Case specific uncertainty assessment in cross well tomography

Odd Kolbjørnsen

Department of mathematical sciences,
Norwegian university of science and technology
Trondheim, Norway

Abstract

The inverse problem in cross well tomography is solved by a Bayesian methodology in a Gaussian framework. A finite element approach is used to resolve the variational structure given by Fermat's principle, as a result the approximate forward map is piecewise affine. In the Gaussian framework the posterior distribution can be calculated as a mixture of truncated Gaussian distributions. A sampling algorithm that exploit this structure is proposed. The methodology is tested in a small synthetic example.

KEY WORDS: *Bayesian statistics, Sampling based inference, piecewise affine inverse problem, nonlinear traveltime tomography, Fermat's principle.*

1 Introduction

Cross well tomography is an important source of information about elastic parameters of the earth. Both the direct problem of wave propagation (Langan, Lerche and Cutler 1985; Vidale 1988; Auld 1990) and the inverse problem in cross well tomography (Menke 1984; Berryman 1990; Langan and Bube 1998) are subject to substantial research interest.

The direct problem in cross well tomography is nonlinear, the solution is given by the minimum of a set of linear functionals. In linearized cross well tomography the solution to the direct problem is approximated by picking one of the linear functionals in the set.

The primary goal in cross well tomography is to stably estimate material parameters of the earth based on traveltime observations. Linearized cross well tomography gives a qualitative understanding of the problem. Menke (1984) show that linearized cross well tomography resolves the material parameters of an isotropic earth poorly, especially in the horizontal direction. The case for anisotropic case is even worse (Bube and Meadows, 1998). Further, related operators such as X-ray and Radon transforms have unbounded inverses (Faridani,

1997). The inverse problem of cross well tomography is hence ill-posed, since it is both underdetermined and unstable.

Ill-posed inverse problems is commonly solved by regularization or equivalently by introducing rigid boundaries on the parameter space (Bertero, 1989). The solution is then obtained by minimizing an objective function. For nonlinear problems such as cross well tomography, iterative solvers are frequently used. Berryman (1990) notice that this formulation does not fully appreciate the variational structure that is present in the problem of cross well tomography. The first arrival time obeys Fermat's principle, i.e. it is the shortest traveltime that is physically possible (Aronsson 1970). Berryman (1990) uses Fermat's principle to construct feasibility constraints for the solution. When he uses the feasibility constraints to determine the step size in his solver he obtain a stable reconstruction.

A secondary and frequently equally important objective in cross well tomography is to assess the uncertainty of the estimate. Common approaches are resolution theory (Menke, 1984) and a singular value decomposition (Michelena 1993), in either case the operator is linearized. In nonlinear problems such as cross well tomography, it is hard to describe the underdetermined and badly determined features exactly, since they do not span a linear space.

The current work apply a Bayesian approach to the inverse problem of cross well tomography. In Bayesian analysis a likelihood is defined according to the statistical link between the parameter of interest and the observations and a prior distribution is defined for the parameter of interest. The prior distribution is frequently criticized by non-Bayesians. However for ill-posed inverse problems, such as cross well tomography, the prior distribution plays an essential role. The prior distribution stabilizes the solution and resolves the problem of underdetermination. The prior hence serve the purpose of regularization and define soft boundaries on the parameter space. In a specific case there is often available information about the scales of the slowness, either based on general geological knowledge or analog reservoirs. This information can be included through the prior distribution. The effect of the assumptions can be visualized by random samples from the prior distribution.

The Bayesian solution to the inverse problem, is the posterior distribution which is formally proportional to the product of the likelihood and the prior. For most practical problems it is beneficial to approximate the posterior distribution by a finite representation. In the current work the posterior distribution is approximated by random samples assigned equal weight. This approach apply to both linear and nonlinear problems. The Bayesian approach achieve both goals in cross well tomography. The posterior mean is a stable estimate. The posterior distribution itself describes the uncertainty of the estimate. Bayesian uncertainty assessment is hence case specific. The posterior distribution is relative to the observation at hand and depend on the prior distribution and the likelihood which are defined such that their characteristics are adapted to the case under study.

Bayesian approaches to problems in tomography is developed by several authors, (Natterer 1980; Carfantan and Mohammad-Djafari 1997), most authors only consider the maximum posterior estimate and does not use the full power of the Bayesian analysis. In the current work the Bayesian inversion, i.e. an algorithm to sample the posterior, is worked out in a Gaussian framework, taking account of the nonlinear features. In Kolbjørnsen and Omre (2002) the theory of piecewise affine inverse problems in a Gaussian framework is presented. The posterior distribution is a mixture of truncated Gaussian distributions in this case. The contribution in the current work is to use the Fermat's principle to phrase cross well tomography as a piecewise affine inverse problem and develop the methodology of Kolbjørnsen and Omre (2002) for this problem.

Section 2 describes the problem of cross well tomography. In section 3 the problem of cross well tomography is formulated as an piecewise affine inverse problem, by using a finite element approach to approximate Fermat's principle. Section 4 describes the statistical models that are used, and section 5 contains the posterior distribution together with the sampling approach. Section 6 discuss a generalization of the approach. In section 7 two small examples are investigated. Section 8 contain a discussion of the results.

2 Problem description

The current section gives a brief introduction to the problem of cross well tomography. The slowness, the inverse of the velocity, is the material parameter of relevance. In the current presentation the medium is assumed to be isotropic, but the approach can easily be extended to media with elliptical anisotropy (Bube and Meadows, 1998).

The objectives in cross well tomography is to reconstruct the slowness field in a region, $\mathcal{R}$, between two wells based on imperfect observations of traveltimes from sources in one well to receivers in the other well, and to assess the uncertainty of the reconstruction. Figure 1 illustrates the situation. A source is placed in one well at the location (x_s, z_s), a receiver is placed in the other well at the location (x_r, z_r). The traveltime is the time it takes for a wave to propagate from the source to the receiver.

For simplicity the earth is considered to vary only with depth, z, and the lateral component describing the inter distance between the two wells, x, i.e. $s(x, y, z) = s(x, z)$. Further the slowness is assumed to be twice continuously differentiable, i.e. $s \in C^2(\mathcal{R})$.

The traveltime between a source and a receiver is denoted the Fermat time because it obeys Fermat's principle. That is, it is the minimum traveltime from the source to the receiver. To make this precise Berryman (1997) introduce two types of functionals for traveltime. Let Γ be the set of continuous paths connecting the source and the receiver. For a given $\gamma \in \Gamma$ define the traveltime

functional, $\tau(\gamma, \cdot)$, associated with this path by its action on a slowness field, s,

$$\tau(\gamma, s) = \int_{\gamma} s(x, z)\, dl^{\gamma}$$

with dl^{γ} being the infinitesimal distance along γ. Define now the traveltime functional, τ^*, corresponding to the Fermat time. For given slowness field, s, this is defined as,

$$\tau^*(s) = \min_{\gamma \in \Gamma} \tau(\gamma, s) \ . \tag{1}$$

The Fermat time is the minimum path integral of the slowness along any continuous path connecting the source and the receiver. The Fermat path, γ^*, is defined as the path where this minimum occur,

$$\gamma^*(s) = \arg\min_{\gamma \in \Gamma} \tau(\gamma, s).$$

The Fermat path need not be unique, but for a given source/receiver pair it almost surely is so. The Fermat time can be expressed as

$$\tau^*(s) = \int_{\gamma^*(s)} s(x, z)\, dl^{\gamma^*(s)} \ ,$$

that is, if the Fermat path is known the traveltime is a linear functional of s.

In a medium of constant slowness, the Fermat paths are straight lines connecting the source and the receiver. A perturbation argument (Boyse and Keller 1995) show that the bending of the Fermat path is a second order effect, hence the traveltime can be approximated to the first order by the line integral along the straight line connecting the source and the receiver. This is the argument used in linearized cross well tomography to pick a particular path. Figure 2 show a slowness field where the perturbation argument is not valid due to large deviations from a constant background. Figure 2(a) show the linear paths for 16 source/receiver pairs. Figure 2(b) show the Fermat paths for the same slowness field. For such cases other approximations are needed.

3 Cross well tomography as a piecewise affine inverse problem

To phrase cross well tomography as a piecewise affine inverse problem, each traveltime is approximated by a piecewise affine functional. In the current work the Fermat time, $\tau^*(s)$ in Expression (1), is approximated by a finite element approach.

$$\tau_0^*(s) = \min_{\gamma \in \Gamma_0} \tau(\gamma, s) \ , \tag{2}$$

with $\tau_0^*(s)$ being the approximate Fermat time; and Γ_0 being the set of finite elements. The set Γ_0 consist of piecewise linear paths, parameterized with d internal nodes. The nodes are equispaced in the lateral directions and free to move in the vertical direction, see Figure 3. Each path is hence parameterized by a d dimensional parameter, $\gamma = (\gamma_1, \gamma_2, ..., \gamma_d)$, being the vertical coordinate of each node. In what follows there will not be made any notationally distinction between the parameter γ and the piecewise linear path that is associated with it. The path parameter is a vector with d components but it is denoted by a normal type letter to avoid confusion when several traveltimes are considered. Further let $\gamma_0^*(s) \in \Gamma_0$ denote the path where the minimum in Expression (2) occur. The path $\gamma_0^*(s)$ is hence the approximate Fermat path.

Figure 4 and 5 visualize the finite element approximation for the slowness field in Figure 2. Figure 4 show how the traveltime approximation improve with an increasing number of internal nodes for the 16 traveltimes indicated in Figure 2. Figure 5 show how one Fermat path change as the number of internal nodes increase. In this particular case the approximation is good even with a low number of internal nodes.

Note that the finite parameterization of the path does not force any particular parameterization of the slowness, this is in contrast to approaches that use block models and Snell's law for ray bending at the block boundaries. The accuracy of the approximation will of course depend on the slowness field. In the continuous formulation of the problem, paths between different source/receiver pairs can cross one time at most. This ordering is forced also in the discrete problem even if several crossings could occur for this case.

According to Kolbjørnsen and Omre (2002) a piecewise affine operator is defined as

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

$$\boldsymbol{K}(z) = \boldsymbol{K}_x z + \boldsymbol{k}_x \quad for \;\; z \in \mathcal{A}_x \; ; \quad x \in \mathcal{X}$$

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

The finite element approximation to the Fermat times is a piecewise affine operator with a continuous index set, Γ_0. For operators having a continuous index set a special type of partition is treated in Kolbjørnsen and Omre (2002).

Definition 2 (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.

The approximate traveltime, $\tau_0^*(s)$ in Expression (2), can be represented as

$$\tau_0^*(s) = \tau(\gamma, s) \quad for \;\; s \in \mathcal{A}_\gamma \; ; \quad \gamma \in \Gamma_0 \tag{3}$$

with

$$\mathcal{A}_\gamma = \{s \in C^2(\mathcal{R}) : \tau(\gamma, s) \leq \tau(\tilde{\gamma}, s) \;\; for \;\; \tilde{\gamma} \in \Gamma_0\} \;\; ,$$

hence $s \in \mathcal{A}_\gamma \Leftrightarrow \gamma = \gamma_0^*(s)$. That is, γ is the approximate Fermat path of s, using the predefined resolution given by Γ_0. Note further $\mathcal{A}_\gamma \subset \{\nabla_\gamma \tau(\gamma, s) = 0\}$ with $\nabla_\gamma \tau(\gamma, s)$ being the gradient of $\tau(\gamma, s)$ with respect to the path, evaluated for the Fermat path, γ. The operator $\nabla_\gamma \tau : \Gamma_0 \times C^2(\mathcal{R}) \to \mathbf{R}^d$ is linear in the second argument, i.e. slowness, for any value of the first, i.e. path. The partition $\{\mathcal{A}_\gamma\}_{\gamma \in \Gamma_0}$ in Expression (3) is hence a restricted linear partition according to Definition 2. Further the Hessian of the traveltime with respect to the path, $\nabla_\gamma \nabla_\gamma \tau$, is of importance. Note that due to the parameterization, the Hessian is tridiagonal.

