough time reduces the uncertainty in future predictions, by analyzing production data from a simulated reservoir cast in the framework of percolation. The two first empirical moments are estimated, and then fitted to a probability density function. There are clear trends in the functional behaviour of the estimated empirical moments in these data, indicating the existence of more general results for the moments, either as analytical functions, or as general curves. As only one percolation system has been analyzed, more work should be done along these lines.

URN:NBN:no-3476

1 Introduction

Hydrocarbon reservoirs are extremely complex geological structures with heterogeneities on all scales, from microns (cores) to tens of kilometres (basins). Analysing the flow of oil, water and gas through the reservoir is very difficult, and yet crucial for development of effective production strategies. In particular, it is important to be able to predict the time it takes for an injected fluid, usually water, to travel from an injection well to a production well. This displacement process is the most common way that oil is produced worldwide. It is also important to be able to predict how rapidly the oil production rate declines with time after the breakthrough has occurred.

There are two important issues that limit our ability to make these predictions. One is that we cannot resolve all the details required in numerical flow simulations. The ratio of volumes of an oil reservoir to pore scale volumes is around 10^{20}, and it is clearly impractical to try to model at this scale. However, even if one considers the scale of core plugs, which is where direct measurements of flow properties are made, the ratio of volumes is of the order of 10^{13}. So, conventionally we have to use a range of upscaled or effective flow properties. Even so, modern reservoir models are very complex, and can take many hours of cpu time to run. It is desireable to be able to reduce this computational time as much as possible.

The other issue is that we really do not have this level of detailed knowledge of where all the reservoir heterogeneities are, so we need to consider a number of possible alternative models in order to estimate the uncertainty arising from this lack of knowledge. As such, we would expect that models with little input or conditioning data would have a larger uncertainty. The key contribution of this paper is to show how knowledge of the breakthrough time reduces the uncertainty in forward predictions.

In this paper we addresss these issues using percolation theory. We shall assume that the rocks in the reservoir can be ascribed to one of two types; either permeable (sand) or impermeable (shale). This is typical of a number of systems, such as fluviale or turbidites. The approach could also be used for fractured systems where the fracture network is the main flow path. For the sand-shale system, the volume fraction of sand is then the net to gross ratio. This corresponds to the probability that a random point in space is lying within the permeable rock. As such, it also corresponds with the percolation occupancy probability p in a continuum percolation model.

The assymptotic mean behaviour of unconditional oil production in an infinite percolation system is discusssed in King et al. (1999a). The probability distribution of the breakthrough itself is analyzed in King et al. (1999b) and King et al. (2002). This paper is also motivated by some of the work shown in Buldyrev (2002).

The paper is outlined as follows; in Section 2 the problem is described in more detail, and a brief analysis a set of production data generated from a given percolation system is

performed; in Section 3 the behaviour of the two first empirical moments for the data set is discussed; in Section 4 these results are put into a more formal statistical framework; in Section 5 we investigate the correlation between the breakthrough time and the half time, the time when the production is half of the initial production; finally, Section 6 contains a summary.

2 Problem Description and Data Analysis

Assume an oil reservoir penetrated by two wells. One oil producing well A, and one well B where some fluid, most often water, is being injected. In the initial fases after the injection of water, only oil will be produced at A. However, as time passes, sooner or later water will find its way from B to A, resulting in the net oil production at A dropping, until only water is being produced. As water flows more easily than oil, it is clear that this drop in oil production will be rapid. We describe this reservoir in the framework of percolation, see Stauffer and Aharony (1994) for an introduction to percolation. For a review of oil reservoirs described by percolation, see King (1990). We let L denote the size of the reservoir, l the shortest path between wells A and B, r the well spacing and p the probability of permeable rock in an arbitrary location.

In Figure 1 production data generated from integrating travelling time distributions of a percolation system with $L/r = 11$, $L = 514$ bonds and $r = 46$, in total 141 production curves, is displayed. The time has not been transformed to correspond with real oil fields. Since the travelling time distributions are smooth, the resulting production data curves are smooth as well. Actually, parts of these data are average curves, implying that the following analysis should be made on rougher production data at some later stage. Also, let $C \in [0, 1]$ denote oil production in fractions of only oil being produced.

First, we note that as the breakthrough time increases, the production curves become steeper. This seems natural. With early breakthrough, a lot of the system is yet to be explored, which then will happen gradually. With late breakthrough, most of the system is already explored, and all possible paths in the percolation system leading from from B to A will then reach the producing well A within a very short time frame after breakthrough.

Figure 2 shows production curves conditioned on breakthrough time $t_{br} \in [50 - \delta, 50 + \delta]$ with $\delta = 3$, a total of 18 traces. A similar δ-adjustment has been used for all conditioning on breakthrough times throughout the paper. As a result, the variances will be slightly overestimated.

In Figure 3 kernel density estimates of the production in Figure 1 at times $[0, 9, 18, \ldots, 135]$ are shown. As expected, at early times the probability mass is located close to production $C = 1$, and as time evolves, the mass slowly shifts towards $C = 0$. In Figure 4 kernel

density estimates for the production data conditioned on $t_{br} = 50$ in Figure 2 at times $[50, 54, \ldots, 110]$ are shown. We observe the same general behaviour as for the unconditional production data: Up until breakthrough the probability mass is located at $C = 1$, then shifts towards $C = 0$ as time passes. However, note that for the conditional data, the transition from $C = 1$ to $C = 0$ is far more defined, and it also takes place within a much shorter time frame than for the unconditional production data. This is as expected.

