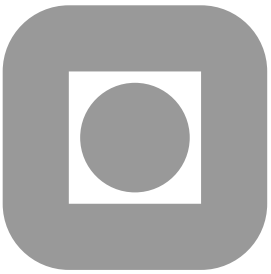
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Piecewise Affine Inverse Problems

by

Odd Kolbjørnsen and Henning Omre

PREPRINT
STATISTICS NO. 6/2001
ISSN: 0804-9173



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S6-2001.ps
Odd Kolbjørnsen has homepage: http://www.math.ntnu.no/∼ok
E-mail: ok@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

Piecewise Affine Inverse Problems.

Odd Kolbjørnsen [1] and Henning Omre [1]
Department of mathematical sciences,
Norwegian university of science and technology
Trondheim
Norway

Abstract

Piecewise affine inverse problems, are solved in a Bayesian framework with a Gaussian random field prior on the model space. Piecewise affine inverse problems denotes the case where the observation is the outcome of a piecewise affine operator applied to the object of interest, possibly with errors. A piecewise affine operator is defined by partitioning the model space and assign a specific affine operator to each part. Both problems with a discrete finite partition and a continuous partition are considered. The main result is that the posterior distribution is found to be a mixture of truncated Gaussian distributions, and the expression for the mixing distribution is partially analytical tractable. The focus is to sample from the posterior, and the decomposition is used to make a sampling algorithm that make use of the affine structures of the problem. The performance of the algorithm is virtually independent of the size of observation error, and work in particular for problems with exact observations, most general purpose algorithms tends to fail in this case. Piecewise affine inverse problems is a general class of nonlinear inverse problems, in particular inverse problems obeying a variational structure such as the Fermat principle of cross well tomography is of this type. Examples included in this paper covers reconstruction based on; the L^1 norm of a vector, extreme values on a closed interval; and travel times from reflection seismic.

KEY WORDS: *Bayesian statistic, Nonlinear inverse problem, Piecewise linear inverse problem, Sampling algorithm.*

[1]Odd Kolbjørnsen is a Ph.D student in statistics,(email: ok@stat.ntnu.no). Henning Omre is Professor of statistics, (email: omre@stat.ntnu.no). Both authors are at the department of mathematical sciences, Norwegian university of science and technology, 7491 Trondheim, Norway. This work was supported by the Research Council of Norway.

1 Introduction

Inverse problems arise in many scientific settings and engineering applications, such as oceanography, astrophysics, medical imaging and geophysical exploration. In many areas inverse problems have become more important, due to the increasing use of remote sensing and indirect measurements.

The feature of inverse problems, that makes it different from parameter estimation in large sample theory as presented in Lehmann (1999), is that the number of unknown parameters is of the same order or larger than the number of observations.

In the statistical literature there are two main strategies for solving inverse problems, namely the minimax and the Bayesian approach. For the early theory of linear inverse problems, the two approaches essentially give the same solution. A classical minimax result of Pinsker (1980), prove that the solution obtained by quadratic regularization, (Tikhonov 1963), have the optimal rate of convergence in the zero noise limit. In the Bayesian approach a quadratic regularizer is formally equivalent with a Gaussian random field prior as described in Tarantola (1987) and Wahba (1990). The conditional expectation in this model will again be the solution found by regularization. This solution is also denoted the maximum posterior (MAP) solution since it is the mode in the posterior distribution. The minimax theory will not be pursued any further, although there are many resent results in this field, (Donoho 1992; Abramovich and Silverman 1998; Johnstone 1999).

In this article a Bayesian approach with a Gaussian random field prior is used, to solve piecewise affine inverse problems. The maximum posterior (MAP) solution is commonly used also for nonlinear inverse problems (O'Sullivan 1986; Tarantola 1987; Wahba 1990). For genuine nonlinear inverse problems this choice is not obvious, since the posterior frequently have multiple modes. Example 1 below, is an example where any single solution would be unsatisfactory. In the Bayesian tradition the full posterior distribution is the solution. Multiple random samples from the posterior is a common Monte Carlo representation of the posterior, and is the approach used in this article.

The class of piecewise affine inverse problems is very general, in particular inverse problems having a variational structure, such as the Fermat principle of travel time tomography (Berryman 1997), can be phrased as piecewise affine inverse problems.

Piecewise affine inverse problems with both a finite discrete index and a continuous index are considered in this article. The main result is that the posterior is calculated to be a mixture of truncated Gaussian distributions in both cases. The decomposition is essential for efficiently exploiting the affine structures of the inverse problem in any sampling algorithm. An algorithm which uses this decomposition to sample the posterior is proposed. Two examples of piecewise affine inverse problems are used to illustrate the concepts as they are introduced.

Example 1: The L^1 norm
Problem: A two dimensional vector is to be recovered, based on one observation of the L^1 norm of the vector. This is a low dimensional synthetic example that gives results that is easy to illustrate graphically. □

Example 2: The meteorologist.
Problem: A meteorologist is supposed to continuously monitor the temperature during a day. After reading the thermometer in the morning, he leaves it unattended the rest of the day. The next day when he return to the office, he reads the thermometer again. In the 24 hour period between the two readings, only the day maximum and the day minimum value can be obtained from the thermometer, not at what time they occurred. The meteorologist need to recover the entire temperature field of the previous day. □

A larger example where the methodology is applied to prediction of a geological horizon based on travel times from reflection seismic is included in Section 4.

The problem of piecewise linear nonlinearities has been encountered by Kolessa (1986), which develop this for recursive filtering in dynamical systems. The case considered by Kolessa may be regarded as a special case of the situation in the current paper. The algorithm presented here, may be adapted as a particle filter for the case considered by Kolessa. This type of model have applications in finance (Liu and Chen 1998).

2 Bayesian approach to inverse problems

To recover a function z is denoted an inverse problem when the observations are related to transforms of the function. When z is assumed to be in a

function space $\mathcal{Z}$, the observations are $\boldsymbol{o} = \boldsymbol{\omega}(z) + \boldsymbol{\varepsilon}$, with $\boldsymbol{\omega} : \mathcal{Z} \to \mathbf{R}^r$ being the known forward map of the inverse problem, and $\boldsymbol{\varepsilon} \in \mathbf{R}^r$ being observation error.

The Bayesian approach to inverse problems, is in principle no different from the Bayesian approach to any other problem (Gelman,Carlin,Stern and Rubin 1995). The analyzer must specify a prior on the parameter space and a likelihood for the observations. The Bayesian framework is very flexible and it allows for stochastic modeling of the prior distribution to quantify lack of knowledge and to merge observations from different sources.

A generic probability distribution is denoted $\mathcal{D}\{\cdot\}$ and $P\{\cdot\}$ denote a generic probability, when $\mathcal{D}\{\cdot\}$ is applied to a random vector it should be given the interpretation as density. In this article, the prior distribution $\mathcal{D}\{Z\}$, is modeled by a Gaussian random field (Vanmarcke 1983).

Example 1 The L^1 norm continued
The random field is defined on the index set $\{1, 2\}$ and is assumed to have a bigaussian prior distribution, hence $\mathcal{D}\{\boldsymbol{Z}\} = N_2(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{Z}})$ with $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{12} = \sigma_{21} = 0.5$. The contour lines of this density is plotted in Figure 1.

□

Example 2 The meteorologist continued
The meteorologist has developed a stochastic model for the temperature during one day, $t \in [0, 24]$.

$$Z(t) = Z_0 + Z_1 \sin \frac{\pi t}{12} + Z_2 \cos \frac{\pi t}{12} + \widetilde{Z}(t),$$

with Z_0, Z_1, Z_2 and $\widetilde{Z}(t)$ being independent; $\mathcal{D}\{Z_0\} = N(15, 5^2)$, $\mathcal{D}\{Z_1\} = N(5, 2^2)$, $\mathcal{D}\{Z_2\} = N(0, 0.5^2)$, and $\widetilde{Z}(t)$ is a zero mean residual stochastic field with covariance function $\mathrm{C}_{\widetilde{Z}}(t, t+h) = \exp\{-3|h/4|^2\}$. Hence Z is a Gaussian random field, being twice continuous differentiable for $t \in [0, 24]$. Figure 3 shows 200 samples from this prior distribution. □

The likelihood model is assumed to contain additive Gaussian error, hence

$$\mathcal{D}\{\boldsymbol{O}|Z = z\} = N_r(\boldsymbol{\omega}(z), \boldsymbol{\Sigma}^z_{\boldsymbol{\varepsilon}}),$$

with N_r being the r dimensional multi Gaussian distribution; $\boldsymbol{\omega}(z)$ being the forward map of the inverse problem and $\boldsymbol{\Sigma}^z_{\boldsymbol{\varepsilon}}$ being the covariance of the observation error, note that this in general may depend on z.