For each traveltime there are two functionals, $\tau(\gamma, s)$ and $\tau_0^*(s)$, and two operators $\boldsymbol{g}(\gamma, s) = \nabla_\gamma \tau(\gamma, s)$ and $\boldsymbol{h}(\gamma, s) = \nabla_\gamma \nabla_\gamma \tau(\gamma, s)$, that are of importance. The operators $\boldsymbol{g}(\gamma, s)$ and $\boldsymbol{h}(\gamma, s)$ are both linear in the second argument and produce row vectors and matrices respectively. When r traveltimes are considered, the paths corresponding to each of the traveltimes are collected to form one large index, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, ..., \gamma_r]$, this should not be confused with the parameterization of the individual paths; i.e. $\gamma_i = (\gamma_{i1}, ..., \gamma_{id})$ for $i = 1, .., r$. The traveltime functionals are stacked to form vector valued operators, $\boldsymbol{\tau}(\boldsymbol{\gamma}, s)$ and $\boldsymbol{\tau}_0^*(s)$, and the relevant operators are joined,

$$\boldsymbol{g}(\boldsymbol{\gamma}, s) \;\; = \;\; \left[\begin{array}{cccc} \boldsymbol{g}(\gamma_1, s) & \boldsymbol{g}(\gamma_2, s) & \dots & \boldsymbol{g}(\gamma_r, s) \end{array} \right] \tag{4}$$

$$\boldsymbol{h}(\boldsymbol{\gamma}, s) \;\; = \;\; \left[\begin{array}{cccc} \boldsymbol{h}(\gamma_1, s) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{h}(\gamma_2, s) & & \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{h}(\gamma_r, s) \end{array} \right] \;\; ,$$

further

$$\mathcal{A}_{\boldsymbol{\gamma}} = \bigcap_{i=1}^{r} \mathcal{A}_{\gamma_i} \; .$$

The traveltimes are hence approximated by a piecewise linear operator with a linear restricted partition, and can thereby be solved in a Gaussian framework by the methodology of Kolbjørnsen and Omre (2002). The Gaussian framework is defined next.

4 Statistical models

In Bayesian analysis knowledge and uncertainty is quantified by probability distributions. A generic distribution and a generic probability is denoted by p an P respectively. The relevant random variable will occasionally be displayed in the argument of p to clarify which distribution that is referred.

The likelihood is the statistical link between the parameter of interest and the observation. In Bayesian analysis it is given the interpretation of being the conditional distribution of traveltimes for a given slowness. In the current Gaussian framework, the observations are assigned additive Gaussian errors. Let $\boldsymbol{T}^*$ denote the random variable that is observed. The conditional distribution of $\boldsymbol{T}^*$ for a given s is then

$$p(\boldsymbol{t}^*|s) = N_r(\boldsymbol{\tau}_0^*(s), \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \ , \tag{5}$$

with $\boldsymbol{t}^*$ being the outcome of $\boldsymbol{T}^*$; s being a slowness field; N_r denoting the r dimensional multinormal distribution; $\boldsymbol{\tau}_0^*(s)$ being the approximate Fermat times for the slowness field s; and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ being the covariance for the observation error. Define also the indexed set of random variables $\boldsymbol{T}(\boldsymbol{\gamma})$, that is defined for each $\boldsymbol{\gamma}$ by the conditional distribution that correspond to observation of the path integrals, $\boldsymbol{\tau}(\boldsymbol{\gamma}, s)$,

$$p(\boldsymbol{t}_{\boldsymbol{\gamma}}|s) = N_r(\boldsymbol{\tau}(\boldsymbol{\gamma}, s), \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \ ,$$

with $\boldsymbol{t}_{\boldsymbol{\gamma}}$ denoting the outcome of $\boldsymbol{T}(\boldsymbol{\gamma})$; s being a slowness field; $\boldsymbol{\tau}(\boldsymbol{\gamma}, s)$ being the traveltimes in s along $\boldsymbol{\gamma}$; and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ being as in Expression (5). The marginal distribution of $\boldsymbol{T}^*$ is dependent on the distribution of the slowness and does not have an explicit representation in the current analysis. The marginal distribution of each of the random variables $\boldsymbol{T}(\boldsymbol{\gamma})$ will however be Gaussian if the slowness is so. The additive error term is modeled by a random error, it includes both observation errors and model errors.

In the current Gaussian framework the slowness, S, is assumed to be a Gaussian random field (Vanmarcke, 1983). The prior distribution is formally denoted $p(s)$, but is symbolic and not a density since S is a random field. The slowness to be reconstructed is assumed to be two times continuously differentiable, see Section 2. Gaussian random fields are well suited to represent different degrees of smoothness. In the presentation below it is assumed that the slowness field is almost surely two times continuously differentiable, i.e. $P(S \in C^2(\mathcal{R})) = 1$, see for example Stein (1999) for details about how to define such a Gaussian random field. This smoothness criterion is somewhat relaxed in Section 6, however.

The random variables that are defined by randomizing $\boldsymbol{g}(\boldsymbol{\gamma}, s)$ and $\boldsymbol{h}(\boldsymbol{\gamma}, s)$, see Expression (4), over the prior distribution of S for a fixed selection of paths, are denoted by capital letters, i.e. $\boldsymbol{G}(\boldsymbol{\gamma})$ and $\boldsymbol{H}(\boldsymbol{\gamma})$. Because the operators, $\boldsymbol{g}(\boldsymbol{\gamma}, s)$ and $\boldsymbol{h}(\boldsymbol{\gamma}, s)$, are linear in the second argument the corresponding random variables are Gaussian. These random variables are used to decompose the posterior distribution.

5 Representing the posterior distribution

The posterior distribution in the inverse problem of cross well tomography is decomposed as a mixture of truncated Gaussian distributions. This representation is in turn used to define an algorithm to sample the posterior distribution. The samples from the posterior distribution yields an approximation of the posterior distribution.

The theory of piecewise affine inverse problems, is developed in a Gaussian framework in Kolbjørnsen and Omre (2002). Using the notation introduced above, the posterior distribution of S can be represented as a mixture distribution

$$\begin{aligned} & p\{S = s, S \in \mathcal{A}_{\boldsymbol{\gamma}}, \boldsymbol{H}(\boldsymbol{\gamma}) = \boldsymbol{h} | \boldsymbol{T}^* = \boldsymbol{t}^*\} \\ & = p\{S = s | S \in \mathcal{A}_{\boldsymbol{\gamma}}, (\boldsymbol{T}(\boldsymbol{\gamma}), \boldsymbol{G}(\boldsymbol{\gamma}), \boldsymbol{H}(\boldsymbol{\gamma})) = (\boldsymbol{t}^*, \boldsymbol{0}, \boldsymbol{h})\} \qquad (6) \\ & \times \mathrm{P}\{S \in \mathcal{A}_{\boldsymbol{\gamma}} | (\boldsymbol{T}(\boldsymbol{\gamma}), \boldsymbol{G}(\boldsymbol{\gamma}), \boldsymbol{H}(\boldsymbol{\gamma})) = (\boldsymbol{t}^*, \boldsymbol{0}, \boldsymbol{h})\} \\ & \times \frac{|det(\boldsymbol{h})| p\{(\boldsymbol{T}(\boldsymbol{\gamma}), \boldsymbol{G}(\boldsymbol{\gamma}), \boldsymbol{H}(\boldsymbol{\gamma})) = (\boldsymbol{t}^*, \boldsymbol{0}, \boldsymbol{h})\}}{p\{\boldsymbol{T}^* = \boldsymbol{t}^*\}} \quad . \end{aligned}$$

The distribution on the left hand side is the posterior distributions of S when $S \in \mathcal{A}_{\boldsymbol{\gamma}}$ and $\boldsymbol{H}(\boldsymbol{\gamma})$ have the value $\boldsymbol{h}$. The marginal posterior distribution of S is obtained by randomizing Expression (6) over $\boldsymbol{\gamma}$ and $\boldsymbol{h}$. The first term in Expression (6) is a truncated Gaussian distribution, since the equality constraint is linear. The second term is a probability and the third term is a non-negative measure on $\Gamma_0 \times \mathbf{R}^{(2d-1)r}$, with d being the number of internal nodes and r being the number of observations. The product of the second and third term is the posterior density for S being in $\mathcal{A}_{\boldsymbol{\gamma}}$ with $\boldsymbol{H}(\boldsymbol{\gamma})$ having the value $\boldsymbol{h}$. Note that $\boldsymbol{h}$ is fully described by $(2d-1)r$ values, see Section 3. In addition a necessary condition for $S \in \mathcal{A}_{\gamma}$ is that $\boldsymbol{H}(\boldsymbol{\gamma})$ is positive definite. The mixing distribution of Expression (6), provides a sampling strategy for the posterior.

Algorithm 1 *CTGM-algorithm (Continuous Truncated Gaussian Mixing)*

1. *Sample* $\boldsymbol{\gamma}^{\#}, \boldsymbol{h}^{\#} \sim q(\boldsymbol{\gamma}, \boldsymbol{h})$ *with*

$$q(\boldsymbol{\gamma}, \boldsymbol{h}) \propto |det(\boldsymbol{h})| p\{(\boldsymbol{T}(\boldsymbol{\gamma}), \boldsymbol{G}(\boldsymbol{\gamma}), \boldsymbol{H}(\boldsymbol{\gamma})) = (\boldsymbol{t}^*, \boldsymbol{0}, \boldsymbol{h})\}$$

2. *Sample* $s^{\#} \sim p\left\{S = s | (\boldsymbol{T}(\boldsymbol{\gamma}^{\#}), \boldsymbol{G}(\boldsymbol{\gamma}^{\#}), \boldsymbol{H}(\boldsymbol{\gamma}^{\#})) = (\boldsymbol{t}^*, \boldsymbol{0}, \boldsymbol{h}^{\#})\right\}$
3. *If* $S^{\#} \in \mathcal{A}_{\boldsymbol{\gamma}^{\#}}$ *stop.*

The algorithm splits the sampling into a nonlinear step, a linear step and an acceptance step. In the nonlinear step a value for the the Fermat paths and

the Hessian of the traveltimes along the Fermat paths is proposed. The matrix $\boldsymbol{h}^{\#}$ is restricted to be positive definite, hence the paths $\boldsymbol{\gamma}^{\#}$ are local minima. In the second step a slowness field, $s^{\#}$, that have local minima along the paths $\boldsymbol{\gamma}^{\#}$, with $\boldsymbol{h}(\boldsymbol{\gamma}^{\#}, s^{\#}) = \boldsymbol{h}^{\#}$ is drawn. In the third step it is controlled that $\boldsymbol{\gamma}^{\#}$ in fact is the Fermat paths, if not the sampled slowness is rejected and a new pair of $(\boldsymbol{\gamma}, \boldsymbol{h})$ must be drawn. Since the proposed paths are guaranteed to be local minima, there is usually a high acceptance rate in the third step. The nonlinear step in the algorithm is the challenge. To sample the distribution $q(\boldsymbol{\gamma}, \boldsymbol{h})$ a McMC algorithm is used. The decomposition given in Expression (6), can also be exploited in other types of algorithms. The benefit of using the decomposition is that it uses the global structure of the inverse problem.

6 Generalization to a non-smooth slowness

The smoothness assumption regarding the slowness is common in a continuous formulation of cross well tomography. In the current work it is however imposed by the solution method and is hence undesirable. In this section the theory is extended to account for small perturbations from a smooth background, let

$$s(x, z) = s_L(x, z) + \epsilon\, s_H(x, z) \ ,$$

with s_L being a lowfrequent background model; and $\epsilon\, s_H$ being a highfrequent perturbation with ϵ being a small number. By a standard perturbation argument, similar to the one used in Boyse and Keller (1995), the traveltime can be expanded in an asymptotic series in powers of ϵ. Including only the first order, this reads

$$\boldsymbol{\tau}^*(s_L + \epsilon\, s_H) = \boldsymbol{\tau}^*(s_L) + \epsilon\, \boldsymbol{\tau}(\boldsymbol{\gamma}^*(s_L), s_H) + \mathcal{O}(\epsilon^2)$$

with $\boldsymbol{\tau}^*(s_L)$ being the Fermat times in the lowfrequent part of the slowness; $\boldsymbol{\gamma}^*(s_L)$ being the Fermat path in the lowfrequent part; $\boldsymbol{\tau}(\boldsymbol{\gamma}^*(s_L), s_H)$ being the line integral of s_H along $\boldsymbol{\gamma}^*(s_L)$; and $\mathcal{O}(\epsilon^2)$ being higher order terms which are neglected in what follows. The likelihood in Expression (5) is now replaced by

$$p(\boldsymbol{t}^*|s_L, s_H) = N_r\left(\boldsymbol{\tau}_0^*(s_L) + \epsilon\, \boldsymbol{\tau}(\boldsymbol{\gamma}_0^*(s_L), s_H), \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \ ,$$

with $\boldsymbol{\tau}_0^*(s_L)$ and $\boldsymbol{\gamma}_0^*(s_L)$ being the approximate Fermat times and paths in the lowfrequent part of the slowness respectively.