3 Empirical Moments

In order to describe the behaviour of the conditional production data in Figure 2, we calculate the empirical mean $\hat{\mu}$ and variance $\hat{\sigma}^2$ for different times conditioned on some breakthrough time;

$$\hat{\mu}(t)|t_{br} = \bar{d}(t)|t_{br} = \frac{1}{n_t}\sum_i d_i(t)|t_{br} \tag{1}$$

$$\hat{\sigma}^2(t)|t_{br} = \frac{1}{n_t - 1}\sum_i \left[d_i(t)|t_{br} - \bar{d}(t)|t_{br}\right]^2 \tag{2}$$

with $d(t)|t_{br} = [d_1(t)|t_{br}, \ldots, d_{n_t}(t)|t_{br}]^T$ being the n_t data points at time t given the breakthrough time t_{br}. We do this for breakthrough times $t_{br} = [40, 50, \ldots, 90]$, resulting in Figure 5; the conditional empirical means are shown in the upper display, and the conditional empirical variances in the lower display.

3.1 Empirical mean

The empirical means $\hat{\mu}(t)|t_{br}$ for different breakthrough times are very similar in their functional behaviour. After breakthrough, they all drop from 1 towards 0 in a similar manner, with a slightly steeper drop as the breakthrough time increases. In Figure 6 three different transformations of the behaviour of $\hat{\mu}(t)|t_{br}$ is shown. The upper left display shows the original empirical means, whereas the upper right display is a log-log plot of these. The lower left and lower right displays show the log of the empirical means plotted against $\log\left(\frac{t}{t_{br}}\right)$ and $\log(t - t_{br})$, respectively. We observe that the shift transform collapses the data nicely. The scaling transform $\frac{t}{t_{br}}$ is not as successful as the shift transform, but one should not exclude transforms like $\frac{t}{t_{br}^{\alpha}}$ with $\alpha \neq 1$. Also, a combination of scaling and shift might turn out to be successful in collapsing the data, as we have observed that the empirical means become slightly steeper as the breakthrough time increases. Other sets of data and other percolation systems should be analyzed, in order to look for a general transform.

3.2 Empirical variance

The empirical variances $\hat{\sigma}^2(t)|t_{br}$ for different breakthrough times also have similar shapes. After breakthrough, they increase from 0, reach a maximum value, before decreasing towards 0 as time evolves. In Figure 7 the same three transforms as for the empirical means are shown; the upper left and right displays show the empirical variances and a log-log plot of these, respectively, whereas the lower left and right displays show the log of the empirical variance plotted against $\log\left(\frac{t}{t_{br}}\right)$ and $\log(t - t_{br})$, respectively. The empirical variances are not as easily collapsed as the empirical means, but the scaling transform almost does the job. As with the mean, the correct scaling might be with respect to t_{br}^{β}, for $\beta \neq 1$. Again, other sets of data and other percolation systems should be analyzed, in order to look for a general transform.

3.3 Prediction of oil production

The collapsions of $\hat{\mu}(t)|t_{br}$ and $\hat{\sigma}^2(t)|t_{br}$, shown in Figures 6 and 7 respectively, indicate that there might exist some universal curves, or general functions, describing the behaviour of the empirical mean and variance of conditional production data. These functions might or might not have simple analytically closed forms. But, knowing the analytical expressions explicitely is not essential. Having the exact universal curves allows for numerical inversion from given breakthrough time to first and second order empirical moments, allowing for controlled prediction of future production, both the mean and the attached variance.

3.4 Empirical moments of unconditional production data

So far in this section, we have focused on the conditional production data. However, we can, of course, calculate the two first empirical moments of the original production data as well. Figure 8 shows the behaviour of the empirical mean and variance for the unconditional production data, and for the production data conditioned on $t_{br} = 50$. We observe that the two empirical moments for the conditional and unconditional data posess the same behaviour as for the conditional data, but that the mean for the conditional data has a more defined drop point and decreases more rapidly towards 0 than the empirical mean for the unconditional production data. As for the variance, it is dramatically reduced when conditioning, both in width and overall level. In conclusion, knowing the breakthrough time clearly increases the precision in the prediction of future oil production.

4 Probability Density Functions

In the previous section, a way of predicting future oil production given the breakthrough time by estimation of the two first empirical moments was outlined. This conditional prediction was also compared to the unconditional prediction. In this section we introduce the empirical moments into a probabilistic framework.

4.1 Fitting a probability density function

Having the emprical mean and variance for some data, one can fit a statistical distribution having these values as their two first moments. In our setting, this implies analytically approximating the kernel density estimates in Figures 3 and 4.

For the conditional production data, we observe a rather well defined, symmetric bell shape travelling from $C = 1$ to $C = 0$ in Figure 4. One of the distributions having this symmetric bell shape is the well-known Gaussian distribution. We may therefore approximate the kernel density estimates in Figure 4 with a truncated Gaussian probability density function (pdf). Truncated Gaussian approximations for the production data conditioned on $t_{br} = 50$, using the associated empirical mean and variance discussed in Section 3, is shown in Figure 9. The Gaussian estimate provides a fairly good fit.

The Gaussian pdf may seem to underestimate the tail behaviour. However, it is not evident whether this is due to the actual behaviour in the data, the smoothing from the kernel density estimate, or just the sparse amount of conditional data available. If the data actually are more heavy-tailed than the Gaussian distribution, one could try fitting a more heavy-tailed, symmetric, bell-shaped distribution, see Evans (2000) for more statistical distributions. However, the Gaussian distribution has the advantage of being analytically tractable, in addition to the fact that it appears to fit the conditional production data reasonably well.

For the unconditional data, Figure 3, the aforementioned peak-travelling behaviour is not similarly apparent. However, we can still fit a Gaussian distribution to the unconditional production data. The result is shown in Figure 10. As the level and width of the empirical variance for the unconditional production data is much larger than for the conditional production data, the result is a much wider Gaussian pdf approximation.