The Bayesian answer to any problem is contained by the posterior distribution, $\mathcal{D}\{Z|\boldsymbol{O}=\boldsymbol{o}\}$, being the conditional distribution of the parameters given the data. The posterior distribution is formally proportional to the product of the prior and the likelihood,

$$\mathcal{D}\{Z|\boldsymbol{O}=\boldsymbol{o}\} \propto \mathcal{D}\{Z\}\mathcal{D}\{\boldsymbol{O}|Z=z\}\ . \tag{1}$$

The object is to sample the posterior distribution to obtain a Monte Carlo representation, since full analytical treatment is impossible for the cases discussed here. Several general purpose sampling algorithms have been developed. Most common is rejection sampling (von Neumann 1951), resampling schemes (Rubin 1988) and Markov chain based methods (Hastings 1970). Most general purpose algorithms become slow if the observations have high precision. The motivation for this article is to speed up sampling algorithms for piecewise affine inverse problems, by use of the affine structures of the inverse problem.

3 Piecewise affine inverse problems

In the first part of this section, piecewise affine inverse problems are defined in general. In the following two parts, attention is drawn to two special cases where the piecewise affine operator have a finite and a continuous index. For both cases the posterior is calculated as a mixing distribution, and a sampling algorithm based on rejection sampling is proposed. The algorithm produces exact independent samples from the posterior distribution. For the case of a finite index, the algorithm is fully specified. For the case of a continuous index, the mixing distribution to be sampled is only known up to a normalizing constant.

3.1 Problem Definition

Piecewise affine inverse problems are defined by the forward map being a piecewise affine operator.

Definition 1 (Piecewise affine operator) *An operator $\boldsymbol{\omega} : \mathcal{Z} \to \mathbf{R}^r$, is said to be piecewise affine, if it can be represented in the following way:*

$$\boldsymbol{\omega}(z) = \boldsymbol{B}_x z + \boldsymbol{b}_x \ \ \textit{for} \ \ z \in \mathcal{A}_x\ ; \quad x \in \mathcal{X}$$

with $\mathcal{X}$ being an index set, $\{\mathcal{A}_x\}_{x\in\mathcal{X}}$ being a partition of $\mathcal{Z}$; $\boldsymbol{B}_x : \mathcal{Z} \to \mathbf{R}^r$ being bounded linear operators on $\mathcal{Z}$ and $\boldsymbol{b}_x$ being r dimensional vectors. The indexed set of triplets $\{\mathcal{A}_x, \boldsymbol{B}_x, \boldsymbol{b}_x\}_{x\in\mathcal{X}}$ are the parameters of the piecewise affine operator.

This definition of piecewise affine operators is very general. Its usefulness depends on the index set $\mathcal{X}$. One special case is obtained by $\mathcal{X} = \{1\}$ in which the class of affine operators are obtained. The other extreme is letting $\mathcal{X} = \mathcal{Z}$, in which any operator can be represented as piecewise affine. In this article intermediate groups are considered, by letting $\mathcal{X} = \{1, ..., m\}$ and $\mathcal{X} \subset \mathbf{R}^d$. The two examples above being one of each type. It will always be assumed that the operator is measurable with respect to the prior measure, $\mathcal{D}\{Z\}$, on $\mathcal{Z}$.

Define also the affine operators of a piecewise affine operator.

Definition 2 (The affine operators of a piecewise affine operator) *Let $\boldsymbol{\omega}$ be a piecewise affine operator with parameters $\{\mathcal{A}_x, \boldsymbol{B}_x, \boldsymbol{b}_x\}_{x\in\mathcal{X}}$. Then the affine operators $\{\boldsymbol{\omega}_x\}_{x\in\mathcal{X}}$ of $\boldsymbol{\omega}$ is defined by*

$$\boldsymbol{\omega}_x(z) = \boldsymbol{B}_x z + \boldsymbol{b}_x \ , \ z \in \mathcal{Z} \ ; \quad x \in \mathcal{X}$$

The affine operators of a given piecewise affine operator is hence defined by $\{\boldsymbol{B}_x, \boldsymbol{b}_x\}_{x\in\mathcal{X}}$, hence extending the affine pieces to the entire $\mathcal{Z}$.

The stochastic variable, $\boldsymbol{O}$, actually observed is defined through its conditional distribution,

$$\mathcal{D}\{\boldsymbol{O}|Z = z, \, z \in \mathcal{A}_x\} = N_r(\boldsymbol{\omega}(z), \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^x), \tag{2}$$

with $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^x$ being the covariance matrix of the observation error, possible dependent on the index. The marginal distribution of $\boldsymbol{O}$ is not obvious due to the nonlinearity of $\boldsymbol{\omega}$. Define also the stochastic variables $\{\boldsymbol{O}_x\}_{x\in\mathcal{X}}$, that would correspond to observing the affine operators of $\boldsymbol{\omega}$,

$$\mathcal{D}\{\boldsymbol{O}_x|Z = z\} = N_r(\boldsymbol{\omega}_x(z), \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^x) \ , \tag{3}$$

with $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^x$ being as for the likelihood in Expression (2). Note that the marginal distribution of $\boldsymbol{O}_x$, is Gaussian, with parameters:

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{O}_x} &= \mathrm{E}\{\boldsymbol{B}_x Z\} + \boldsymbol{b}_x \\ \boldsymbol{\Sigma}_{\boldsymbol{O}_x} &= \mathrm{Cov}\{\boldsymbol{B}_x Z\} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^x \end{aligned}$$

In general $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}^{x}$ may be singular, but $\mathbf{\Sigma}_{\boldsymbol{O}_x}$ should be of full rank.

Example 1 The L^1 norm continued
The piecewise affine operator of this problem is the L^1 norm $\boldsymbol{\omega}(\boldsymbol{z}) = \|\boldsymbol{z}\|_1$, defined as:

$$\boldsymbol{\omega}(\boldsymbol{z}) = \begin{cases} z_1 + z_2 & z_1 \geq 0 \; ; \; z_2 \geq 0 \\ z_1 - z_2 & z_1 \geq 0 \; ; \; z_2 < 0 \\ -z_1 + z_2 & z_1 < 0 \; ; \; z_2 \geq 0 \\ -z_1 - z_2 & z_1 < 0 \; ; \; z_2 < 0 \end{cases} .$$

This problem has a discrete index $\mathcal{X} = \{1, 2, 3, 4\}$. Since both $\mathcal{X}$ and the index set of $\boldsymbol{Z}$ is finite, the parameters of the piecewise affine operator can be written explicitly and is listed in Table 1. In Figure 1, the contour $\|\boldsymbol{z}\|_1 = 2$ is plotted as a solid line, the extensions of the affine operators at the same level are plotted as dotted lines. Assume that $o = 2$ is observed and that the observation error is distributed as $N(0, 0.1^2)$.

□

Example 2 The meteorologist continued
The observations the meteorologist have, are the values each morning and the global extrema in between. To define this operator, consider its action on the function, $z(t)$. The piecewise affine operator is:

$$\boldsymbol{\omega}(z) = \begin{cases} z(0) \\ z(t_{\max}) \\ z(t_{\min}) \\ z(24) \end{cases} \qquad z \in \mathcal{A}_{t_{\max}, t_{\min}} ,$$

with

$$t_{\max} = \arg \max_{t \in [0,24]} z(t) \quad , \quad t_{\min} = \arg \min_{t \in [0,24]} z(t) ,$$

and $\mathcal{A}_{t_{\max}, t_{\min}} = \{z \in C^2([0, 24]) : z(t_{\max}) \geq z(t) \geq z(t_{\min}) \; ; \; \forall t \in [0, 24]\}$. Hence the operator is indexed by a continuous index $\mathcal{T} = [0, 24] \times [0, 24]$. The index correspond to the location where the maximum and the minimum occur, this is of course not observed. Assume $z(0) = 19.78$, $z(24) = 19.74$, $\max_{0 \leq t \leq 24} z(t) = 26.04$ and $\min_{0 \leq t \leq 24} z(t) = 16.61$ is observed, and that the observations are exact.

□

3.2 Finite Index

The objective is to assess the posterior distribution. The full posterior will not be Gaussian due to the nonlinearity in $\boldsymbol{\omega}$, although within each $\mathcal{A}_x$ the posterior will be a truncated Gaussian distribution. The following theorem characterizes the posterior in terms of a mixing of these truncated Gaussians.