The sampling of the the lowfrequent and highfrequent part is done sequentially. For a fixed low frequent part the problem of sampling the highfrequent part is the linearized problem for a non-constant back ground. The challenge is hence to sample the lowfrequent part in the presence of the highfrequent part. This can be done by computing the marginal likelihood of s_L. Assuming S_H to be a Gaussian random field independent of s_L this can be done analytically. If S_H is centered the marginal likelihood for $s_L \in \mathcal{A}_{\boldsymbol{\gamma}}$ is,

$$p(\boldsymbol{t}^*|s_L) = N_r(\boldsymbol{\tau}_0^*(s_L), \boldsymbol{\Sigma}(\boldsymbol{\gamma}) + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \ ,$$

with $\mathbf{\Sigma}(\boldsymbol{\gamma})$ being the covariance of line integrals of ϵS_H along the paths, $\boldsymbol{\gamma}$. Assuming that S_L is a Gaussian random field, this formulation is still within the scope of the theory of piecewise affine inverse problems developed by Kolbjørnsen and Omre (2002).

The characteristic that allows for the generalization is that the highfrequent part have an additive effect for which the statistical properties only depend on the index of the piecewise affine inverse problem, i.e. the Fermat paths. Sølna and Papanicolaou (2000) find a similar result for a different type of deviation from a smooth background.

7 Example

In the current section a synthetic example is investigated to highlight some of the differences between the current approach and a linearized problem. The traveltimes investigated relates to the slowness in Figure 2.

The slowness is a stationary Gaussian random field and is defined by its mean, variance and spectral density, these are denoted by μ_S, σ_S^2 and $\phi_S(k_x, k_z)$ respectively. It is convenient to specify the correlation in terms of the spectral density since this makes it easier to control the differentiability of the random field. The spectral density is assumed to have the form

$$\phi_S(k_x, k_z) \propto \left(1 + (k_z\, L_z)^2 + (k_x\, L_x)^2\right)^{-(\nu+2)/2},$$

with k_z and k_x being spatial frequencies; L_z and L_x being scales in depth and lateral direction respectively; and ν being the parameter that controls the smoothness. In the subsequent examples the prior distribution is defined by $\mu_S = 0.5\,\text{ms/m}$, $\sigma_S = 0.06\,\text{ms/m}$, $L_z = 225\,\text{m}$, $L_x = 130\,\text{m}$ and $\nu = 18$. Figure 6 show the resulting covariance function for the depth, z, and the lateral component, x. The slowness field in Figure 2 is a random sample from this prior distribution.

The observations have variance $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma^2 \boldsymbol{I}_{r\times r}$, with $\sigma = 0.1\,\text{ms}$ being the standard deviation of the error; $\boldsymbol{I}_{r\times r}$ being the $r \times r$ identity matrix; and r being the number of observations. The observations are hence recorded with a high precision since the travel times are ranging from 48.1 ms to 76.6 ms

7.1 One observation

In this paragraph only one observation is considered. The source is in the left well at the depth 150 m and the receiver is in the right well at the depth 50 m. The approximation of the Fermat path in the true slowness field is displayed in Figure 5 for a variable number of internal nodes. The approximation of the Fermat time as a function of the number of internal nodes is displayed in the

top right corner in Figure 4. The observed traveltime is 76.6 ms. One traveltime observation hardly provide any information regarding the slowness field, hence no features of the true slowness can be expected to be retrieved. The example highlight differences between linear and nonlinear cross well tomography, however.

The inversion procedure is carried out for zero, one and seven internal nodes. The case with no internal nodes correspond to the linear case and is not discussed in any further detail. When only one internal node is considered, the mixing distribution $q(\gamma, h)$ is two dimensional, the density for the current case is displayed in Figure 7. To sample the mixing distribution for the case of seven internal nodes, a Markov chain is constructed. The algorithm use a diffusion step to sample the path. For a given path, γ, the distribution $q(\gamma, \boldsymbol{h})$ is approximated by a Gaussian distribution, this distribution is sampled sequentially to assure $\boldsymbol{h}$ to be positive definite. One sample is extracted for every 200 iteration, extracting a total of 3000 samples. Figure 8(a) show the value of the 2end, 4th and 6th internal node, and Figure 8(b) show the corresponding diagonal elements of $\boldsymbol{h}$. The plots show that the algorithm is slowly mixing. Figure 9(a)-(c) show the paths used in the inversion when zero, one and seven internal nodes are used respectively. For the case of one and seven internal nodes these are the samples from the corresponding mixing distribution in Step 1 of the CTGM-algorithm. For comparison the true path is plotted in the same figures. The uncertainty of the path is clearly illustrated by the figures. The acceptance rate in the third step of the CTGM-algorithm is 96% and 92% for the case with one and seven nodes respectively.

Figure 10(a)-(c) show the final estimates using the three strategies. The estimate for zero internal nodes is obtained analytically. Visually the estimates appear to be similar. All estimates increase the slowness along the line connecting the source and the receiver. The main effect of the internal nodes are better seen in cross sections of the estimates. Figure 11(a)-(c) show cross sections of the estimates at $x = 10\,\mathrm{m}$, $x = 50\,\mathrm{m}$ and $x = 75\,\mathrm{m}$ respectively. The nonlinear estimates are consistently larger, and have a larger region of influence. The deviation from the background is 20% larger for the case with seven internal nodes than it is for the linear estimate. Much of the nonlinear effect on the estimate is present in the case with only one internal node.

The main effect of the nonlinearity is however hidden by the averaging that is done in the estimation. The nonlinearity is present in the individual samples. To illustrate the differences, 500 samples from the three conditional distributions are used. Let γ_0 denote the direct line from the source to the receiver. For each sample, $s^{\#}$, the two traveltime functionals $\tau(\gamma_0, s^{\#})$ and $\tau^*(s^{\#})$ are computed. That is the line integral of the slowness along γ_0 and the Fermat time. Figure 12(a)-(c) display the scatter plot of these two functionals evaluated for each sample using three approaches. In linear tomography the line integral remains stable and the Fermat time fluctuates, whereas in the case with seven internal

nodes the opposite effect is observed. For the case with one internal node the Fermat times are quite stable, but some large deviations are present. Compare also some of the conditional probabilities that is illustrated in Figure 12. The percentage of the samples having Fermat time less than 75 ms is 29%, 5% and 0%, and the percentage of the samples having have line integral larger than 78 ms is 0%, 16% and 24% for the case with zero, one and seven internal nodes respectively. Much of the nonlinear effect is hence gained by including just one internal node.

In the algorithm, the Fermat path is drawn conditioned to the observed traveltime. The mixing distribution of the Fermat path, $q(\gamma, h)$, is hence dependent on the observed value of the traveltime. Figure 13(a) and (b) visualize this effect in the case with one internal node. The figures show $q(\gamma, h)$ for $t = 50$ ms and $t = 100$ ms. Note that $q(\gamma, h)$ is not the posterior distribution of (γ, h) since the acceptance probability is factored out, but $q(\gamma, h)$ still indicate the general shape of the distribution since the acceptrate in the third step of the CTGM-algorithm is large. When the observed value of t is small, i.e. $t = 50$ ms, it is likely that the path has followed the direct line from source to the receiver. This is illustrated in Figure 14(a) where 1000 paths sampled from $q(\gamma, h)$ are displayed. Notice the low spread of the samples that, indicate a channel of high velocity connecting the source and receiver. When the observed value of of t is large, i.e. $t = 100$ ms, it is likely that the Fermat path is bent either up or down as is indicated by the bi-modality in Figure 13(b). This is illustrated in Figure 14(b) where 1000 paths sampled from $q(\gamma, h)$ are displayed. Notice how most paths avoid the middle of the figure. This indicate a bump of low velocity located on the direct line connecting the source and the receiver.

7.2 Several observations

In this paragraph all 16 traveltimes, see Figure 2, are considered. Compared to the results of the previous paragraph, more of the structure of the slowness field is expected to be recovered.

The inversion procedure is carried out for zero and one internal node. The results for zero internal nodes are obtained analytically. The results for one internal node is obtained using the CTGM-algorithm. To sample the mixing distribution in Step 1 of the CTGM-algorithm a Markov chain is constructed in the same manner as in the previous paragraph. In each step a change is proposed in all the paths simultaneously. For the given path proposal the distribution $q(\gamma, \boldsymbol{h})$ is approximated by a Gaussian distribution and sampled sequentially to assure $\boldsymbol{h}$ to be positive definite. A sample is extracted after every 400 iteration, extracting a total of 2000 samples. Figure 15 show the mixing plot of the 32 random variables that are sampled. In general the mixing plots are satisfactory, but the internal node in the path that start in the left well at depth 150 m and arrive in the right well at depth 50 m to is however mixing slightly slower than

the other parameters. The mixing plot for this parameter is in the top right corner in Figure 15(a). The acceptance rate in the third step of the CTGM-algorithm is 98%.

Figure 16(a) and (b), show the estimates from the two models. Visually the estimates appear to be similar and have captured some of the features of the slowness. A high slowness region in the true slowness is located is located from depth 60 m to 120 m and at lateral position 30 m to 100 m. This is also present in the estimates, but the shape is slightly wrong. At the depth of 150 meters the estimates have a high value at the left and a low value at the right. This is also so for the true slowness. Comparing the two different estimates closely, the features are more diffuse in the nonlinear estimate than in the linear estimate. When the deviation from the true surface is measured, the nonlinear estimate improve the quadratic loss by 10%. It is however substantially more time consuming to compute the nonlinear estimate.

As a measure of the variability, the pointwise variance is integrated. The posterior in the linear case has 30 % lower integrated variance than the posterior in the nonlinear case. This does not necessarily mean that the full posterior is better determined in the linear case, since the integrated variance only respond to the marginal posterior distributions. In the linear case each observation will reduce the posterior integrated variance. This is generally not true for nonlinear observations.

8 Discussion

The inverse problem in nonlinear cross well tomography is solved by a Bayesian methodology in a Gaussian framework. The traveltimes obey Fermat's principle. This variational structure is approximated by a finite element method. Under the finite element approximation the forward map of nonlinear cross well tomography is piecewise affine. For a test example the approximation is reasonable even for a coarse resolution of the finite elements.

The estimate is taken to be the posterior expectation which is optimal under quadratic loss. The posterior distribution is explored by sampling and the expectation is approximated by the sample average. When the conditional expectation is used as estimator, the estimated slowness field will not reproduce the Fermat times in the case of exact observations. This is due to the convexity of the problem. Each individual sample will have a Fermat time corresponding to the observed time, but the Fermat path will differ between samples. When all the samples are averaged the Fermat time will be a lower bound for the path integral along any path, hence the Fermat time in the average medium will be larger. In general it is difficult to preserve nonlinear properties in an estimator. In nonlinear cross well tomography this can however be done by estimating the Fermat paths and then average the slowness for the given Fermat times under

the given selection of Fermat paths. This will however raise issues on estimating the Fermat paths. This is further complicated by the fact that the posterior distribution of the Fermat paths can be multi modal, see Figure 13.

The ray paths are flexible in the current nonlinear approach whereas in a linear approach, they are fixed. When studied in an example with one observation the nonlinear estimate have a deviation from the background that is 20% larger than the linear estimate. When studied in an example with 16 observations, the nonlinear estimate perform 10% better in terms of quadratic loss compared to the linear estimate. In both cases however the estimates look similar and only a small amount is gained by using the methodology in this respect. The major impact of the nonlinearity is however regarding typical deviations from the estimate, i.e. in the uncertainty.