4.2 Comparing conditional and unconditional probability density functions

Let $f(\cdot)$ denote a pdf and $f(\cdot|\cdot)$ a conditional pdf. The problem described in Section 2 can now be restated as finding $f\left(C_t|t_{br}, r\right)$ and $f\left(C_t|r\right)$, and comparing the two. Here C_t denotes oil production at time t. Note that the reservoir parameters $\rho = \{l, L, p\}$ are absent, as we want the pdfs to be independent of the percolation system. We assume that r is deterministically known. From probability theory, we have

$$f\left(C_t|t_{br}, r\right) = \int f\left(C_t|\rho, t_{br}, r\right) f(\rho|t_{br}, r)\mathrm{d}\rho. \tag{3}$$

Bayes' law yields

$$f(\rho|t_{br}, r) = \frac{f(t_{br}|\rho, r)f(\rho|r)}{f(t_{br}|r)}, \tag{4}$$

where

$$f(t_{br}|r) = \int f(t_{br}|\rho, r)f(\rho|r)\mathrm{d}\rho \tag{5}$$

is merely a scaling constant. The pdf for the breakthrough time conditioned on the system, $f(t_{br}|\rho, r)$, is discussed in King et al. (1999b), whereas the pdf for the system parameters, $f(\rho|r)$, is discussed in Nikolay et al. (1999). Thus, having some functional form for $f(C_t|\rho, t_{br}, r)$, for example the Gaussian pdf, as was discussed in the previous subsection, the pdf in expression (3) may be found. Not having explicit functional forms for the two first moments in $f\left(C_t|\rho, t_{br}, r\right)$, but universal curves, the integration has to be performed numerically. As for the pdf for the unconditional production, we can perform

$$f\left(C_t|r\right) = \int f\left(C_t|t_{br}, r\right) f(t_{br}|r)\mathrm{d}t_{br}, \tag{6}$$

again by numerical integration.

Neither of the pdfs (3) nor (6) will have simple, known closed analytical forms. Meaning that in order to draw samples from these distribution to visualize the uncertainty in the pdf, one will have to rely on stochastic simulation techniques, see Ripley (1994) for more on this topic.

5 Prediction of the Half-Time

The conditional production data curves, whether conditioned on an early or late breakthrough time, are very similar. Once water has broken through in the producing well,

the production drops rapidly and with similar shape, almost unaffected of what t_{br} is. The empirical means and variances are also very similar for different breakthrough times. This indicates that the half time $t_{1/2}$, or any time $t > t_{br}$, is highly correlated with the breakthrough time. For our data, the correlation between the t_{br} and $t_{1/2}$ is as high as 0.9747.

Figure 11 displays kernel density estimates of t_{br} and $t_{1/2}$. We see that the density estimates are almost identical, only with the density estimate for $t_{1/2}$ slightly scaled and shifted compared to the density estimate for t_{br}. The log-log plot of $t_{1/2}$ against t_{br} in Figure 12 indicates a clear linear trend. Stated as a probabilistic linear regression problem, we have

$$(\log t_{1/2})_i = a + b(\log t_{br})_i + \epsilon_i \ ; \ i = 1, \ldots, n \tag{7}$$

with a and b constants, $\epsilon = [\epsilon_1, \ldots, \epsilon_n]^T$ independent zero mean stochastic error terms, and n the number of data points. Using standard estimators, we find the estimates $\hat{a} = 0.5414$ and $\hat{b} = 0.7637$. The fact that $\hat{b} < 1$ is no surprise; as the breakthroug time increases, the production curves become steeper. However, the fact that $\hat{b} < 1$ also implies that the linearity assumption in expression (7) has a limited range of validity, as we have the lower limit $t_{1/2} > t_{br}$.

The upper display of Figure 13 shows a histogram of the residuals $\epsilon_i, i = 1, \ldots, n$, whereas the lower display shows a kernel density estimate together with a Gaussian pdf assuming independent errors, $\epsilon_i \sim N(0, \sigma_\epsilon^2), i = 1, \ldots, n$. For our data, $\hat{\sigma}_\epsilon^2 = 8.833 \times 10^{-4}$ using the standard estimator. This is very low. Thus, given the breakthrough time, expression (7) states that

$$(\log t_{1/2} | \log t_{br})_i \sim N\Big(a + b(\log t_{br})_i, \sigma_\epsilon^2\Big) \tag{8}$$

with $\sim$ meaning 'distributed as'. The mean of the distribution (8) differs only slightly from the time where the variance is at its maximum value, see Figure 14. As $t_{1/2}$ appears where the production curves are at their steepest, this seems reasonable. Further, this tells how small the variance actually becomes when knowing the breakthrough time. As a result, predicting t_q for q small or large compared to $1/2$ will be even more precise.

Taking the exponential of expression (7) results in

$$t_{1/2} \quad \propto \quad 10^{\varepsilon} t_{br}^{b}. \tag{9}$$

with $\propto$ meaning 'proportional to'. In the framework of percolation, expression (9) states that the half time scales as a power of the breakthrough time. Further, expression (9) also states that if one knows the analytical distribution of the errors ϵ as a function of the reservoir parameters, one can caclulate the pdf for $t_{1/2}$. As the variance of residuals $\epsilon_i, i = 1, \ldots n$ decreases, the pdf $f(\epsilon)$ approaches a Dirac delta function, resulting in the distribution of $t_{1/2}$, as shown in expression (9), being merely a scaled power transform of the distribution of t_{br}. From Figure 11, this kind of behaviour is expected, as for this set of data $\hat{\sigma}_\epsilon^2$ is very small.

6 Summary and Discussion

In this report we have shown how knowing the breakthrough time increases precision in prediction of future oil production. This is done by estimating and analyzing the two first empirical moments. We have then used these two empirical moments to fit a Gaussian pdf to the production data, placing the analysis in a more probabilistic setting. We have then shown how this can be used for quantitatively comparing the decrease in variance which the increased knowledge t_{br} results in. Finally we have shown that there is a strong correlation between the breakthrough time and the half-time, or any other time after breakthrough, thus allowing for precise prediction of when a given fraction of the reservoir is produced.