Theorem 1 (Finite partition) *Let Z be a Gaussian random field on $\mathcal{Z}$ with distribution $\mathcal{D}\{Z\}$, such that $P\{Z \in \mathcal{Z}\} = 1$, and $\boldsymbol{\omega} : \mathcal{Z} \to \mathbf{R}^r$ be a $\mathcal{D}\{Z\}$-measurable piecewise affine operator with index set $\mathcal{X} \subset \boldsymbol{N}$ with $|\mathcal{X}| < \infty$ and parameters $\{\mathcal{A}_x, \boldsymbol{B}_x, \boldsymbol{b}_x\}_{x\in\mathcal{X}}$ having affine operators $\{\boldsymbol{\omega}_x\}_{x\in\mathcal{X}}$. Further let $\boldsymbol{O}$ and $\{\boldsymbol{O}_x\}_{x\in\mathcal{X}}$ be defined by Expressions (2) and (3) above. Assume further that:*

$$\begin{array}{ll} \forall x \in \mathcal{X} & \mathit{rank}\{\boldsymbol{\Sigma}_{\boldsymbol{O}_x}\} = r \\ \exists x \in \mathcal{X} & P\{Z \in \mathcal{A}_x | \boldsymbol{O}_x = \boldsymbol{o}\} > 0 \end{array}$$

Then,

$$\begin{aligned} \mathcal{D}\{Z|\boldsymbol{O}=\boldsymbol{o}\} &= \sum_{x\in\mathcal{X}} \mathcal{D}\{Z|\boldsymbol{O}_x=\boldsymbol{o}, Z\in\mathcal{A}_x\} \\ &\quad \times \frac{P\{Z\in\mathcal{A}_x|\boldsymbol{O}_x=\boldsymbol{o}\}\cdot\mathcal{D}\{\boldsymbol{O}_x=\boldsymbol{o}\}}{\mathcal{D}\{\boldsymbol{O}=\boldsymbol{o}\}} \end{aligned} \tag{4}$$

Proof :
Use the law of total probability to factorize the distribution into a sum over the partitions;

$$\mathcal{D}\{Z|\boldsymbol{O}=\boldsymbol{o}\} = \sum_{x\in\mathcal{X}} \mathcal{D}\{Z|\boldsymbol{O}=\boldsymbol{o}, Z\in\mathcal{A}_x\} P(Z\in\mathcal{A}_x|\boldsymbol{O}=\boldsymbol{o}).$$

By definition,

$$\mathcal{D}\{Z|\boldsymbol{O}=\boldsymbol{o}, Z\in\mathcal{A}_x\} = \mathcal{D}\{Z|\boldsymbol{O}_x=\boldsymbol{o}, Z\in\mathcal{A}_x\},$$

Further let $\boldsymbol{I}(z\in\mathcal{A}_x)$ be the indicator function of $\mathcal{A}_x$. Then,

$$\begin{aligned} \mathrm{P}\{Z\in\mathcal{A}_x|\boldsymbol{O}=\boldsymbol{o}\} &= \mathrm{E}\{\boldsymbol{I}(Z\in\mathcal{A}_x)|\boldsymbol{O}=\boldsymbol{o}\} \\ &= \frac{\mathcal{D}\{\boldsymbol{O}_x=\boldsymbol{o}\}}{\mathcal{D}\{\boldsymbol{O}=\boldsymbol{o}\}}\mathrm{E}\{\boldsymbol{I}(Z\in\mathcal{A}_x)|\boldsymbol{O}_x=\boldsymbol{o}\} \\ &= \frac{\mathrm{P}\{Z\in\mathcal{A}_x|\boldsymbol{O}_x=\boldsymbol{o}\}\cdot\mathcal{D}\{\boldsymbol{O}_x=\boldsymbol{o}\}}{\mathcal{D}\{\boldsymbol{O}=\boldsymbol{o}\}}. \end{aligned}$$

QED.

Note that the first term in Expression (4) is a truncated Gaussian distribution conditioned to the linear constraint $\boldsymbol{O}_x = \boldsymbol{o}$. The second term gives the posterior probability of being in $\mathcal{A}_x$. The rank criterion is to assure that the mixing distribution on $\mathcal{X}$ do not have any singularities. The positivity criterion assures that there exists a solution to the problem.

The mixing distribution of Theorem 1, provides a sampling strategy to sample the posterior distribution.

Algorithm 1 *FTGM-algorithm (Finite Truncated Gaussian Mixing)*

- *Sample* $x^* \sim q(x) \propto \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}\}$
- *Sample* $Z^* \sim \mathcal{D}\{Z|\boldsymbol{O}_{x^*} = \boldsymbol{o}\}$
- *If* $Z^* \in \mathcal{A}_{x^*}$ *stop.*

The algorithm is a variant of rejection sampling. The first objective is to obtain the posterior probability of being in $\mathcal{A}_x$. In Theorem 1 this probability is calculated as:

$$p(x) = \frac{P\{Z \in \mathcal{A}_x|\boldsymbol{O}_x = \boldsymbol{o}\} \cdot \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}\}}{\mathcal{D}\{\boldsymbol{O} = \boldsymbol{o}\}} .$$

Note that $\mathcal{D}\{\boldsymbol{O} = \boldsymbol{o}\}$ is just the normalizing constant in the expression and $0 < \mathrm{P}\{Z \in \mathcal{A}_x|\boldsymbol{O}_x = \boldsymbol{o}\} < 1$, hence proposing x according to $q(x) \propto \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}\}$ and accepting according to $\mathrm{P}\{Z \in \mathcal{A}_x|\boldsymbol{O}_x = \boldsymbol{o}\}$, is a standard rejection algorithm for sampling $p(x)$. The algorithm does not give optimal acceptance rate for $q(x)$, since the acceptance probability in the algorithm $\alpha(x) = \mathrm{P}\{Z \in \mathcal{A}_x|\boldsymbol{O}_x = \boldsymbol{o}\} < 1 \quad \forall x \in \mathcal{X}$. The value $\mathrm{P}\{Z \in \mathcal{A}_x|\boldsymbol{O}_x = \boldsymbol{o}\}$ is not easy to calculate in general, however it is possible to create an event with this probability by sampling Z^* from $\mathcal{D}\{Z|\boldsymbol{O}_x = \boldsymbol{o}\}$ and accepting if $Z^* \in \mathcal{A}_x$. The second objective is to sample $\mathcal{D}\{Z|\boldsymbol{O}_x = \boldsymbol{o}, Z \in \mathcal{A}_x\}$, this is a truncated Gaussian distribution. By sampling $\mathcal{D}\{Z|\boldsymbol{O}_x = \boldsymbol{o}\}$ and accepting if $Z \in \mathcal{A}_x$ this distribution is sampled correctly. Note that because of the positivity constraint, the algorithm produces a sample in finite time. Note also that the two acceptance criterions are the same for the two parts of the algorithm, hence they are coupled in the final algorithm.

The acceptance rate of the algorithm, can be calculated as:

$$p_{accept} = \frac{\mathcal{D}\{\boldsymbol{O} = \boldsymbol{o}\}}{\sum_{x \in \mathcal{X}} \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}\}}. \tag{5}$$

In comparison, the standard rejection algorithm with proposals drawn from the prior and accepted with probability proportional to the likelihood, have acceptance rate:

$$p_{accept} = \mathcal{D}\{\boldsymbol{O} = \boldsymbol{o}\}(2\pi)^{r/2}|\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}|^{1/2}, \tag{6}$$

for the case with fixed $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. This is heavily dependent on the scale of the observation error, whereas the sampling algorithm proposed is virtually independent of the size of the observation error. The dependence of observation error in the performance, is also known to be present in resampling techniques and algorithms based on naive use of Markov Chains. Since $\mathcal{X}$ contains a finite number of states, $q(x)$ can be calculated exactly. The FTGM-algorithm is hence fully specified and provides exact independent samples from the posterior.

Example 1. The L^1 norm continued
The FTGM-algorithm was implemented for this example. Figure 1 shows a scatter plot of the samples obtained using the algorithm. The posterior have four modes that are visible in the scatter plot. Note that the algorithm is exact and the samples are independent. The observed acceptance rate was 88.5% in this example.

From Expression (5) it is clear that the acceptance probability is dependent on the observed value, o. Figure 2, shows this dependence for Example 1. For comparison the acceptance rate for naive rejection sampling is plotted as a dotted line in the same figure. □

3.3 Continuous Index

The objective is to assess the posterior. In the case of a finite index one is free to choose the partition $\{\mathcal{A}_x\}_{x \in \mathcal{X}}$, and the operator $\boldsymbol{B}_x$, as long as the resulting operator $\boldsymbol{\omega}(Z)$ is measurable. For a continuous index further assumptions about the partition and the operator must be made.

Definition 3 (Restricted linear partition) *A partition $\{\mathcal{A}_x\}_{x \in \mathcal{X}}$ with $\mathcal{X} \subset \mathbf{R}^d$ of $\mathcal{Z}$ is a restricted linear partition if,*

$$\mathcal{A}_x = \{\boldsymbol{R}_x z + \boldsymbol{r}_x = \boldsymbol{0}\} \cap \mathcal{C}_x,$$

with $\boldsymbol{R}_x : \mathcal{Z} \to \mathbf{R}^d$ *being a bounded linear operator;* $\boldsymbol{r}_x$ *being a* d*-dimensional vector function; and* $\mathcal{C}_x$ *is any subset of* $\mathcal{Z}$*. The set of triplets* $\{\boldsymbol{R}_x, \boldsymbol{r}_x, \mathcal{C}_x\}_{x\in\mathcal{X}}$ *are the parameters of the restricted linear partition.*

Hence for a restricted linear partition, having parameters $\{\boldsymbol{R}_x, \boldsymbol{r}_x, \mathcal{C}_x\}_{x\in\mathcal{X}}$, the part $\mathcal{A}_{x^*}$ contain only those z's in the subset $\mathcal{C}_{x^*}$ for which $\boldsymbol{R}_{x^*}z + \boldsymbol{r}_{x^*} = \mathbf{0}$. Except from the fact that $\{\mathcal{A}_x\}_{x\in\mathcal{X}}$ should make up a partition, the parameters of the partition, can be chosen freely. Assumptions must also be made about the dependence of the index for the linear operators involved.