The challenge in the methodology is to sample the mixing distribution $q(\boldsymbol{\gamma}, \boldsymbol{h})$, see Algorithm 1. In the current work this is done by a naive implementation of a McMC algorithm, the resulting chain is slowly mixing. Efficient exploration of $q(\boldsymbol{\gamma}, \boldsymbol{h})$ is of high importance for further development of the methodology.

The prior distribution of the slowness field is Gaussian. Gaussian random fields constitute a large class of prior distributions and is in particular well suited for modeling of smoothness. The methodology can also be extended to priors being mixtures of Gaussian distributions.

Acknowledgments

This work was supported by the Research Council of Norway. The author thanks Henning Omre for helpful comments.

References

Aronsson, G. (1970) "Axiomatic derivation of Fermat's principle", SIAM Journal on applied Mathematics, Vol. 18, no. 3, pp 675-681.

Auld, B.A (1990) "Acoustic Fields and Waves in solids",Vol. 1, Robert E. Krieger, Melbourn, F1.

Berryman J.G. (1990) "Stable iterative reconstruction algorithm for nonlinear traveltime tomography", Inverse problems 6, pp 21-42.

Berryman, J.G. (1997) "Variational structure of inverse problems in wave propagation and vibration", Inverse problems in wave propagation (Minneapolis, MN, 1995), 13–44,IMA Vol. Math. Appl., 90, Springer, New York.

Bertero, M. (1989), "Linear inverse and ill-posed Problems", Advances in Electronics and Electron Physics, Academic Press, New York.

Boyse, W. and Keller, J.B. (1995) "Short acoustic, electromagnetic, and elastic waves in random media", J. Opt. Soc. Amer. A., Vol 12 , no. 2, pp 380-389.

Bube K.P. and Meadows M.A. (1998) "Characterization of the null space of a generally anisotropic medium in linearized cross well tomography", Geophys.J.Int, 133 pp 65-84.

Carfantan, H. and Mohammad-Djafari, A. (1997) "An overview of nonlinear diffraction tomography within the Bayesian estimation framework", Inverse problems of wave propagation and diffraction, 107–124, Lecture Notes in Phys., 486, Springer, Berlin.

Faridani A. (1997) "Results, old and new, in computed tomography, in inverse problems in wave propagation", G. Chevalent et al. (editors), The IMA Volumes in mathematics and its Applications, vol 90, Springer Verlag New York, pp 167-193.

Kolbjørnsen, O and Omre H. (2002) "Bayesian inversion of piecewise affine operators in a Gaussian framework"

Langan R.T., Lerche I. and Cutler R.T. (1985) "Tracing rays through heterogeneous media an accurate and efficient procedure", Geophysics, 50 pp 1456-1465.

Langan R.T. and Bube, K.P. (1998) "A resolution analysis of cross well seismic tomography and its implications for the formulation of a traveltime inversion algorithm", in Mathematics of Reflection Seismology, ed Symes W.W.SIAM Frontiers in Applied Mathematics.

Menke, W. (1984) "The resolving power of cross-borehole tomography", Geophys. Res.Lett., 11, pp 105-108.

Michelena, R.J (1993) "Singular value decomposition for cross well tomography", Geophysics 58, 1655-1661.

Natterer, F. (1980) "Efficient implementation of "optimal" algorithms in computerized tomography", Math. Methods Appl. Sci. 2, no. 4, 545–555.

Stein, M. L. (1999) "Interpolation of spatial data. Some theory for Kriging", Springer Series in Statistics. Springer-Verlag, New York.

Sølna, K. and Papanicolaou, G. (2000) "Ray Theory for a Locally Layered Random Medium", Waves in Random Media, vol 10, pp. 155-202.

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

Vidale J.E. (1988) "Finite difference traveltime calculation", Bull.seism. Soc. Am, 78 pp 2062-2076.

Tables and figures

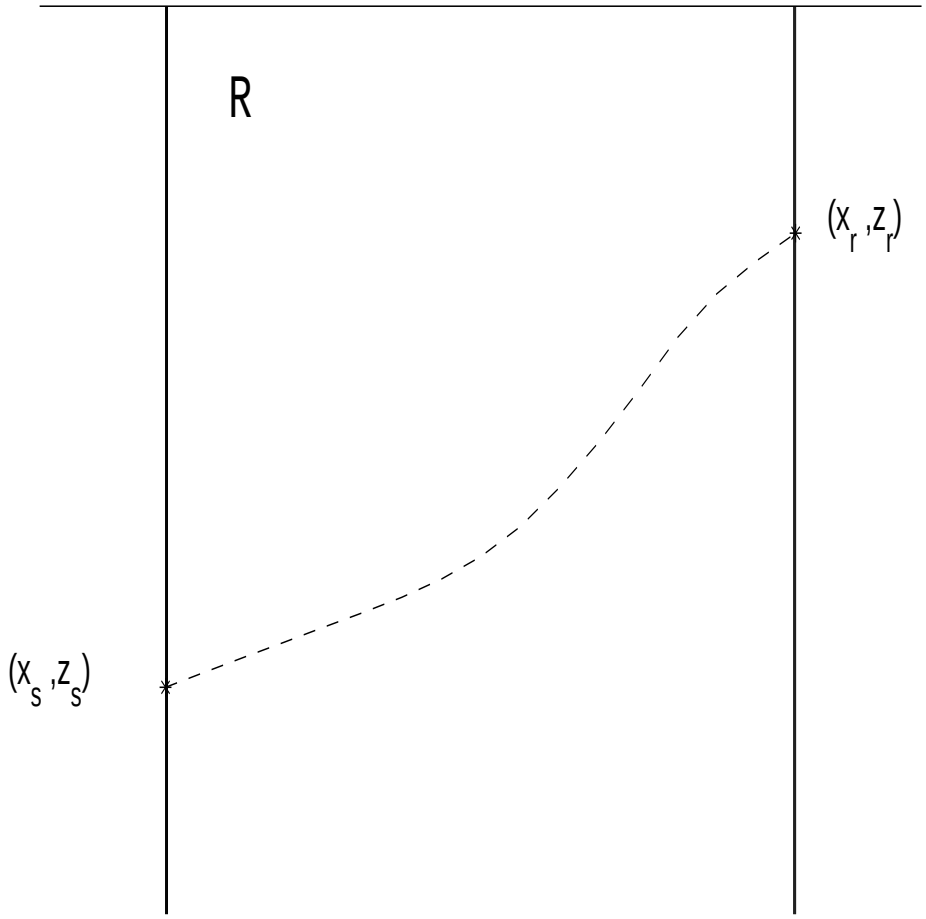
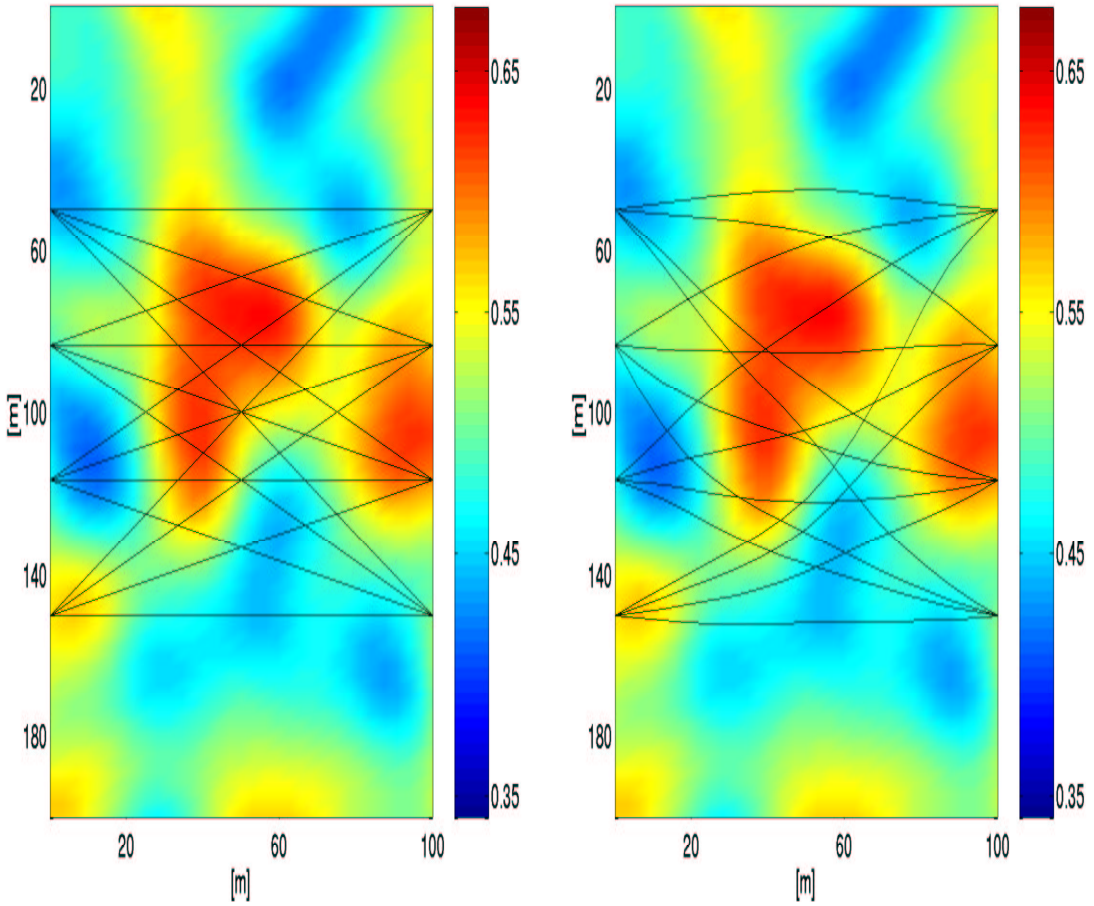


Figure 1: Cross well tomography. The two vertical lines are boreholes, the region between the two boreholes is $\mathcal{R}$. A source is situated in one well at the the location (x_s, z_s), a receiver is situated in the other well at the the location (x_r, z_r). The time it takes for a pulse to propagate from the source to the receiver is recorded.



(a) (b)

Figure 2: Linear paths and Fermat paths. The paths used for linear tomography (a); The Fermat paths for the superimposed slowness field (b).