Note that we have only analyzed average production data from one specific percolation system. In order to make our findings more general, work must be done on several different percolation systems with different system parameters. However, our analysis indicates that there exists some systematic behaviour in empirical means and variances within a percolation system, as well as strong correlations between the half-time and the breakthrough time.

We are aware that there are problems with our set of data. These problems divide into two different categories. First of all, our production data are integrated from travel time distributions. That is, they are not real production curves, but represent a smooth, average beaviour. Thus, the systematic behaviour might be more apparent in our data than would be in real production data. Secondly, we have very few data. The total of 141 traces result in only 10-20 traces for each of the conditional sets of data. Thus our analysis, particularly in the empirical variances, is very sensitive to outliers, and the difference in numbers of data for each breakthrough time highly affects possible trends and, thus, conclusions. We would assume that the empirical variances would have some sort of trends, like the empirical means have.

These shortcomings aside, the analysis shows that more work should be done along these lines.

References

[1] Buldyrev, S.V. (2002) History Matching for Oil Production, *Research Note*, Center for Folkymer Studies and Department of Physics, Boston University.

[2] Nikolay, V.D., Buldyrev, S.V., Havlin, S., King, P.R., Lee, Y. and Stanley, H.E (1999) Distribution of Shortest Paths in Percolation, *Physica A*, **266**, 55-61

[3] Evans, M., Hastings, N. and Peacock, B. (2000) *Statistical Distributions*, John Wiley & Sons, Inc.

[4] King, P.R. (1990) The Connectivity and Conductivity of Overlapping Sand Bodies, *North Sea Oil and Gas Reservoirs-II*, Norwegian University of Science and Technology.

[5] King, P.R., Andrade Jr.,J.S., Buldyrev, S.V., Dokholyan, N., Youngki, L., Havlin, S. and Stanley, H.E. (1999a) Predicting oil recovery using percolation, *Physica A*, **266**, 107-114.

[6] King, P.R., Buldyrev, S.V., Dokholyan, N.V, Havlin, S., Lee, Y., Paul, G. and Stanley, H.E. (1999b) Applications of statistical physics to the oil industry: predicting oil recovery using percolation theory, *Physica A*, **274**, 60-66

[7] King, P.R., Buldyrev, S.V., Dokholyan, N.V, Havlin, S., Lopez, E., Paul, G. and Stanley, H.E. (2002) Uncertainty in oil production predicted by percolation theory, *Physica A*, **306**, 376-380

[8] Ripley, B.D. (1994) *Stochastic Simulation*, John Wiley & Sons, New York

[9] Stauffer, D. and Aharony, A. (1992) *Introduction to Percolation Theory*, Taylor and Francis, London, 2nd edition

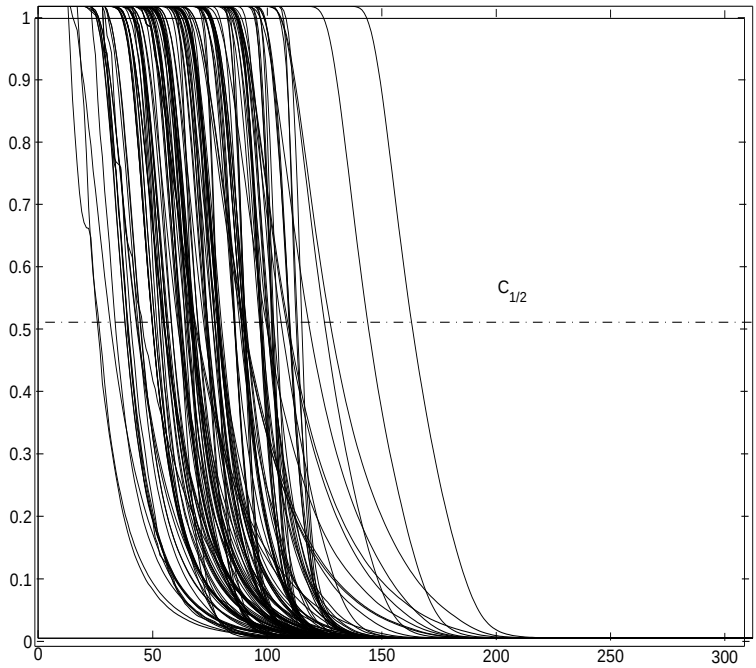

Figure 1: Production data, integrated from travelling time distributions.

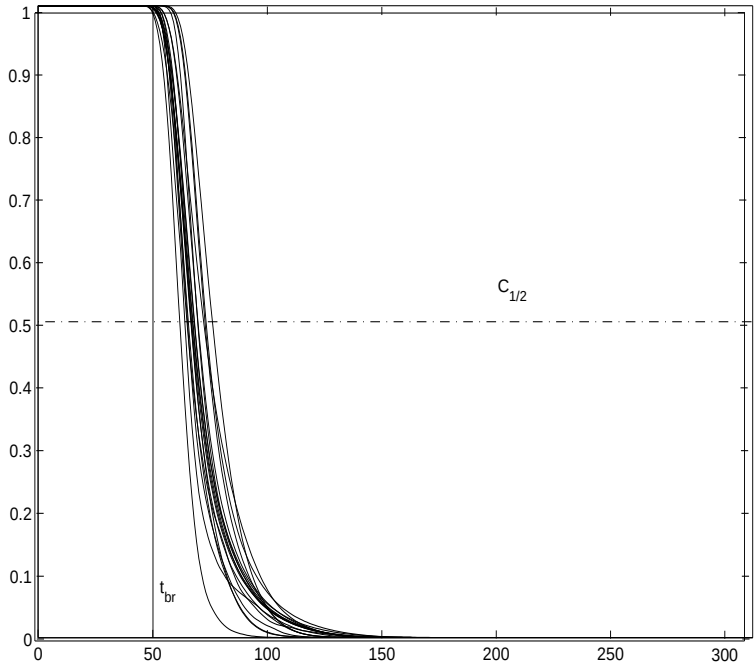

Figure 2: Production data, integrated from travelling time distributions and conditioned on breakthrough time $t_{br} = 50$.