Definition 4 (Index continuity of operator) *An indexed set of bounded operators,* $\boldsymbol{B}_x : \mathcal{Z} \to \mathbf{R}^r$ *with* $x \in \mathcal{X} \subset \mathbf{R}^d$ *is continuous in the index with respect to a probability measure* $\mathcal{D}\{Z\}$ *on* $\mathcal{Z}$*, if:*

$$P\Big\{ \lim_{\Delta x \to 0} \|\boldsymbol{B}_x Z - \boldsymbol{B}_{x+\Delta x} Z\| = 0 \,;\, \forall x \in \mathcal{X} \Big\} = 1$$

That is, define a d-parameter, r-dimensional random field on $\mathcal{X}$, by for each $x \in \mathcal{X}$ associating the r-dimensional vector $\boldsymbol{B}_x Z$. Definition 4 implies sample path continuity for this field on $\mathcal{X}$.

To define an expression for the posterior in the continuous index case, similar to the result of Theorem 1, further regularity conditions must be made about the restricted linear partition

Definition 5 (Regular restricted linear partition) *Let* $Leb(\cdot)$ *denote the Lebesgue-measure on* $\mathbf{R}^d$*. A restricted linear partition with parameters* $\{\boldsymbol{R}_x, \boldsymbol{r}_x, \mathcal{C}_x\}_{x\in\mathcal{X}}$ *is regular with respect to a prior measure* $\mathcal{D}\{Z\}$ *on* $\mathcal{Z}$ *if:*

i) $rank(Cov\{\boldsymbol{R}_x Z\}) = d \,;\, \forall x \in \mathcal{X}$

ii) $\boldsymbol{\nabla}_x \boldsymbol{R}_x : \mathcal{Z} \to \mathbf{R}^{d\times d}$ *is a bounded linear operator,continuous in index.*

iii) $P\{Leb(\{x : \det\{\boldsymbol{\nabla}_x \boldsymbol{R}_x Z\} = 0\}) = 0\} = 1$

iv) $\boldsymbol{r}_x$ *is continuously differentiable in each component.*

v) $\mathcal{C}_x$ *is* $\mathcal{D}\{Z|\boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}\}$ *measurable*

vi) $\exists \mathcal{B} \in \mathcal{X}$ *with* $Leb(\mathcal{B}) > 0 : P\{Z \in \mathcal{C}_x | \boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}\} > 0 \,;\, \forall \boldsymbol{x} \in \mathcal{B}$

An interpretation of a regular restricted linear partition, is as follows. The linear part of the partition, $\boldsymbol{R}_x Z + \boldsymbol{r}_x$, constitutes a non stationary d - dimensional Gaussian process on $\mathcal{X}$. The restriction, $\boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}$, in the partition

are the zero crossings of this process. The level crossings of $\boldsymbol{R}_x Z + \boldsymbol{r}_x$ is a point process due to the match of the dimension of the parameter space and the value space of the process. For one specific realization z each point is also marked with $\boldsymbol{I}\{z \in \mathcal{C}_x\}$, the indicator of $\mathcal{C}_x$ being one if $z \in \mathcal{C}_x$, zero otherwise. This mark determines the value of x uniquely among the points of the point process, hence determines the part $\mathcal{A}_x$, z belongs to.

Example 2 The meteorologist continued
Assume that the maximum value does not occur at the boundary, in current case this possibility is ruled out by the value of the observations. The global maximum and minimum during the 24 hour period are hence obtained in the interior of the region at critical points. Since z is twice continuously differentiable, the derivative of z is zero at critical points, hence the parameters of the partition are :

$$\boldsymbol{R}_{t_{\max},t_{\min}} z = [z'(t_{\max}), z'(t_{\min})] \ , \ \boldsymbol{r}_{t_{\max},t_{\min}} = \mathbf{0}$$

$$\mathcal{C}_{t_{\max},t_{\min}} = \mathcal{A}_{t_{\max},t_{\min}}$$

The partition is regular with probability one. Criterion i) is fulfilled if $t_{\max} \neq t_{\min}$, this possibility is ruled out by the value of the observations. Criterion ii) and iii) are fulfilled since:

$$\boldsymbol{\nabla}_{t_{\max},t_{\min}} \boldsymbol{R}_{t_{\max},t_{\min}} z = \begin{bmatrix} z''(t_{\max}) & 0 \\ 0 & z''(t_{\min}) \end{bmatrix}$$

Criterion iv) is fulfilled by the definition of $\boldsymbol{r}_{t_{\max},t_{\min}}$. Criterion v) is fulfilled since z is twice continuous differentiable. The number of critical points within the domain will be finite with probability one. By assumption one of these points are the global maximum and one the global minimum point hence criterion vi) is also fulfilled. □

Theorem 2 (Continuous partition) *Let Z be a Gaussian random field on $\mathcal{Z}$ with distribution $\mathcal{D}\{Z\}$, such that $P\{Z \in \mathcal{Z}\} = 1$. Further let $\boldsymbol{\omega} : \mathcal{Z} \to \mathbf{R}^r$ be a $\mathcal{D}\{Z\}$-measurable piecewise affine operator with index set $\mathcal{X} \subset \mathbf{R}^d$, $0 < Leb(\mathcal{X}) < \infty$ and parameters $\{\mathcal{A}_x, \boldsymbol{B}_x, \boldsymbol{b}_x\}_{x\in\mathcal{X}}$ having affine operators $\{\boldsymbol{\omega}_x\}_{x\in\mathcal{X}}$,and a regular restricted linear partition, with parameters $\{\boldsymbol{R}_x, \boldsymbol{r}_x, \mathcal{C}_x\}_{x\in\mathcal{X}}$. Further let $\boldsymbol{O}$, $\{\boldsymbol{O}_x\}_{x\in\mathcal{X}}$ be as defined by Expressions (2)*

and (3) above, with $\boldsymbol{B}_x$ *being index continuous; and* $\boldsymbol{b}_x, \boldsymbol{\Sigma_\varepsilon}^x$ *being continuously dependent on* x*, and that* $rank\{Cov\{\boldsymbol{\nabla}_x \boldsymbol{R}_x Z\}\} = n$*. Assume further:*

$$
\begin{aligned}
&rank\{Cov\{[\boldsymbol{O}_x, \boldsymbol{R}_x Z, \boldsymbol{\nabla}_x \boldsymbol{R}_x Z]\}\} = r + d + n\,; \\
&\exists \mathcal{B} \times \mathcal{R} \subset \mathbf{R}^d \times \mathbf{R}^n \text{ with } Leb(\mathcal{B} \times \mathcal{R}) > 0: \quad \forall (x, \boldsymbol{r}) \in \mathcal{B} \times \mathcal{R} \\
&\qquad P\{Z \in \mathcal{C}_x | \boldsymbol{O}_x = \boldsymbol{o}, \boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}, \boldsymbol{\nabla}_x (\boldsymbol{R}_x Z + \boldsymbol{r}_x) = \boldsymbol{r}\} > 0
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\mathcal{D}\{Z | \boldsymbol{O} = \boldsymbol{o}\} = \\
&\int_{x \in \mathcal{X}} \int_{\mathbf{R}^n} \mathcal{D}\{Z | \boldsymbol{O}_x = \boldsymbol{o}, \boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}, \boldsymbol{\nabla}_x (\boldsymbol{R}_x Z + \boldsymbol{r}_x) = \boldsymbol{r}, Z \in \mathcal{C}_x\} \\
&\qquad \times P\{Z \in \mathcal{C}_x | \boldsymbol{O}_x = \boldsymbol{o}, \boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}, \boldsymbol{\nabla}_x (\boldsymbol{R}_x Z + \boldsymbol{r}_x) = \boldsymbol{r}\} \qquad (7) \\
&\qquad \times \frac{|\det(\boldsymbol{r})| \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}, \boldsymbol{R}_x Z + \boldsymbol{r}_x = \mathbf{0}, \boldsymbol{\nabla}_x (\boldsymbol{R}_x Z + \boldsymbol{r}_x) = \boldsymbol{r}\}}{\mathcal{D}\{\boldsymbol{O} = \boldsymbol{o}\}} \, dx \, d\tilde{\boldsymbol{r}}
\end{aligned}
$$

The first term in Expression (7) is a truncated Gaussian distribution. The second term is the acceptance criterion and the third term is the proposal distribution on $\mathcal{X} \times \mathbf{R}^n$. Note that there are only linear equality constraints in the first term and that the second and third term is the posterior density for being in $\mathcal{A}_x$ with $\boldsymbol{\nabla}_x(\boldsymbol{R}_x Z - \boldsymbol{r}_x)$ having the value $\boldsymbol{r}$. The random variable $\boldsymbol{\nabla}_x(\boldsymbol{R}_x Z - \boldsymbol{r}_x)$ has a Gaussian distribution in $\mathbf{R}^{d^2}$, but the distribution lives on a n-dimensional hyper plane in $\mathbf{R}^{d^2}$. In Expression (7), the variable $\tilde{\boldsymbol{r}}$ is introduced to denote integration along this hyper plane. For a given problem the relation between $\boldsymbol{r}$ and $\tilde{\boldsymbol{r}}$ is evident.