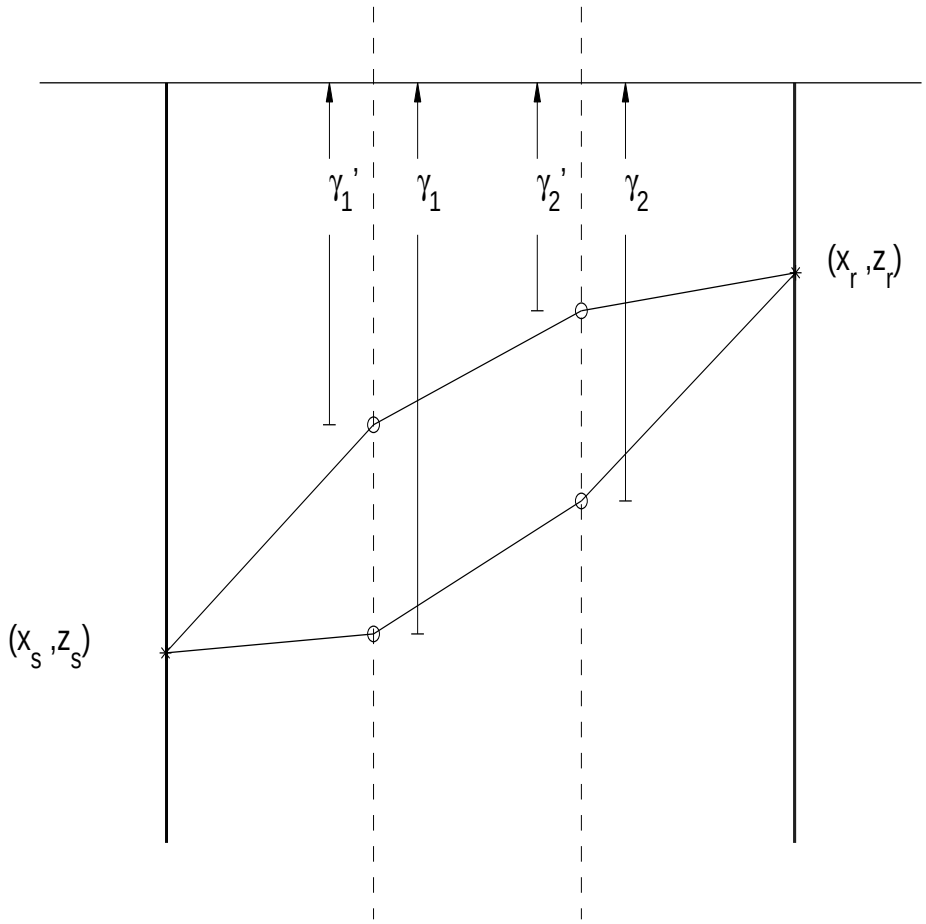


Figure 3: Finite element approximation to Fermat's principle. The path between the source and the receiver is restricted to be piecewise linear between internal nodes. Two different paths are displayed for the case with two internal nodes. The path is parameterized by the vertical distance to the knot point,(γ_1, γ_2) and (γ_1', γ_2') for the two paths respectively

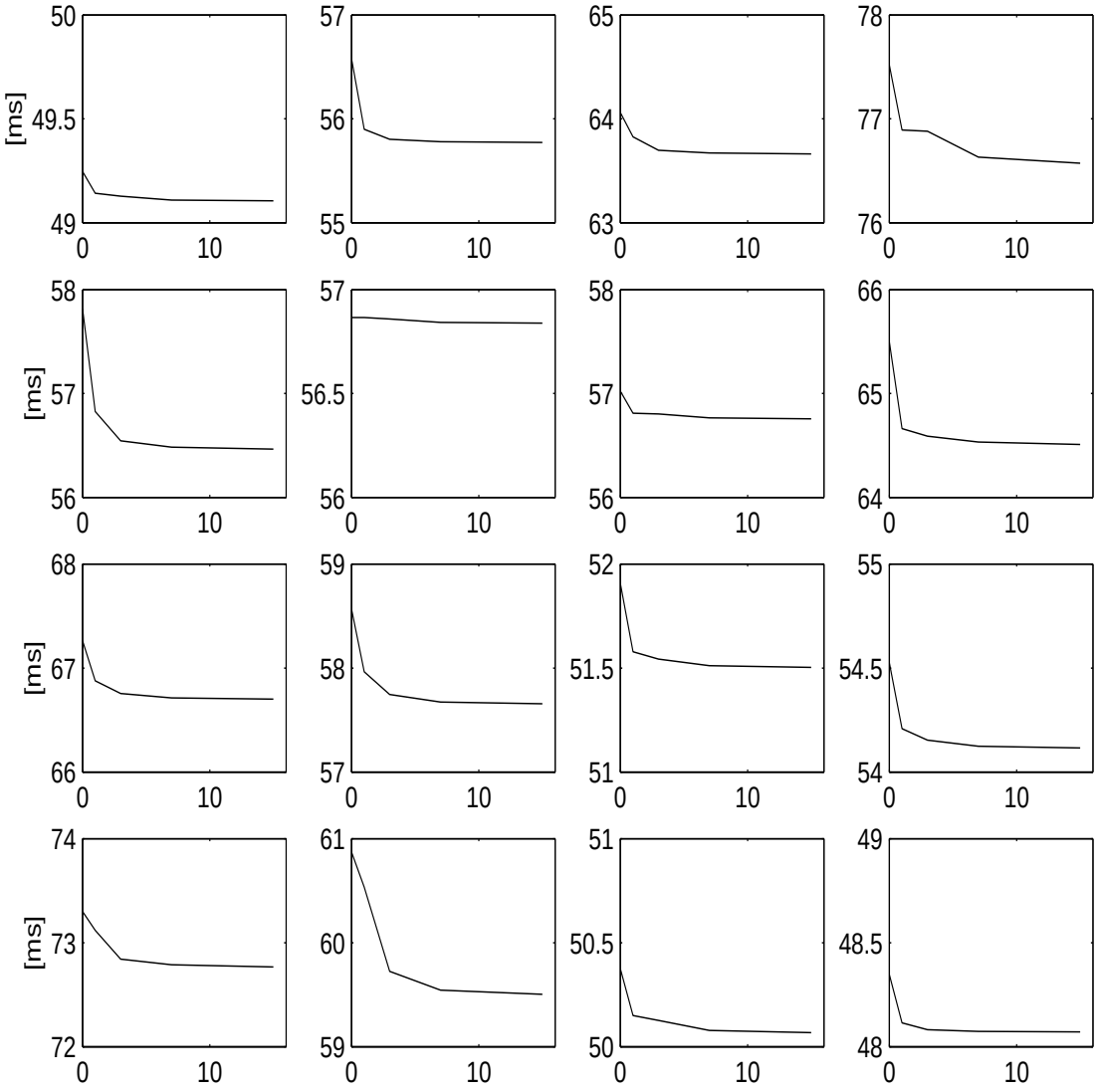


Figure 4: Minimum traveltime, dependence of number of internal nodes. For each of the 16 traveltimes, the minimum traveltime is plotted as a function of the number of internal nodes. Increasing column number correspond to increasing depth of starting point. Increasing row number correspond to increasing depth of end point. The values where computed for zero, one, three, seven and 15 internal nodes to have a monotone decay.

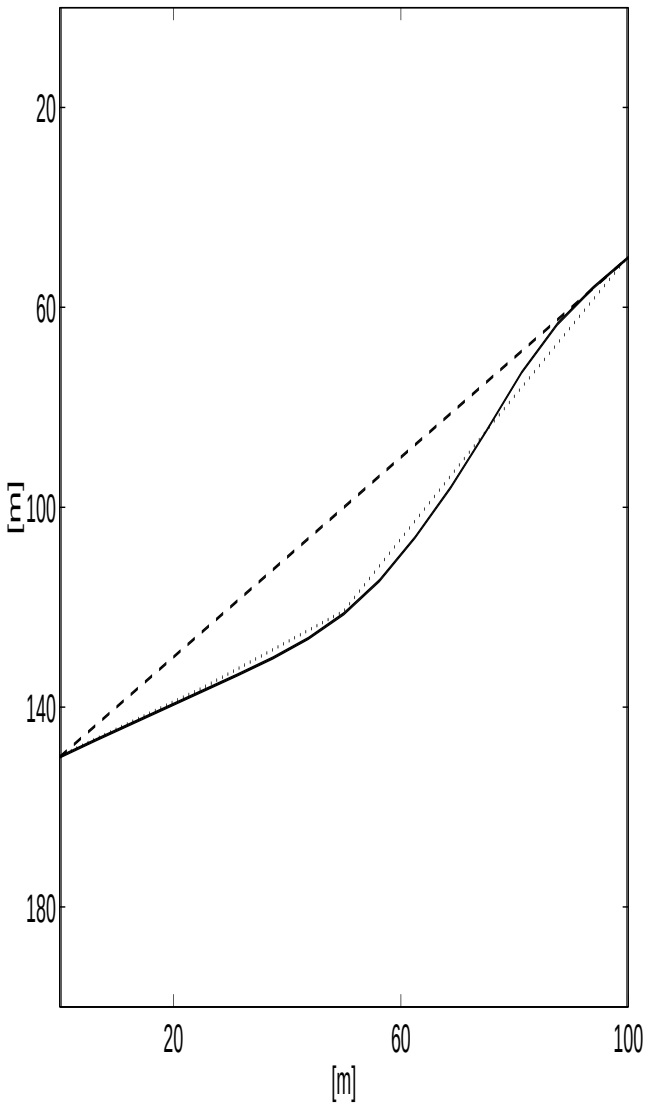


Figure 5: Minimum path, dependence of number of internal nodes. For one set of endpoints, the minimum path is plotted for zero (dashed line), three (dotted line) and 15 (full line) internal nodes.

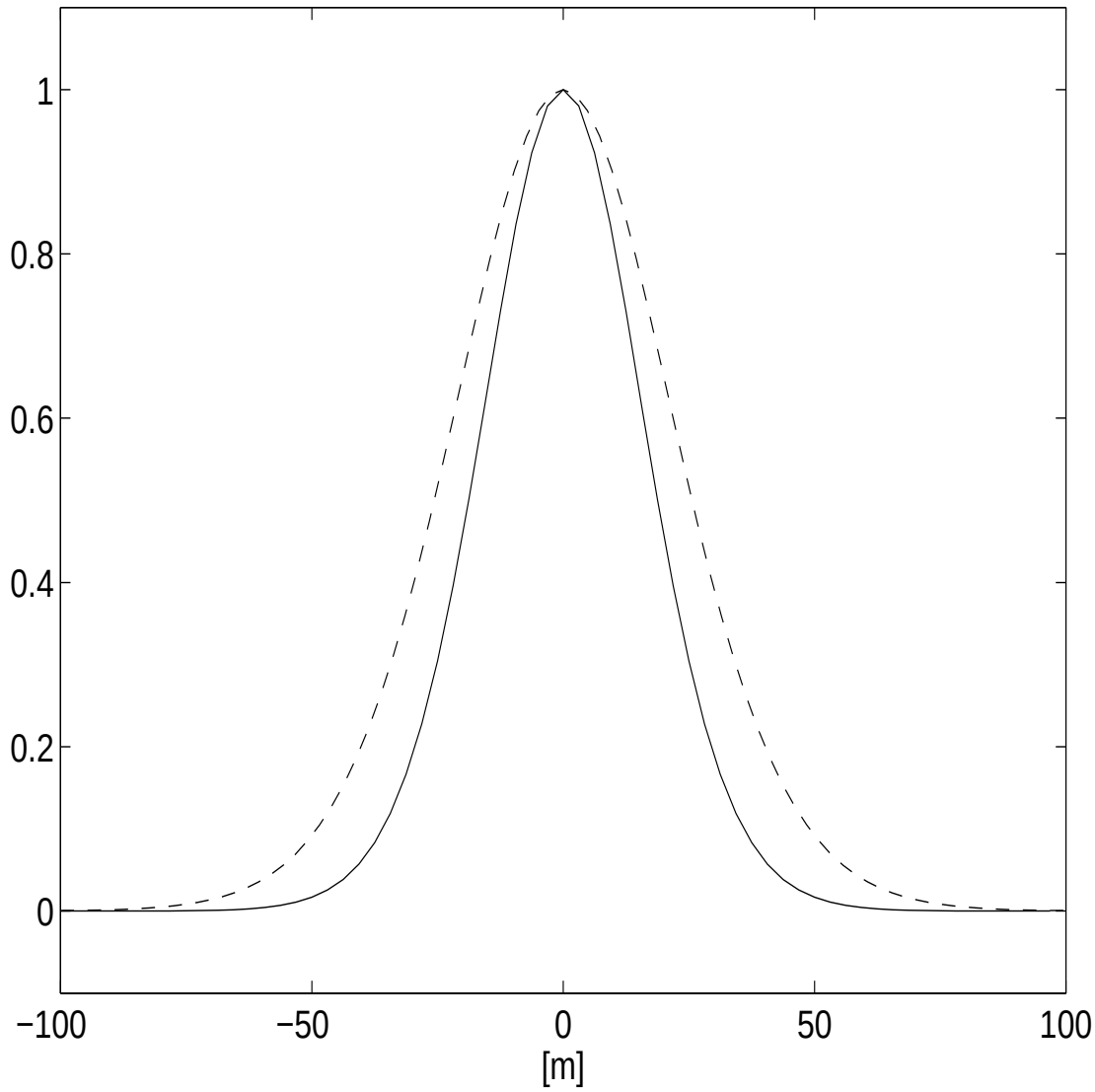


Figure 6: The correlation function used in the example of Section 7. The dashed line being for the depth, z, and the full line being for the lateral direction, x.