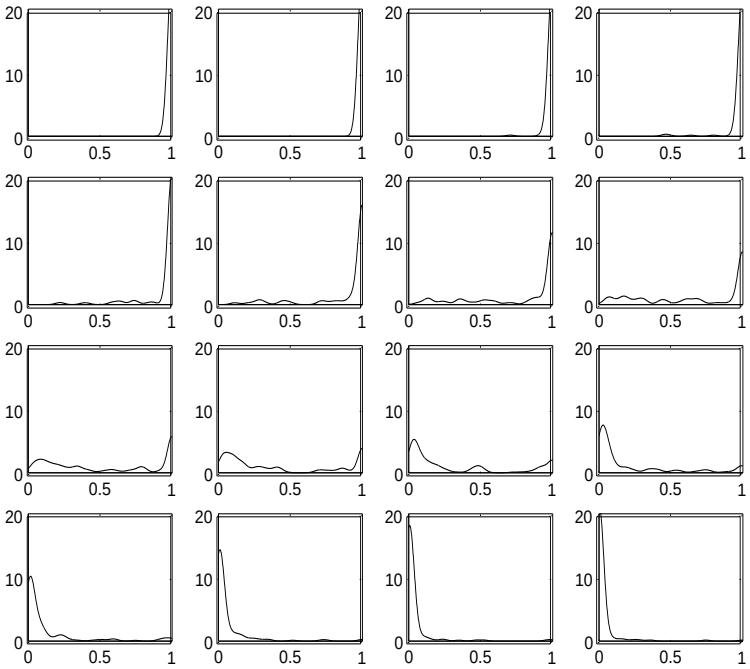

Figure 3: Kernel density estimates for the unconditional production data in Figure 1 for $t = [0, 9, 18, \ldots, 135]$.

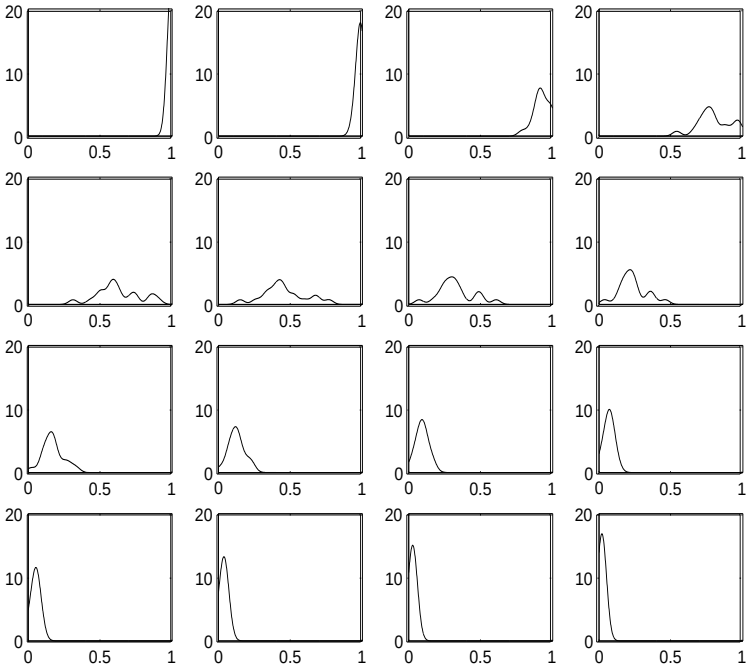

Figure 4: Kernel density estimates for the conditional production data in Figure 2 for $t = [50, 54, \ldots, 110]$.

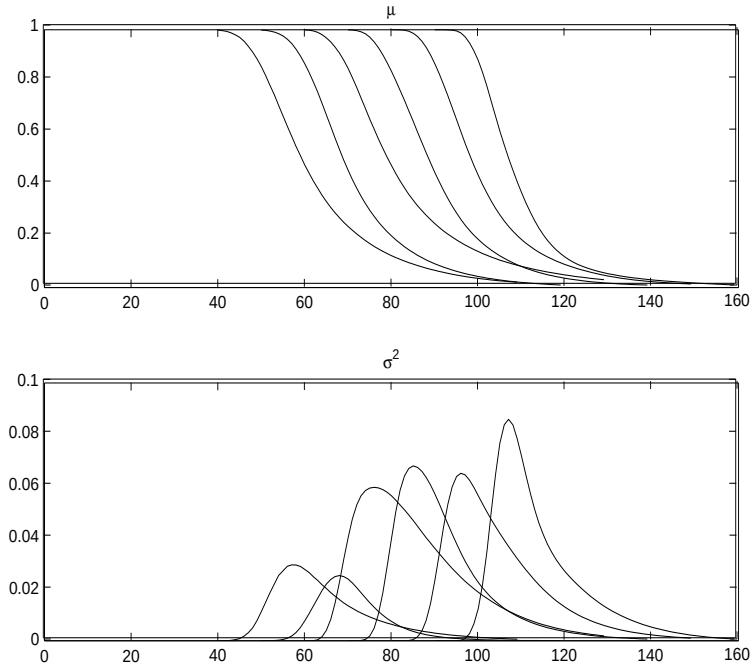

Figure 5: Empirical mean $\hat{\mu}(t)$ (upper display) and empirical variance $\hat{\sigma}^2(t)$ (lower display) when conditioning on breakthrough times $t_{br} = [40, 50, \ldots, 90]$.