In comparison to the finite index case, Expression (7) contains a Jacobian. Due to this Jacobian, the mixing distribution must be extended. Instead of just drawing the region $\mathcal{A}_x$ containing z, the value of the gradient of the linear part of the restriction, $\boldsymbol{R}_x Z + \boldsymbol{r}_x$, must be sampled simultaneously. The appearance of a Jacobian when turning from the case of a discrete index, to the case of a continuous index, is similar to what appends for transforms of random variables. The proof of Theorem 2 is left to Appendix A.

The mixing distribution of Theorem 2, provides a sampling strategy for the posterior.

Algorithm 2 *CTGM-algorithm (Continuous Truncated Gaussian Mixing)*

- *Sample* $x^*, \tilde{\boldsymbol{r}}^* \sim q(x, \tilde{\boldsymbol{r}})$

$$q(x, \tilde{\boldsymbol{r}}) \propto |\det(\boldsymbol{r})| \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}, \boldsymbol{R}_x Z + \boldsymbol{r}_x = \boldsymbol{0}, \boldsymbol{\nabla}_x(\boldsymbol{R}_x Z + \boldsymbol{r}_x) = \boldsymbol{r}\}$$

- *Sample* $Z^* \sim \mathcal{D}\{Z | \boldsymbol{O}_{x^*} = \boldsymbol{o}, \boldsymbol{R}_{x^*} Z + \boldsymbol{r}_{x^*} = \boldsymbol{0}, \boldsymbol{\nabla}_{x^*}(\boldsymbol{R}_{x^*} Z + \boldsymbol{r}_{x^*}) = \boldsymbol{r}^*\}$
- *If* $Z^* \in \mathcal{C}_{x^*}$ *stop.*

The algorithm is a variant of rejection sampling, and works exactly as for the case of a finite index. The only difference is that the index has been extended. The acceptance rate of the algorithm can be calculated as

$$p_{accept} = \frac{\mathcal{D}\{\boldsymbol{O} = \boldsymbol{o}\}}{\int_{x \in \mathcal{X}} \int_{\mathbf{R}^n} |\det(\boldsymbol{r})| \mathcal{D}\{\boldsymbol{O}_x = \boldsymbol{o}, \boldsymbol{R}_x Z - \boldsymbol{r}_x = \boldsymbol{0}, \boldsymbol{\nabla}_x(\boldsymbol{R}_x Z - \boldsymbol{r}_x) = \boldsymbol{r}\}\, dx\, d\tilde{\boldsymbol{r}}}$$

The difference from the finite index case, is that the normalizing constant of $q(x, \tilde{\boldsymbol{r}})$ is unknown, hence the proposal distribution for the mixing, must be sampled by the use of general purpose algorithms. Since the CTGM-algorithm requires use of a general purpose algorithm to sample $q(x, \tilde{\boldsymbol{r}})$, the use of a general purpose algorithm to sample Z directly, should be considered an alternative. The CTGM-algorithm compare favorable in two respects: firstly the dimension of $(x, \tilde{\boldsymbol{r}})$ is finite, whereas the dimension of z is infinite; secondly under the assumptions of Theorem 2, $q(x, \tilde{\boldsymbol{r}})$ will be nonsingular even if the observations are exact, for this case the direct use of a general purpose algorithm fail, and provides no alternative.

Example 2. The meteorologist continued
The CTGM-algorithm was implemented for this case. The proposal distribution q was sampled using a SIR-algorithm (Rubin 1988) using 500000 proposals. Figure 4 show a contour plot of the marginal proposal distribution of location of the maximum and minimum value. Figure 5 show 400 independent samples from the SIR approximation to the posterior. The true curve is plotted in white. The acceptance rate in step 3 of the CTGM-algorithm was observed to be 43.8% in this example. The correct distribution is sampled within the discretisation effect of the SIR-algorithm, even for this case where the observations are exact. The naive use of any general purpose algorithm fail for this case, due to the exactness of the observations. □

4 Event migration of reflection seismic

In this section a geological horizon is predicted based on traveltimes from reflection seismic.

The geological model is assumed to have a constant velocity above a reflector, Figure 6 illustrates the situation. A 10 km large region is considered. Zero offset traveltimes are recorded every 100 m in the middle 8 km of the region. The vertical lines in Figure 6 are the ray paths connecting the shot/receiver location on the surface with the reflecting point on the horizon. From Figure 6 one can see that the valleys of the horizon are under sampled. For a given shot/receiver location, the two way reflection time is a nonlinear multi valued functional of the geological horizon. There will always be at least one reflection to each shot/receiver location, being the minimum distance from the travel time to the reflector, only this first arrival is considered.

The travel time measured at location x_s is

$$B_x z = t_x(z) = \frac{2}{v}\sqrt{z(x)^2 + (x - x_s)^2}\,, \tag{8}$$

with v being the velocity in the top layer; z being the vertical component of the horizon; and x being the location of the reflecting point on the horizon. At the reflecting point on the horizon, $(x, z(x))$, the horizon is tangent to the sphere centered at x_s trough the point. This property is described by the relation:

$$R_x z + r_x = z(x) \cdot z'(x) + x - x_s = 0\,. \tag{9}$$

In addition only the first arrival is considered, hence

$$\mathcal{C}_x = \{z \in \mathcal{Z} : z(x)^2 + (x - x_s)^2 < z(u)^2 + (u - x_s)^2;\, x, u \in \mathbf{R}^1\,, u \neq x\} \tag{10}$$

The parts of the piecewise affine operator are identified by Expressions (8), (9) and (10) above, but the example does not automatically conform to the framework of the piecewise affine operators since B_x, R_x are nonlinear operators. However since $Z(x)$, $Z'(x)$, and $Z''(x)$ can be solved exactly when $B_x Z = t$, $R_x Z + r_x = 0$ and $\nabla_x(R_x Z + r_x) = r$ is known, the only adjustment in the CTGM-algorithm is for sampling x and $\tilde{\boldsymbol{r}}$. The exact details for this is in Appendix B.

When having observations of the first reflection time at several shot/receiver locations, the ordering of the reflection locations are strictly monotone as a function of the shot/receiver location. Criterion i) of Definition 5 is hence fulfilled. The remaining conditions may also be controlled to assure the partition to be regular. Both uncertainty in the shot/receiver location, and observation error in the traveltime may be introduced in the model. To correctly account for these effects, these parameters should be sampled simultaneously with x and r.

The prior model for the reflector is defined to be a stationary Gaussian random field,

$$Z(x) = Z_0 + \widetilde{Z}(x) \quad ; \quad x \in [0,\, 10];$$

with $\mathcal{D}\{Z_0\} = N(2,1)$; and $\widetilde{Z}(x)$ being a zero mean Gaussian random field with covariance $\mathrm{Cov}\{\widetilde{Z}(x), \widetilde{Z}(x+h)\} = 0.15^2 \exp\{-3 \cdot (h/2)^2\}$. Several samples from the prior distribution are shown in Figure 7. The realization used in this example is plotted with a thick line.

The uncertainty in shot location, x_s, is assigned standard deviation $\sigma_{x_s} = 10$ m. The observation error of the traveltime is assigned standard deviation $\sigma_t = 0.05$ sec. The interval velocity is assumed to be $v = 1$ km/sec. In Figure 8 the observed travel times and assumed shot locations are plotted together with the true traveltimes.

Figure 9 and 10 displays the results of a simulation study done by using the CTGM-algorithm. The samples in Figure 9 was obtained by using only the five marked observations in Figure 8. The proposal algorithm was sampled by use of a Markov chain based algorithm. The acceptance rate in the final step of the algorithm, was 98%, leaving the sampling of q as the most time consuming part. The samples in Figure 10 was obtained by using all 81 observations displayed in Figure 8. The proposal distribution was sampled by use of lineraization hence the samples are approximate. However since the proposal distribution can be fully specified, and the actual density is known to proportionality, the algorithm could be put into the frame work of Markov chains or importance sampling to asymptotically remove bias in the samples. The observed acceptance rate was 4.95 % in the final step of the CTGM-algorithm. The relatively low acceptance rate is partially due to numerical instability in this particular problem having a large number of observations.

5 Conclusions

A Bayesian approach to solve piecewise affine inverse problems, has been introduced. Piecewise affine inverse problems have an intuitive definition and are easy to picture mentally. Although piecewise affine inverse problems, only are a small step into the world of nonlinearity, they possess genuine nonlinear features. There are several problems that may be posed as piecewise affine inverse problems. In particular inverse problems having a variational structure, such as the Fermant principle of traveltime tomography possess this structure. Event migration of traveltimes from reflection seismic is also of this type and is used as as an example in this article.

The posterior for piecewise affine inverse problem with a finite and a continuous index has been shown to be a mixture of truncated Gaussian distributions. There are several ways to exploit this decomposition when sampling the posterior. Compared with other general purpose algorithms, the real benefit using this decomposition is expected to be when the observations have high precision.

An algorithm that uses the decomposition and is based on rejection sampling has been developed. For the discrete index case, the algorithm requires to sample a known finite set of point probabilities and to sample a Gaussian random field conditioned to a linear constraint. For the continuous index case, the algorithm requires sampling from a density that is known up to a normalizing constant, and to sample a Gaussian random field conditioned to a linear constraint. When tested for example problems, the algorithm gives reasonable acceptance rates, and solves problems that can not be solved by direct use of general purpose sampling algorithms.