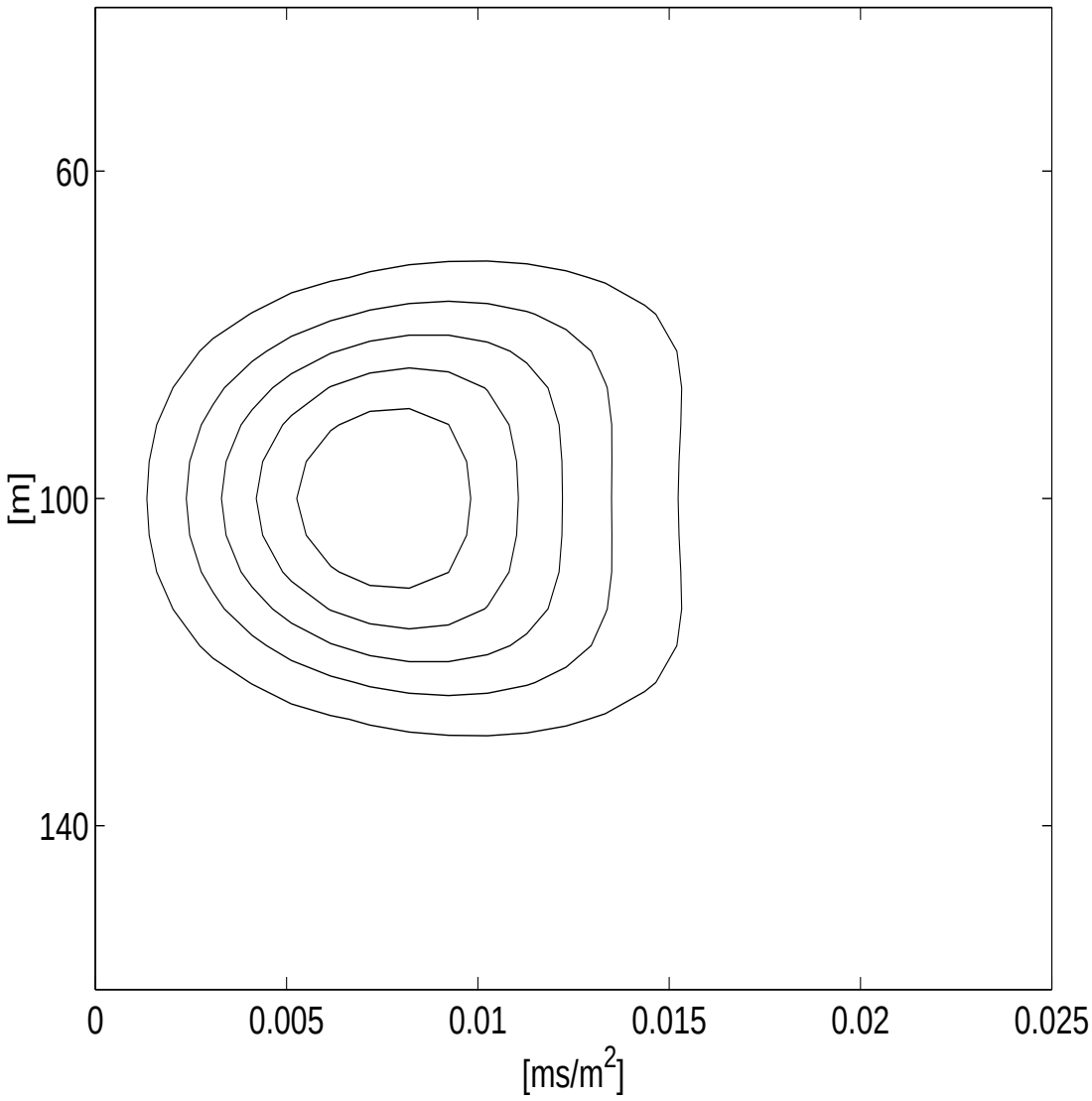
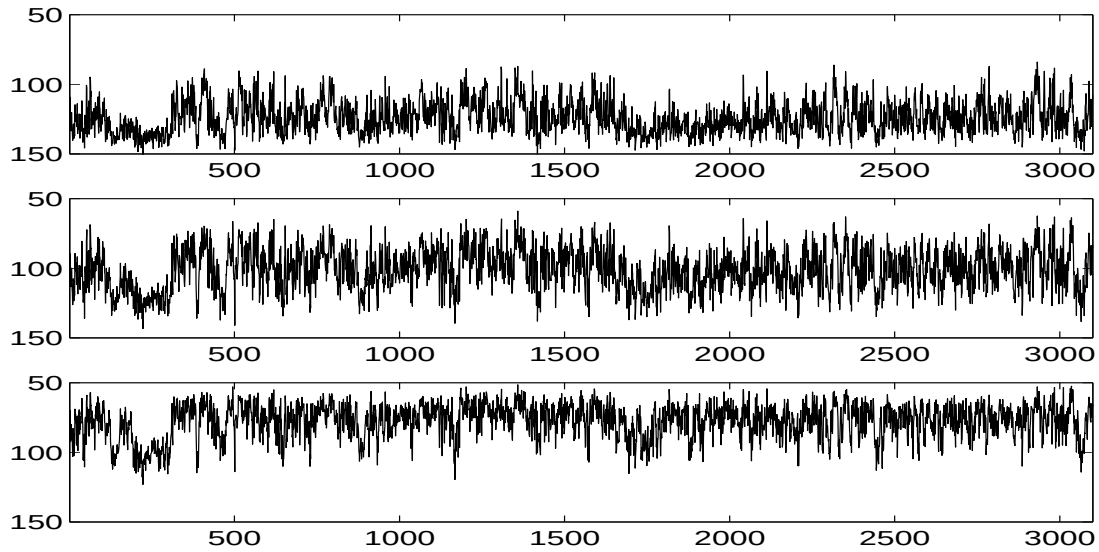
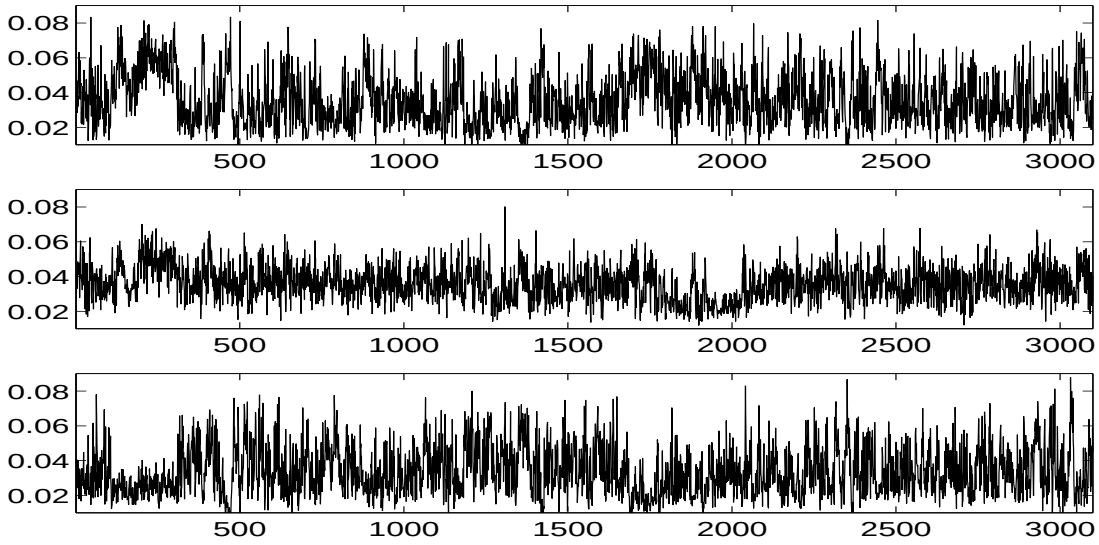


Figure 7: Proposal distribution for one internal node. The proposal distribution in the nonlinear step in the CTGM algorithm for the actual observation; i.e. $t = 76.6\,\text{ms}$.

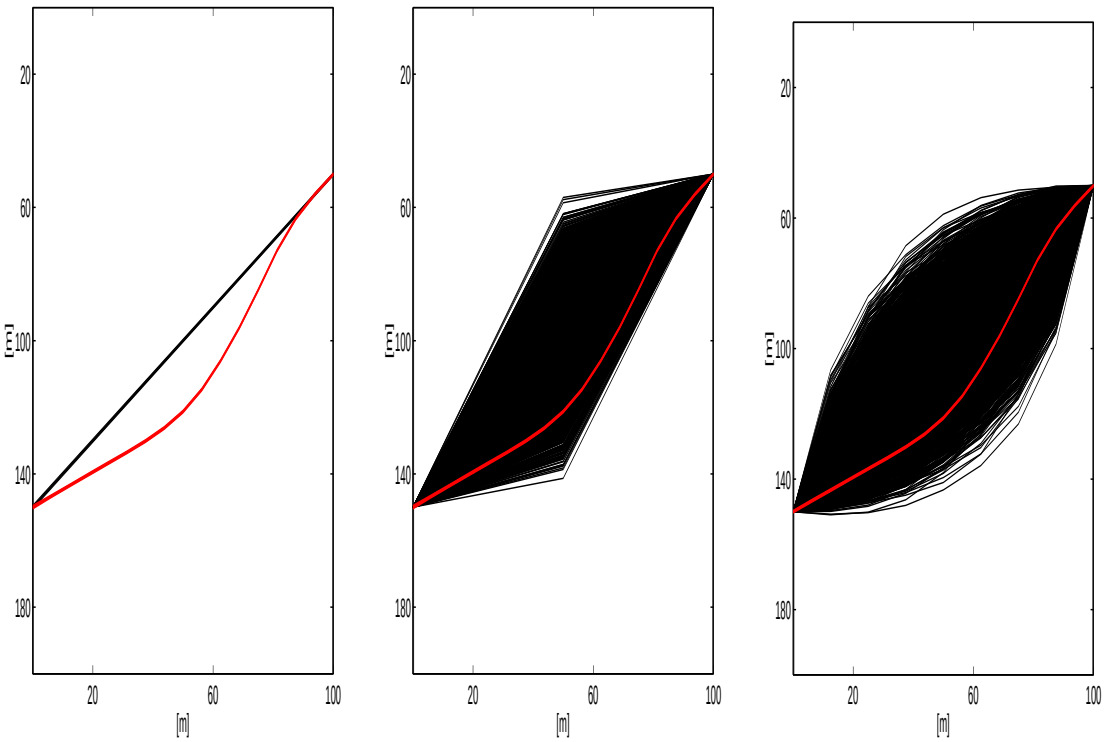


(a)



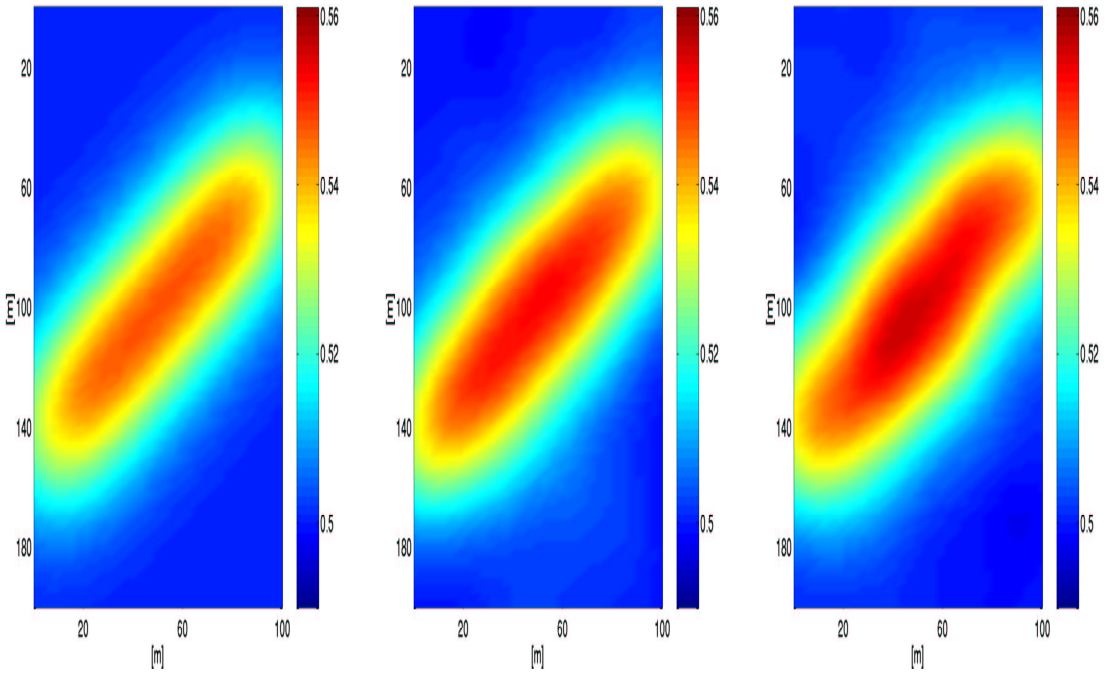
(b)

Figure 8: Mixing plot for the Markov chain used in nonlinear inversion for seven internal nodes. Three path parameters (a); three parameters for the gradient of the constraint (b). In both (a) and (b) the top is for the parameter corresponding to $x = 25\,\mathrm{m}$ the middle corresponding to $x = 50\,\mathrm{m}$ and the bottom corresponding to $x = 75\,\mathrm{m}$.