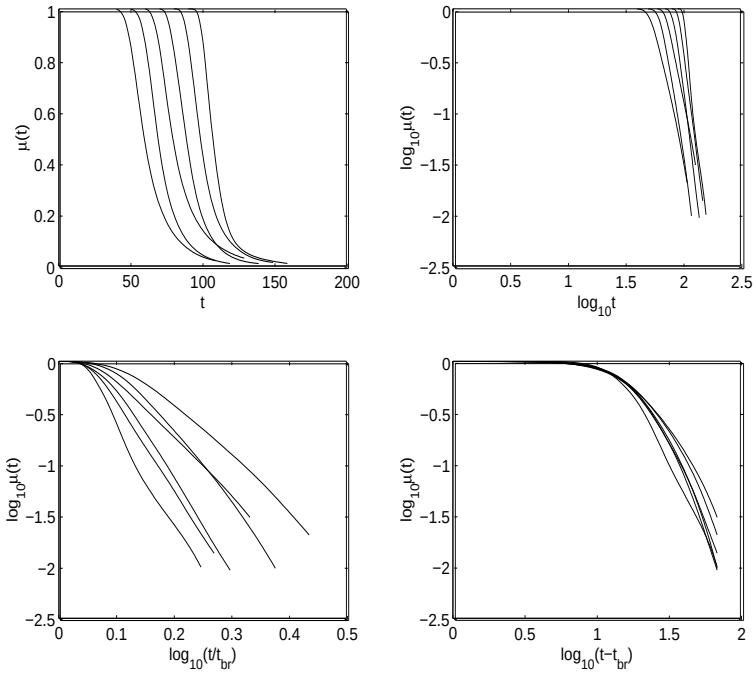

Figure 6: Different transforms of the empirical means in Figure 5.

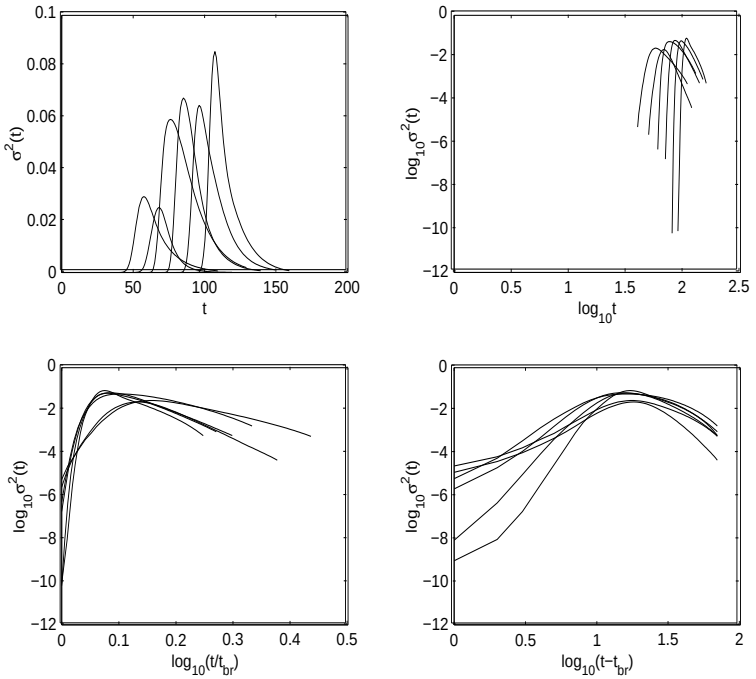

Figure 7: Different transforms of the empirical variances in Figure 5.

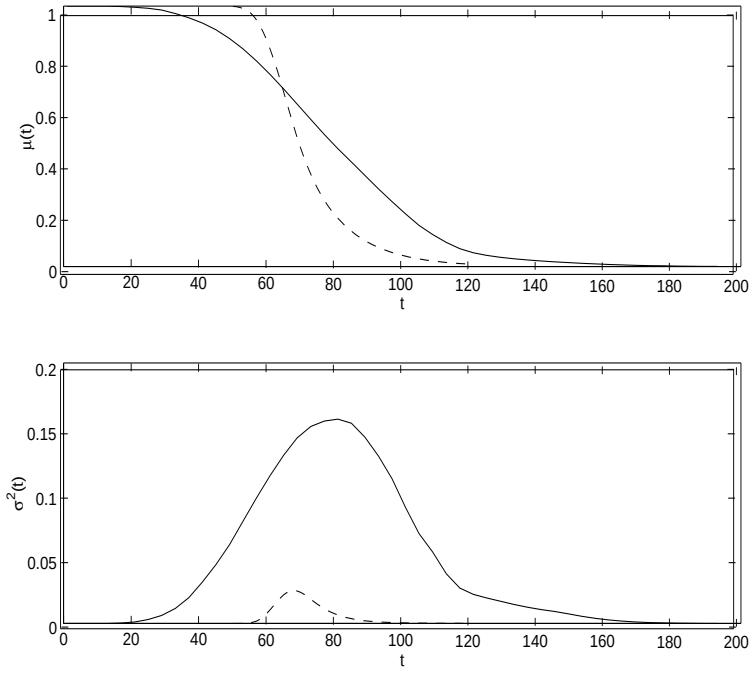

Figure 8: Empirical mean $\hat{\mu}(t)$ (upper display) and empirical variance $\hat{\sigma}^2(t)$ (lower display) for the unconditional data in Figure 1 (solid line) vs. the conditional data in Figure 2 (dotted line).