A Proof of theorem 2

To simplify notation, define the random variable X implicitly by $Z \in \mathcal{A}_{X(Z)}$, and the random vector fields $\boldsymbol{R}(x) = \boldsymbol{R}_x Z + \boldsymbol{r}_x$, $\boldsymbol{O}(x) = \boldsymbol{O}_x$; $x \in \mathcal{X}$. In the current notation the content of Theorem 2 is restated as:

$$\mathcal{D}\{X = x, \boldsymbol{O} = \boldsymbol{o}\} = \int_{\mathbf{R}^n} |det(\boldsymbol{r})| P\{Z \in \mathcal{C}_x | \boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{\nabla}_x \boldsymbol{R}(x) = \boldsymbol{r}, \boldsymbol{O}(x) = \boldsymbol{o}\} \times \mathcal{D}\{\boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{\nabla}_x \boldsymbol{R}(x) = \boldsymbol{r}, \boldsymbol{O}(x) = \boldsymbol{o}\}\, d\tilde{\boldsymbol{r}} \,.$$

Proof :
Consider the following identity

$$P\{X \in \mathcal{B} \cap \boldsymbol{O} \in \Omega\} = P\{\exists x \in \mathcal{B} : \boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{O}(x) \in \Omega, Z \in \mathcal{C}_x\}$$

Define a random counting process by

$$N_\Omega(\mathcal{B}) = \#\{x \in \mathcal{B} : \boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{O}(x) \in \Omega, Z \in \mathcal{C}_x\}$$

Note that $N_\Omega(\mathcal{B}) \in \{0, 1\}$ since $\{\boldsymbol{R}(x) = \mathbf{0}\} \cap \mathcal{C}_x$ constitutes a partition, hence:

$$\mathrm{E}\{N_\Omega(\mathcal{B})\} = P\{\exists x \in \mathcal{B} : \boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{O}(x) \in \Omega, Z \in \mathcal{C}_x\}$$

This expectation can be obtained under suitable regularity conditions for the random fields involved (Adler 1981). First note that

$$N_\Omega(\mathcal{B}) = \lim_{\varepsilon \to 0} \int_{\mathcal{B}} \delta_\varepsilon(\boldsymbol{R}(x)) I(Z \in \mathcal{C}_x \cap \boldsymbol{O}(x) \in \Omega) |\det(\boldsymbol{\nabla}_x \boldsymbol{R}(x))|\, dx \,, \quad (11)$$

with δ_ε being a delta sequence; $I(\cdot)$ being the indicator function and $\boldsymbol{\nabla}_x \boldsymbol{R}(x)$ being the Jacobian of $\boldsymbol{R}(x)$. The result is obtained by integrating the delta sequence in the value set of $\boldsymbol{R}(x)$ and then flip the variable of integration to the definition set, resulting in a local Jacobian. The zero crossings of $\boldsymbol{R}(x)$ is a point process on $\mathcal{X}$. Each point in the point process, is marked with $I(Z \in \mathcal{C}_x \cap \boldsymbol{O}(x) \in \Omega)$, $N_\Omega(\mathcal{B})$ count only those points for which this mark have the value 1, hence $N_\Omega(\mathcal{B})$ is a thinning of the point process being the zero crossings of $\boldsymbol{R}(x)$. The expectation of Expression (11) is obtained by integrating the limit of the expected value of the integrand, the interchange

of limit, integration and expectation is valid under the regularity conditions of Theorem 2.

$$P\{N_\Omega(\mathcal{B}) = 1\} = \int_{\mathcal{B}} E\left\{|\det(\boldsymbol{\nabla}_x \boldsymbol{R}(x))| \cdot I(Z \in \mathcal{C}_x \cap \boldsymbol{O}(x) \in \Omega) \,|\boldsymbol{R}(x) = \mathbf{0}\right\} \times \mathcal{D}\{\boldsymbol{R}(x) = \mathbf{0}\}\, dx$$

The final result is obtained by using the sentence of double expectation twice to get

$$P\{N_\Omega(\mathcal{B}) = 1\} = \int_{\mathcal{B}}\int_{\Omega}\int_{\mathbf{R}^n} |det(\boldsymbol{r})| \cdot E\left\{I(Z \in \mathcal{C}_x)\,|\boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{O}(x) = \boldsymbol{o}, \boldsymbol{\nabla}_x \boldsymbol{R}(x) = \boldsymbol{r}\right\} \times \mathcal{D}\{\boldsymbol{R}(x) = \mathbf{0}, \boldsymbol{O}(x) = \boldsymbol{o}, \boldsymbol{\nabla}_x \boldsymbol{R}(x) = \boldsymbol{r}\}\; d\tilde{\boldsymbol{r}}\, d\boldsymbol{o}\, dx$$

QED.

Note that the dimension of the mixing distribution in the general case is d^2+d, but for most problems there will be many zeros off the diagonal in the Jacobian. The effective dimension of the mixing distribution will therefore usually be of order d.

B Details of section 4

This appendix contain detailed calculations to fit traveltime observations from reflection seismic into the frame work piecewise affine operators. Firstly assume that the observed traveltimes do not have any observation error. The mixing distribution to be sampled in the CTGM-algorithm is then,

$$q(x, r) = \text{const} \times |r| \mathcal{D}\{B_x Z = t, R_x Z + r_x = 0, \nabla_x (R_x Z + r_x) = r\},$$

with $B_x Z$, $R_x Z + r_x$ and $\nabla_x (R_x Z + r_x)$ referring to Expressions (8) and (9). Rename the random variables ($B_x Z$, $R_x Z + r_x$, $\nabla_x (R_x Z + r_x)$) by (T, R_0, R_1) and the random variables ($Z(x)$, $Z'(x)$, $Z''(x)$) by (Z_0, Z_1, Z_2) for a given x. The relation between (T, R_0, R_1) and (Z_0, Z_1, Z_2) is given by:

$$\begin{aligned} T &= \frac{2}{v}\sqrt{Z_0^2 + (x - x_s)^2}\,, \\ R_0 &= Z_0 \cdot Z_1 + x - x_s\,, \\ R_1 &= Z_0 \cdot Z_2 + Z_1^2 + 1\,. \end{aligned} \tag{12}$$

This relation can be inverted to find :

$$\begin{aligned} Z_0 &= \sqrt{\left(\frac{Tv}{2}\right)^2 - (x - x_s)^2}, \\ Z_1 &= \frac{x_s - x - R_0}{Z_0}, \\ Z_2 &= \frac{R_1 - Z_1^2 - 1}{Z_0}. \end{aligned} \tag{13}$$

The variables (Z_0, Z_1, Z_2), have a known Gaussian distribution, hence the distribution of (T, R_0, R_1) may be calculated by the usual transformation rule. Assuming $(T, R_0, R_1) = (t, 0, r)$, the outcome of (Z_0, Z_1, Z_2) is given by the following expressions:

$$\begin{aligned} z_0(t, x, x_s) &= \sqrt{\left(\frac{tv}{2}\right)^2 - (x - x_s)^2}, \\ z_1(t, x, x_s) &= \frac{x_s - x}{z_0(t, x, x_s)}, \\ z_2(t, x, x_s, r) &= \frac{r - z_1(t, x, x_s)^2 - 1}{z_0(t, x, x_s)}. \end{aligned}$$

Using the transformation rule yields the result

$$q(x, r; x_s, t) = \text{const} \times \left| \frac{r\frac{tv}{2}}{z_0(t, x, x_s)^3} \right|$$

$$\times \mathcal{D}\{Z(x) = z_0(t, x, x_s), Z'(x) = z_1(t, x, x_s), Z''(x) = z_2(t, x, x_s, r)\},$$

with x_s and t regarded as given parameters; and the joint probability distribution of $(Z(x),\ Z'(x),\ Z''(x)\,)$ being known from the prior.

When including observation error in traveltime and shot location, the mixing distribution must be extended to include sampling of the true traveltime and the true shot location simultaneously. Hence by denoting the observed traveltime o, the task is to sample the distribution

$$\tilde{q}(x, r, t, x_s; o) = q(x, r; x_s, t)\mathcal{D}\{X_s = x_s\}\mathcal{D}\{\varepsilon = o - t\},$$

with; $\mathcal{D}\{\varepsilon = o - t\}$ being the likelihood of the observation; and $\mathcal{D}\{X_s = x_s\}$ being a prior distribution for the shot location. The calculations above are

made for the case of one observation, but the extension to several observations is obvious, since the transforms apply locally to each set of variables corresponding to one observation.

The equivalence between sampling $\mathcal{D}\{Z|Z(x), Z'(x), Z''(x)\}$ and $\mathcal{D}\{Z|B_xZ, R_xZ+r_x, \nabla_x(R_xZ+r_x)\}$, is due to the one to one relation between $(Z(x), Z'(x), Z''(x))$ and $(B_xZ, R_xZ+r_x, \nabla_x(R_xZ+r_x))$, given by Expression (12) and (13). It is not equivalent to sample $\mathcal{D}\{Z|Z(x), Z'(x), Z''(x)\}$ and $\mathcal{D}\{Z|O_x, R_xZ + r_x, \nabla_x(R_xZ + r_x)\}$, since an exact relationship is needed for the equivalence to hold.