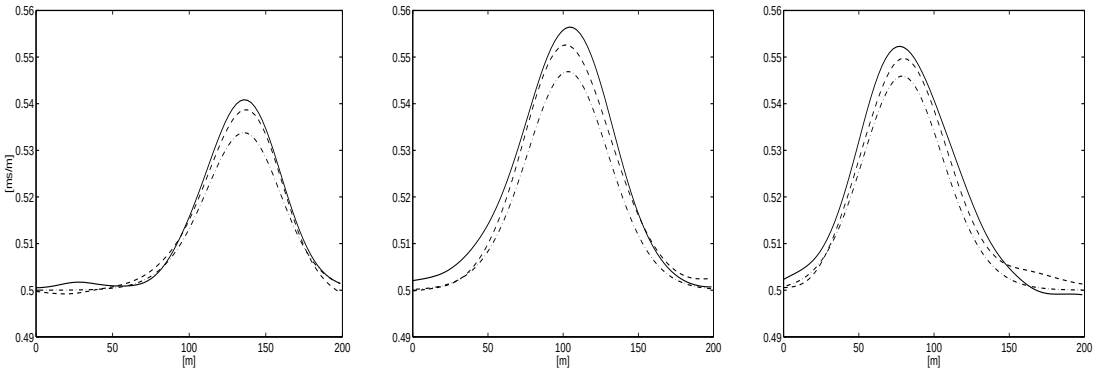
(a) (b) (c)

Figure 9: Comparison of paths used for reconstruction. The red line is the actual Fermat path in the problem, the black lines are the paths used in (a) linear inversion; (b) nonlinear inversion with one internal node;(c) nonlinear inversion with seven internal nodes. In (b) and (c) 3000 paths are displayed.



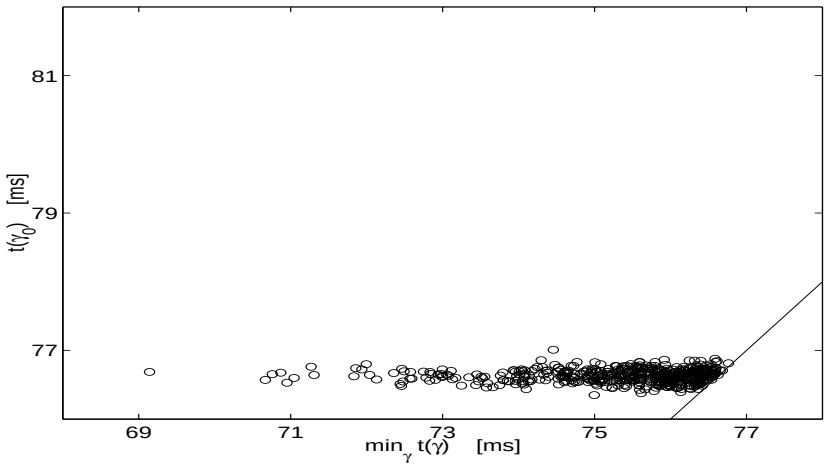
(a) (b) (c)

Figure 10: Comparison of estimates. The conditional expectation using (a) linear tomography; (b) nonlinear tomography one internal node; (c) nonlinear tomography seven internal nodes.

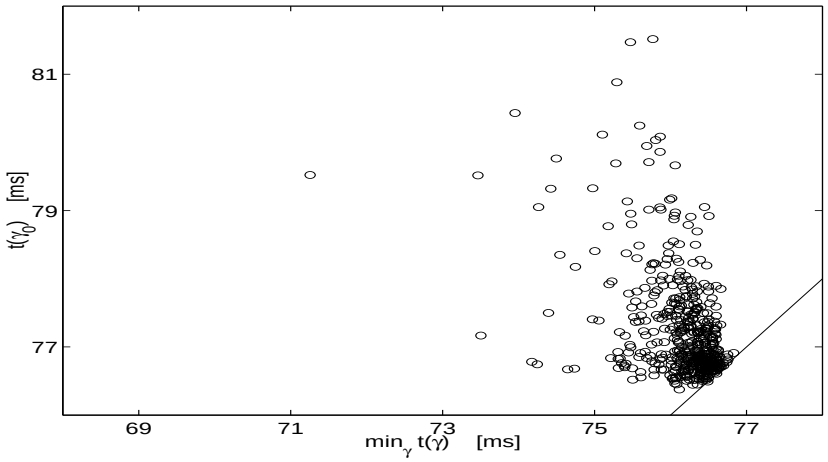


(a) (b) (c)

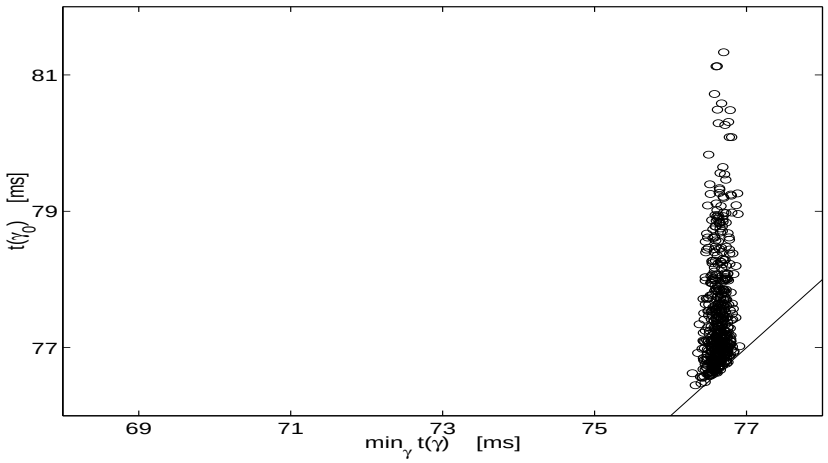
Figure 11: Cross sections of estimates. Dash/dot line - linear tomography; dashed line - nonlinear tomography with one internal node; full line - nonlinear tomography with seven internal nodes. The cross sections show vertical slices for lateral components (a) $x = 10\,\mathrm{m}$;(b) $x = 50\,\mathrm{m}$; (c) $x = 75\,\mathrm{m}$.



(a)

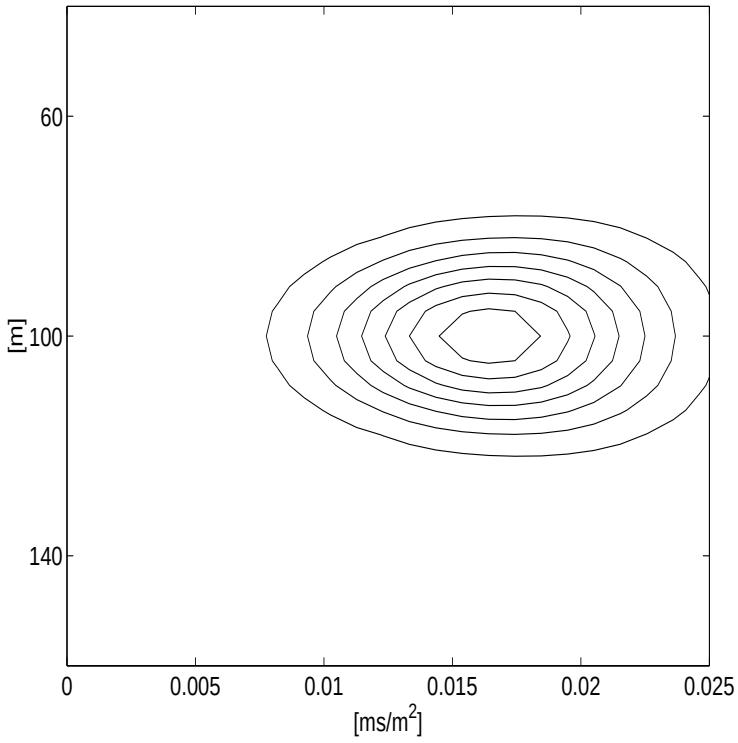


(b)

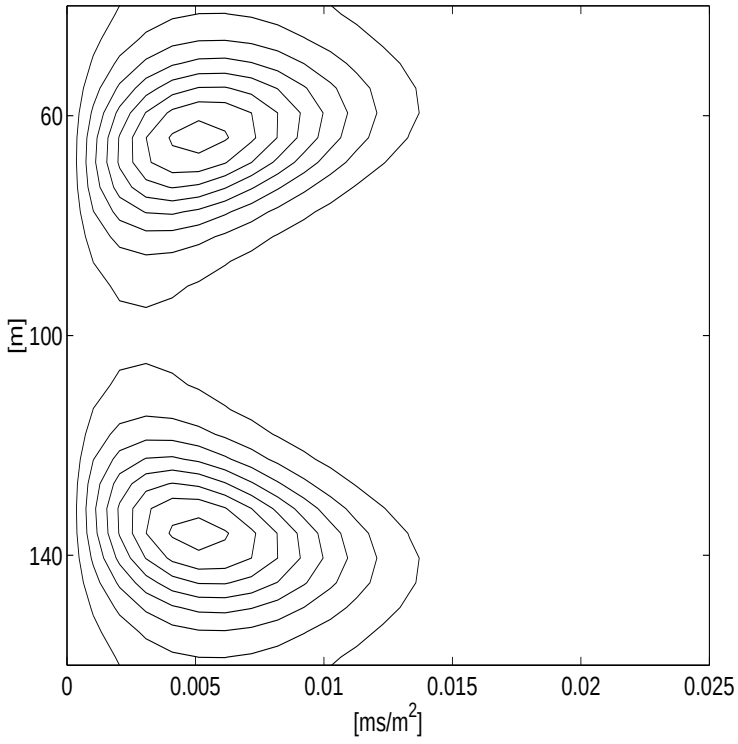


(c)

Figure 12: Comparison of results for linear and nonlinear inversion. The scatter plot of the posterior distribution of the traveltime along the linear path and the Fermat path for (a) linear tomography; (b) nonlinear tomography with one internal node; (c) nonlinear tomography with seven internal nodes.



(a)



(b)

Figure 13: Proposal distribution for one internal node. The proposal distribution in the nonlinear step in the CTGM algorithm for extreme observations (a) $t = 50\,\text{ms}$; (b) $t = 100\,\text{ms}$.