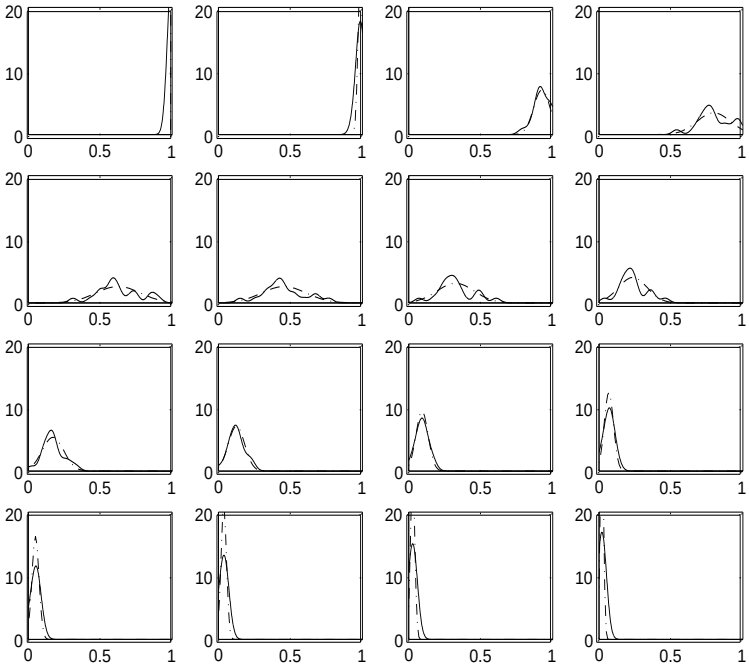

Figure 9: Kernel density estimates for the conditional data in Figure 2 (solid line), together with Gaussian probability density function estimates based on empirical mean and variance (dotted line).

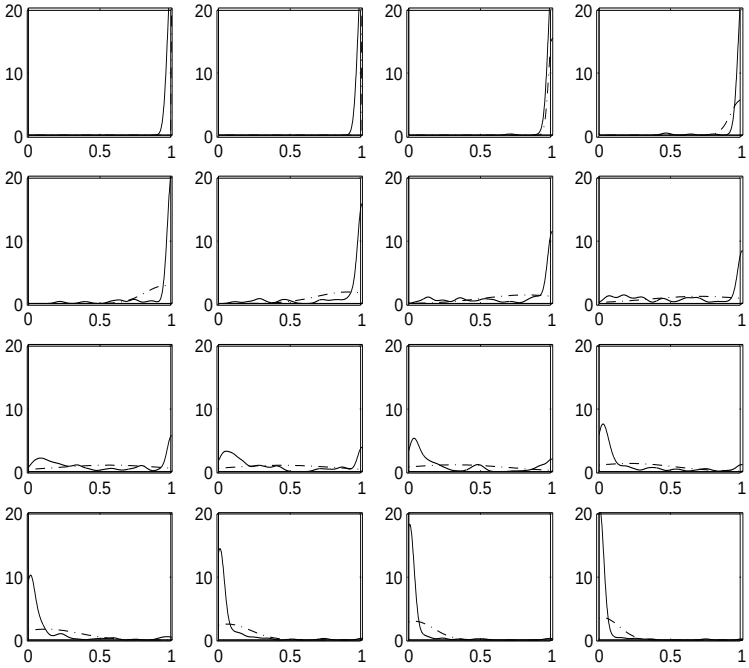

Figure 10: Kernel density estimates for the unconditional data in Figure 1 (solid line), together with Gaussian probability density function estimates based on empirical mean and variance (dotted line).

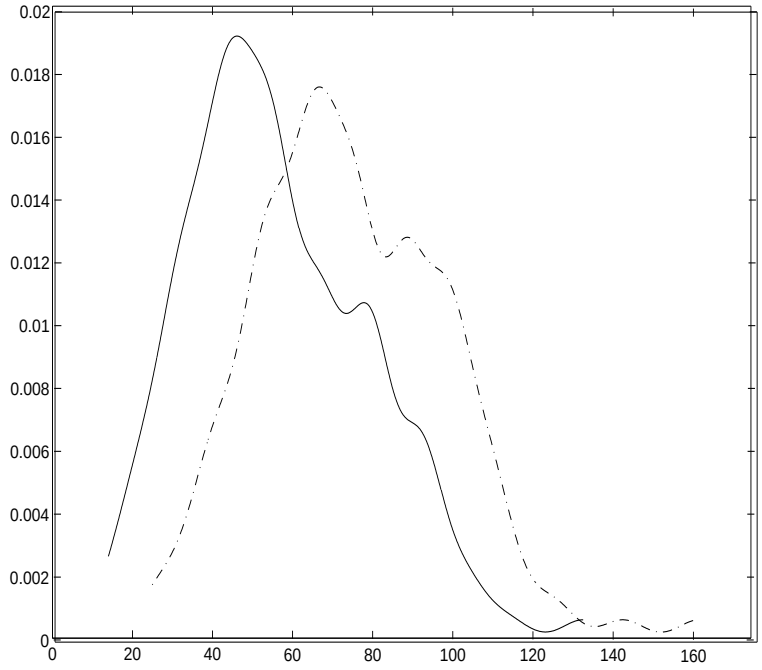

Figure 11: Kernel density estimate for the breakthrough time (solid line) and half time (dotted line) for the 141 traces of the production data in Figure 1.

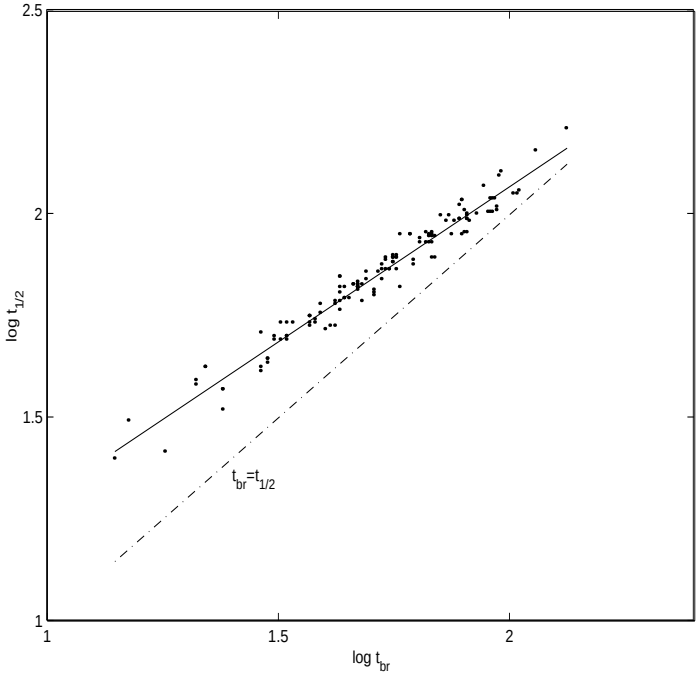

Figure 12: Log-log plot of $t_{1/2}$ vs t_{br}, together with dotted line having slope 1 and solid line having slope 0.7637.