References

Abramovich, F., and Silverman B.W. (1998), "Wavelet decomposition approaches to statistical inverse problems," *Biometrika*, 85, 115-129.

Adler, R.J. (1981), "The geometry of random fields"; Wiley.

Berryman, J.G.(1997), "Variational structure of inverse problems in wave propagation and vibration," in *Inverse Problems in Wave Propagation*, ed. G. Chavent, G. Papanicolaou, P. Sacks, and W. W. Symes; New York; Springer; pp. 13-44.

Donoho, D.L. (1992), "Nonlinear solution of linear inverse problems by wavelet - vaguelet decomposition," *Technical report No. 403*; Stanford University.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.B. (1995), "Bayesian data analysis"; London; Chapman & Hall.

Hastings, W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97-109.

Johnstone, I. M. (1999), "Wavelet shrinkage for Correlated data and inverse problems: adaptivity results," *Statistica Sinca*, 1, 51- 83.

Kolessa A.E. (1986), "Recursive filtering algorithms for systems with piecewise linear non-linearities",*Avtomat.Telemekh.*, 4, 48-55.

Lehmann, E.L. (1999), "Elements of large-sample theory"; New York; Springer.

Liu J.S., and Chen,R. (1998), "Sequential Monte Carlo Methods for Dynami-

cal Systems," *Journal of the American Statistical Association*, 93, 1032-1044.

O'Sullivan, F. (1986), "A statistical perspective on ill-posed inverse problems," *Statistical Science*, 1, 502-518.

von Neumann, J. (1951), "Various techniques in connection with random digits," *NBS Appl.Math Ser.*, 12, 36-38.

Pinsker, M. (1980), "Optimal filtering of square integrable signals in Gaussian white noise," *Problems of Information Transmission* 16, pp. 120-133. Originally in Russian in, *Problemy Peredatsii Informatsii*, 16, pp. 52-68.

Rubin, D.B. (1988), "Using SIR algorithm to simulate posterior distributions," in *Bayesian statistics 3* ed. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith; New York: Oxford University Press; pp. 395-402.

Tarantola, A. (1987), "Inverse problem Theory"; Elsevier.

Tikhonov, A. (1963), "Solution to incorrectly formulated problems and the regularization method," *Soviet Math. Doklady*, 5, 1035-1038.

Vanmarcke, E. (1983), "Random fields"; The MIT press.

Wahba, G. (1990), "Spline models for Observational data", *NSF-CBMS Regional Conference in Mathematics*, (Vol. 59); Philadelphia; SIAM.

Tables and figures

Table 1: Parameters of the piecewise affine operator in Example 1.

x	$\mathcal{A}_x$	$\boldsymbol{B}_x$	b_x
1	$\{z_1 \geq 0\ ;\ z_2 \geq 0\}$	$[\ 1,\ \ 1]'$	0
2	$\{z_1 \geq 0\ ;\ z_2 < 0\}$	$[\ 1, -1]'$	0
3	$\{z_1 < 0\ ;\ z_2 \geq 0\}$	$[-1,\ \ 1]'$	0
4	$\{z_1 < 0\ ;\ z_2 < 0\}$	$[-1, -1]'$	0

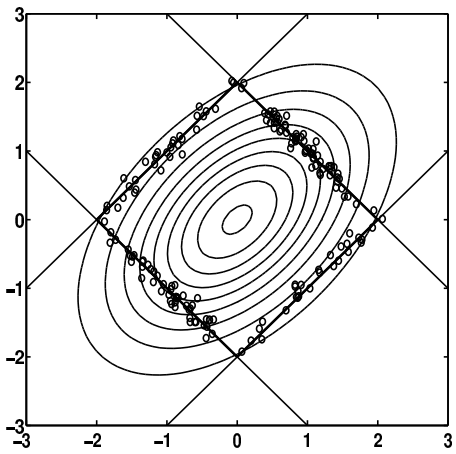


Figure 1: Conditioning to the L^1 norm. A scatter plot containing 177 values sampled from the posterior by the use of FTGM-algorithm. Superimposed on this is the contour lines from prior distribution; a solid square showing the contour of the piecewise affine operator at the observed level, $\|\boldsymbol{z}\|_1 = 2$; and dashed lines showing the extension of the affine operators at the same level.

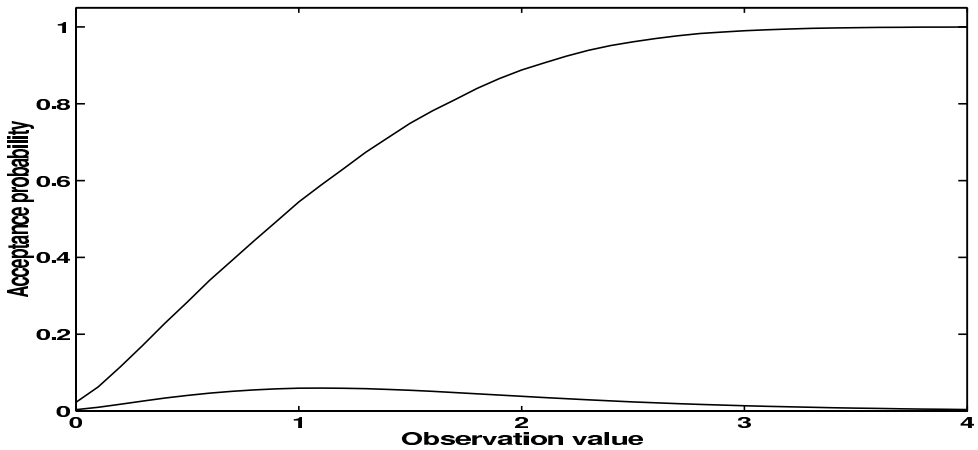


Figure 2: Comparison to naive rejection sampling. The acceptance rate is plotted as a function of the observed value in Example 1. The solid line is the accept probability for the FTGM-algorithm, the dotted line is the corresponding for naive rejection sampling.

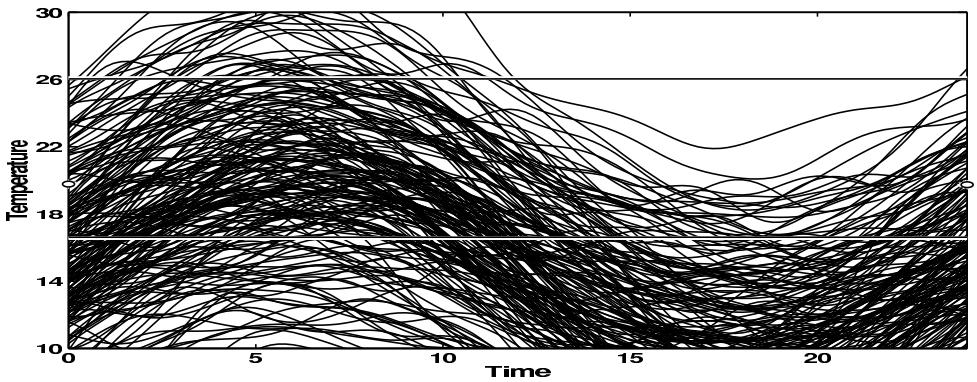


Figure 3: Conditioning to extreme value observations; observations and prior. The horizontal lines shows the observed level of maximum and minimum values. The white circles at each end, show the values observed at the boundaries, in addition 200 samples from the prior distribution used in Example 2, is displayed, some of the samples are partially or completely outside the scaling of the figure.

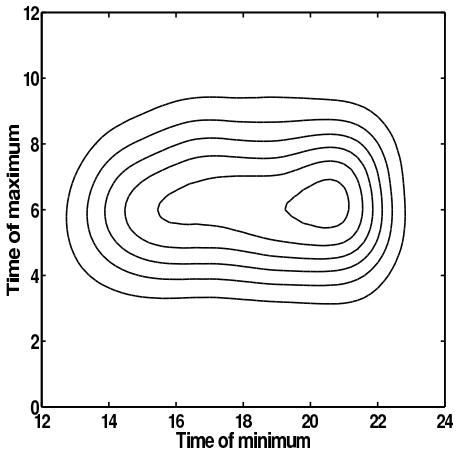


Figure 4: Conditioning to extreme value observations; the mixing distribution. The CTGM-algorithm proposes the time for the maximum temperature and minimum temperature according to the displayed distribution, in addition two more parameters are sampled. The displayed distribution is the estimated marginal distribution $q(t_{\max}, t_{\min})$, based on the SIR-approximation of the full mixing distribution in Example 2.