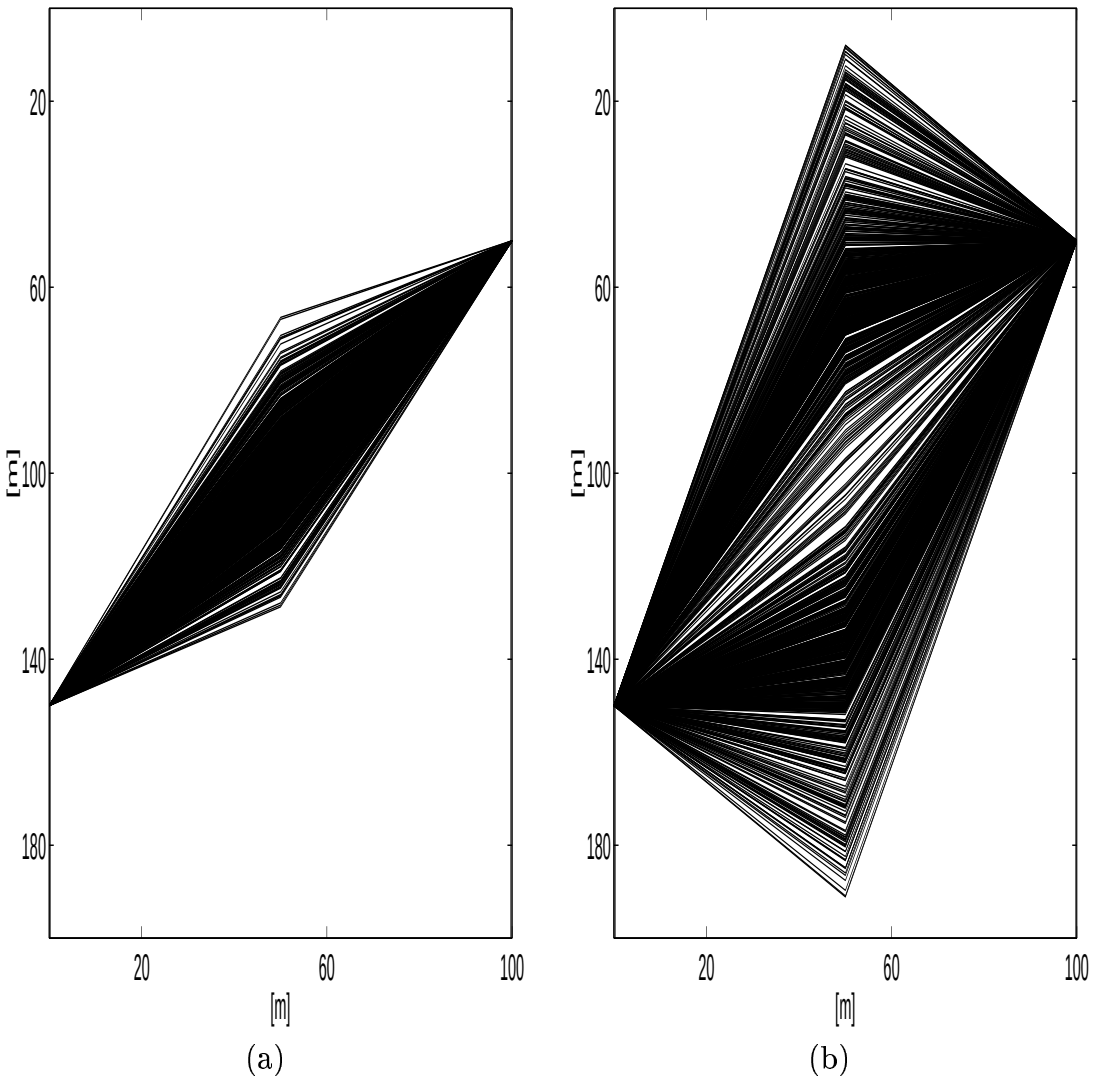
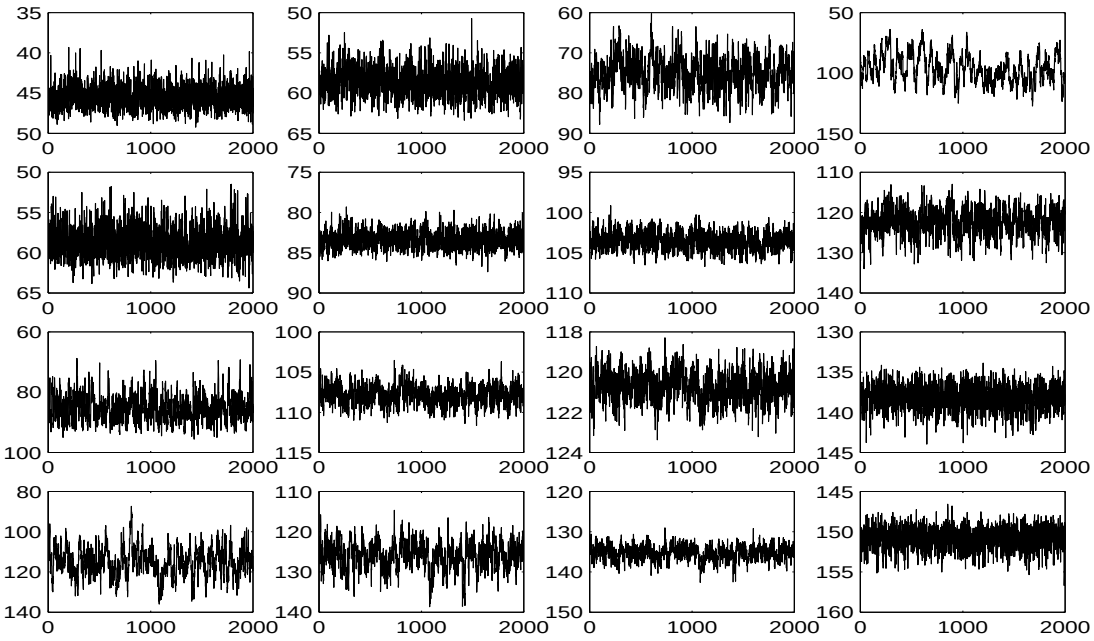
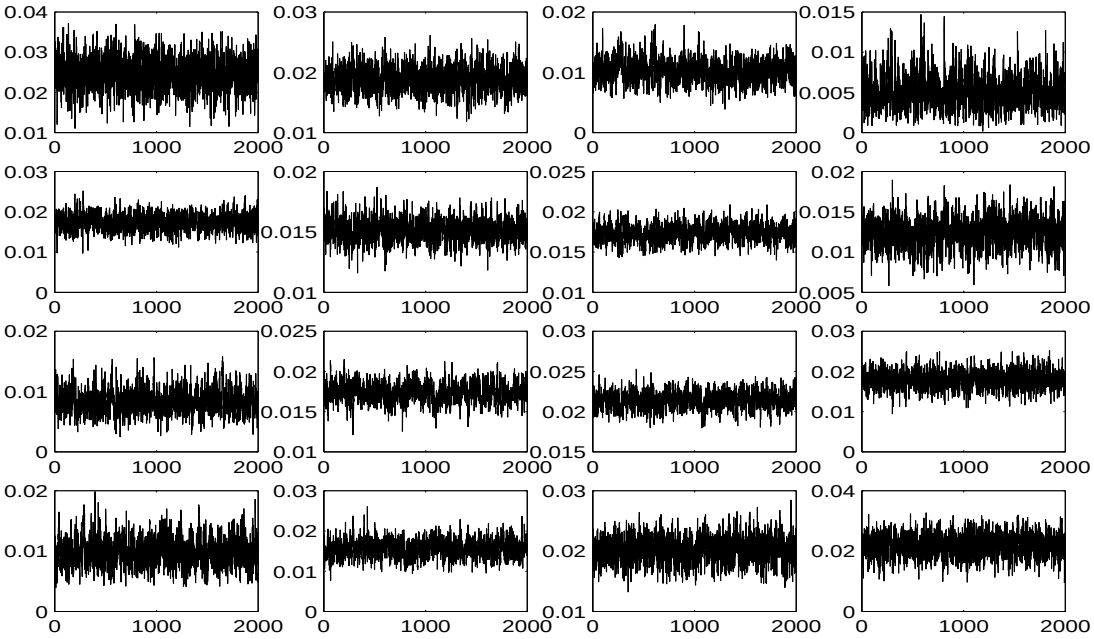


Figure 14: Proposed paths with one internal node and extreme observations. 1000 proposed paths sampled from the proposal distributions plotted in Figure 13 in the nonlinear step in the CTGM algorithm for extreme observations (a) $t = 50.0\,\text{ms}$; (b) $t = 100\,\text{ms}$.



(a)



(b)

Figure 15: Mixing plot for the Markov chain used in nonlinear inversion for one internal node and 16 observations. The path parameters i.e. $\boldsymbol{\gamma}$ (a); the second derivative i.e. $\boldsymbol{h}$ (b). In both (a) and (b) the ordering of the figures is such that an increasing column number correspond to increasing depth of starting point of the path. Increasing row number correspond to increasing depth of end point of the path.

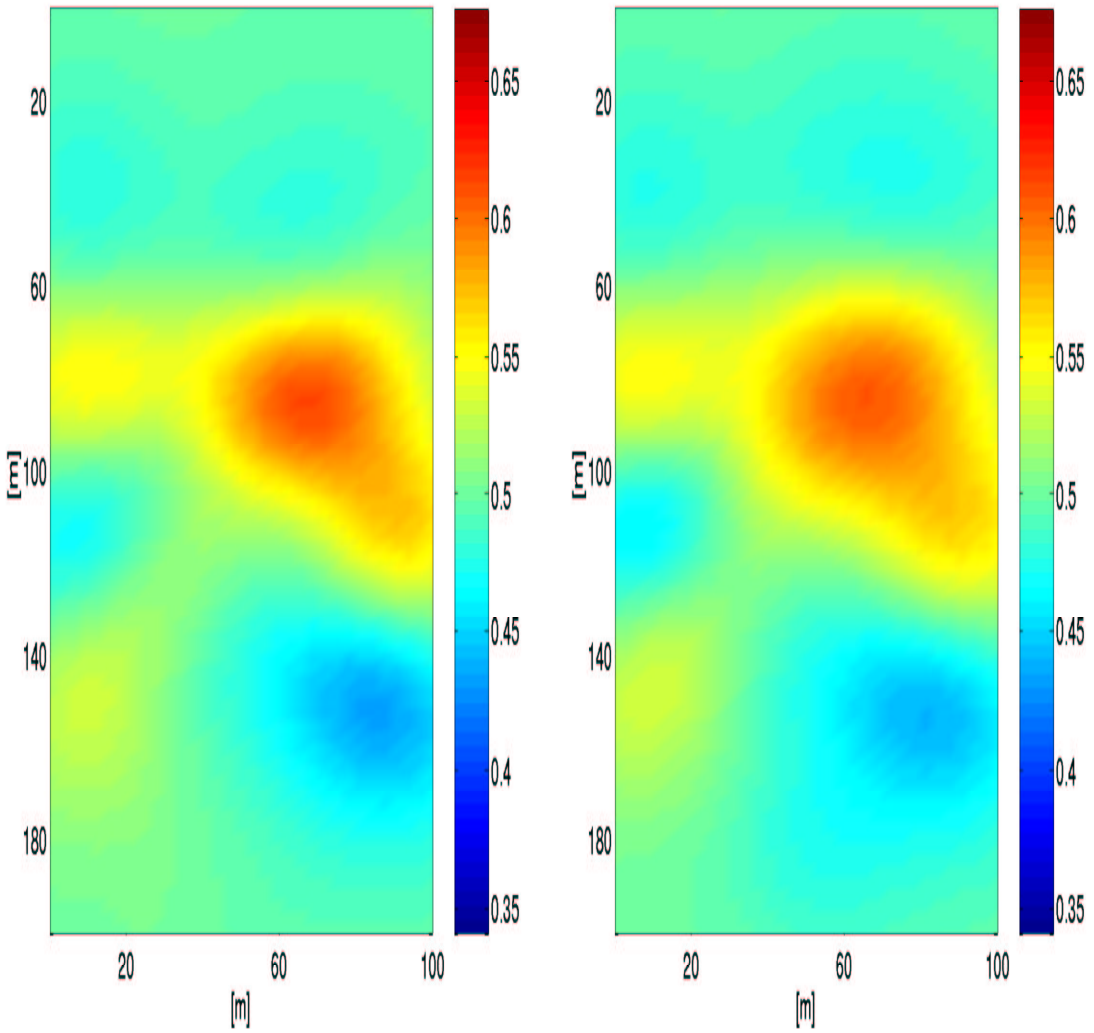


Figure 16: Comparison of estimates 16 observations. The conditional expectation using (a) linear tomography; (b) nonlinear tomography one internal node.