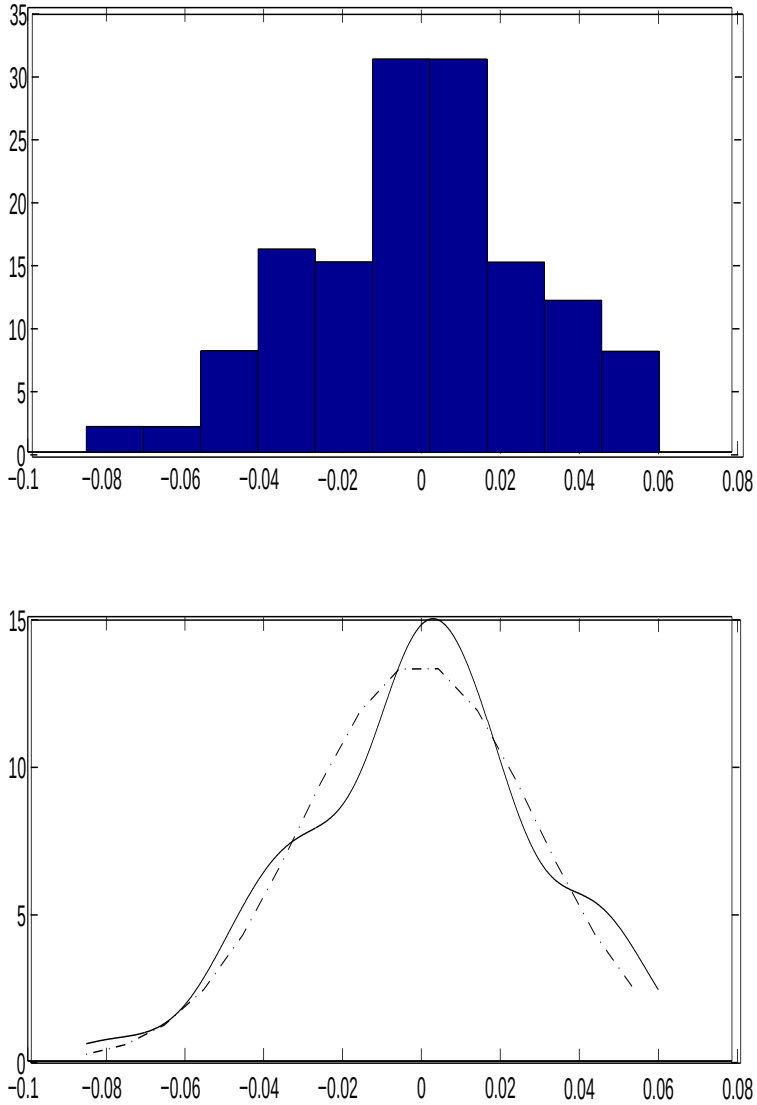

Figure 13: Upper display: Histogram of residuals from the linearization in Figure 12. Lower display: Kernel density estimate of residuals from the linearization in Figure 12 (solid line) and Gaussian probability density function approximation (dotted line).

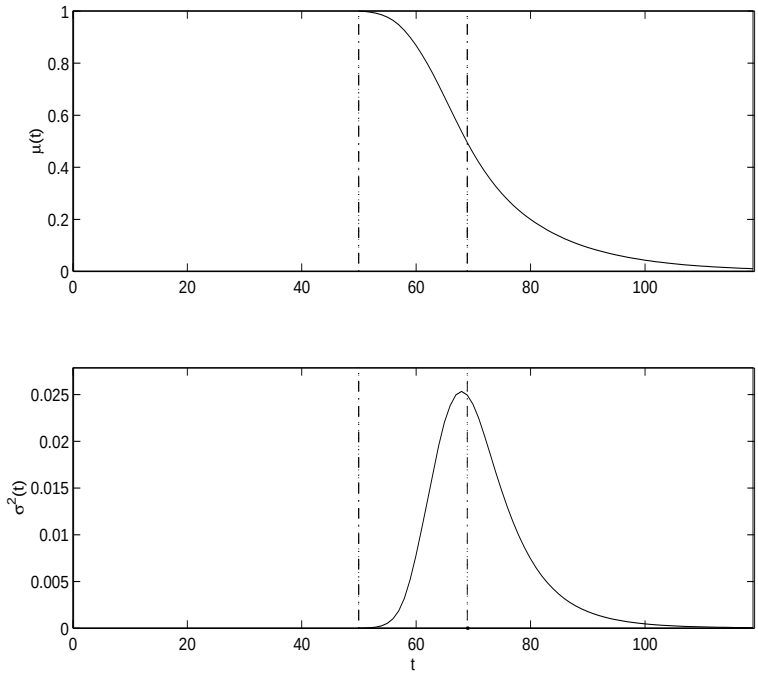

Figure 14: Empirical mean $\hat{\mu}(t)$ (upper display) and empirical variance $\hat{\sigma}^2(t)$ (lower display) given breakthrough time $t_{br} = 50$ (left dotted line), together with the linearly predicted $t_{1/2}$ (right dotted line).