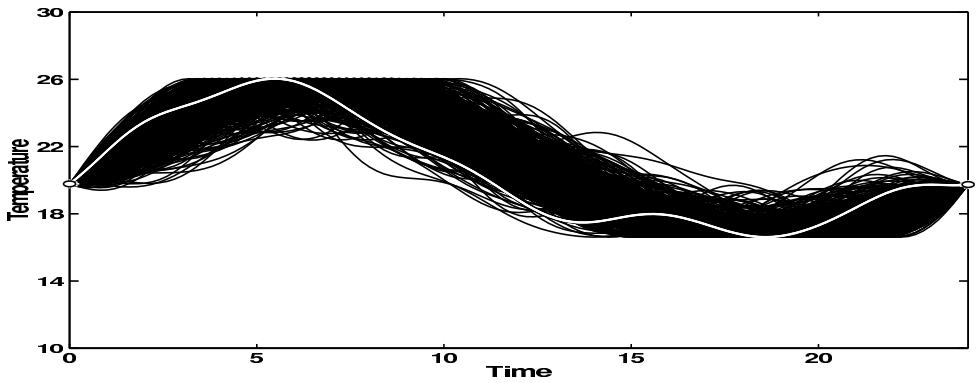


Figure 5: Conditioning to extreme value observations; the posterior. The actual temperature is displayed as a white curve, together with 400 realizations from the posterior distribution, sampled by the CTGM-algorithm in Example 2.

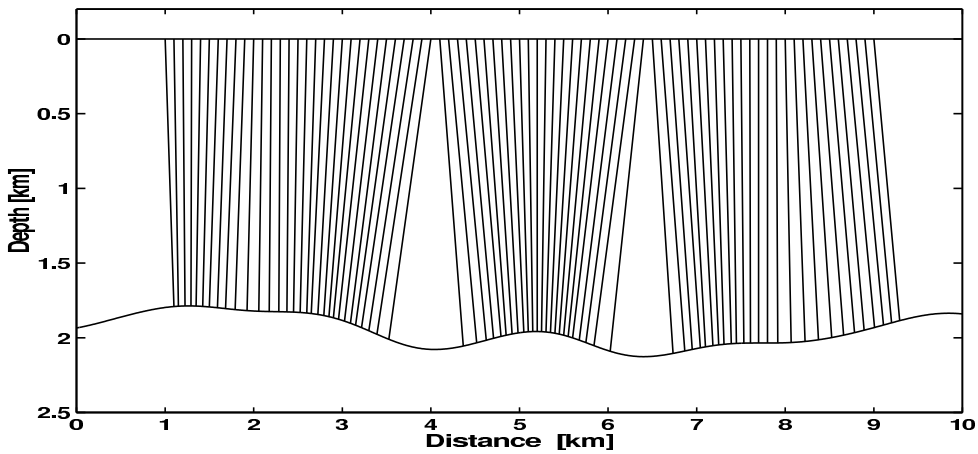


Figure 6: Traveltime geometry. The upper straight horizontal line is the surface, the geological horizon to be recovered in the example of section 4 is the lower curved horizontal line. The vertical lines are the ray paths connecting the shot location at the surface to to the nearest point at the geological horizon.

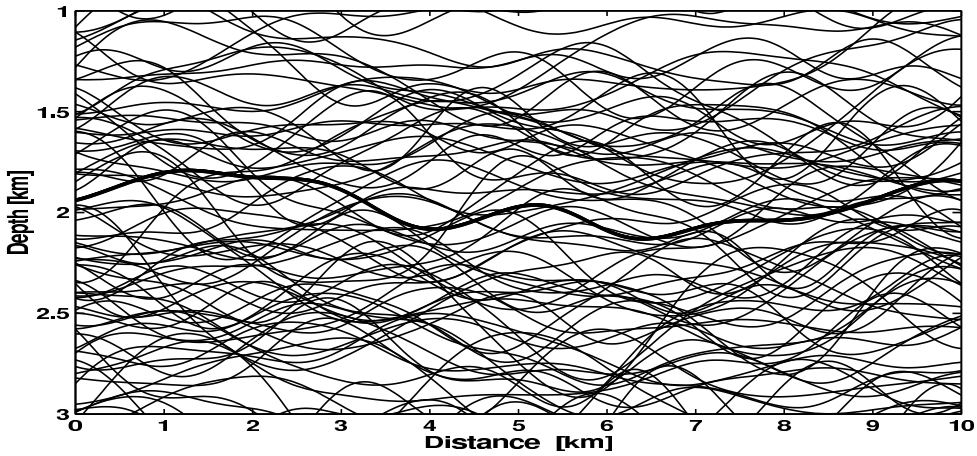


Figure 7: Traveltime migration; the prior. The actual curve is shown with a thick black line together with 100 samples from the prior model for the geological horizon of Section 4, several of the samples are partially or completely outside the scales of the figure.

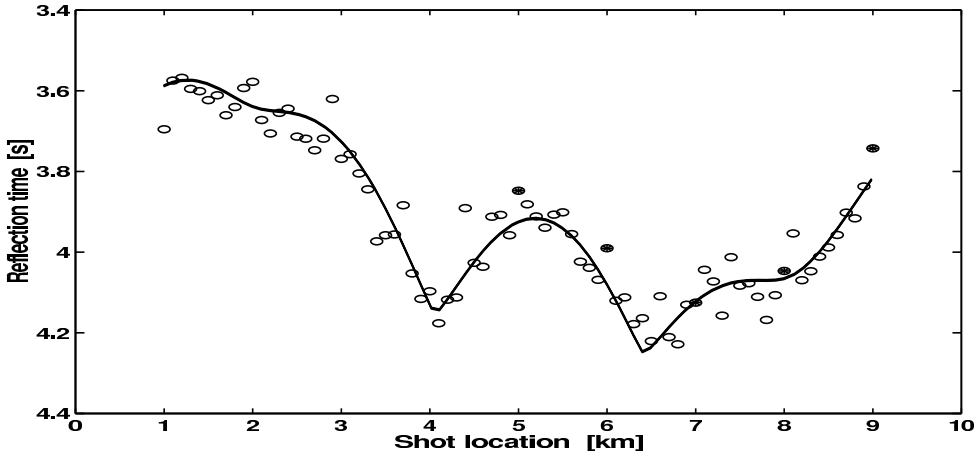


Figure 8: Traveltime migration; the observations. The observed travel times are plotted as circles together with the true travel times as a solid line. The observations have errors in both x_s and t. The filled circles is the subset of observations used when only five observations are considered.

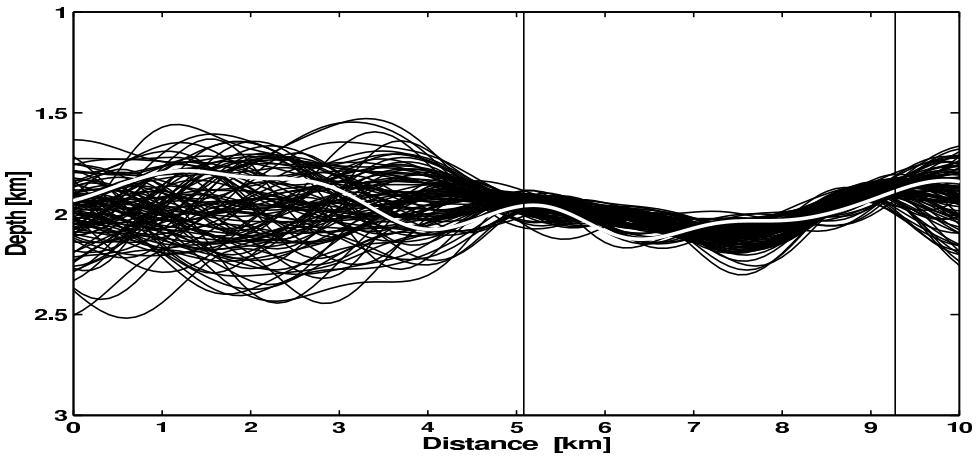


Figure 9: Traveltime migration; the posterior. The actual horizon is displayed in whit together with 99 samples from the posterior, using only the five observations marked in Figure 8. The vertical lines indicates the region with reflections.

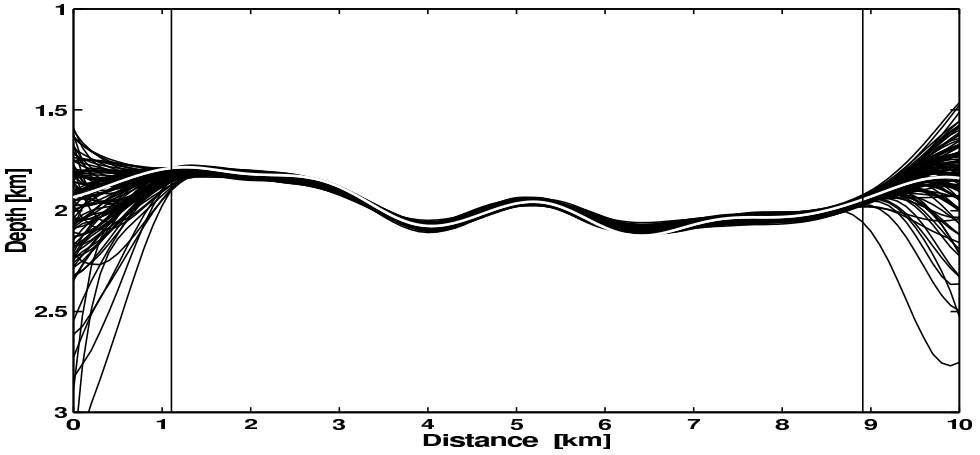


Figure 10: Traveltime migration; the posterior. The actual horizon is displayed in whit together with 99 samples from the posterior, using all 81 observations in Figure 8. The vertical lines indicates the region with reflections.