SIAM/ASA J. Uncertainty Quantification
Vol. 9, No. 2, pp. 420–445

Bayesian Spatial Inversion and Conjugate Selection Gaussian Prior Models*

Henning Omre† and Kjartan Rimstad†

Abstract. We study conjugate prior models in Bayesian spatial inversion. The spatial Kriging model may be phrased in a conjugate Bayesian inversion setting with a Gaussian prior model and a Gauss-linear likelihood function, resulting in a Gaussian posterior model. Spatial variables with unimodal, symmetric spatial histograms can be represented by this Kriging model. We generalize this Gaussian prior model by a selection mechanism, and this selection Gaussian prior model may represent multimodal, skewed, and/or peaked spatial variables. Also this selection Gaussian prior model is conjugate with respect to Gauss-linear likelihood functions. Hence the posterior model is selection Gaussian and analytically tractable. Efficient algorithms for simulation of and prediction in the selection Gaussian posterior model are defined. Model parameter inference in a maximum likelihood setting, which is simplified by the conjugate property, is also discussed. Moreover, we demonstrate that any conjugate prior model can be generalized by selection and still remain conjugate with respect to the actual likelihood function. Lastly, a seismic inversion case study is presented, and improvements of 20–40% in prediction mean-square-error, relative to traditional Gaussian inversion, are found.

Key words. inverse problems, spatial prediction, conditional simulation, spatial inference, seismic inversion

AMS subject classification. 62H11

DOI. 10.1137/19M1302995

1. Introduction.

Inversion constitutes a challenge in many mathematical engineering problems. Observations from the variable of interest are often indirectly collected by some complex acquisition device. The objective is naturally to predict the variable of interest based on the available observations. We consider spatial variables in this study and examples of inverse problems can be found in geophysics, image analysis, and remote sensing. Inversion of seismic data is presented as a case study later in the paper.

The spatial variable of interest is $\{r(\mathbf{x}); \mathbf{x} \in \mathcal{D} \subset \mathcal{R}^m\}$ with $r(\mathbf{x}) \in \mathcal{R}$ being the variable, having spatial reference $\mathbf{x}$ running over the reference domain $\mathcal{D} \subset \mathcal{R}^m$, which naturally has dimension m equal to one, two, or three. The variable is discretized $\{r(\mathbf{x}); \mathbf{x} \in \mathcal{L}_\mathcal{D}\}$ where $\mathcal{L}_\mathcal{D}$ is a regular grid, of size n_r, covering $\mathcal{D}$, and the spatial variable is represented by the n_r-vector $\mathbf{r} \in \mathcal{R}^{n_r}$. Assume further that an n_d-vector of observations $\mathbf{d} \in \mathcal{R}^{n_d}$ related to the variable of interest is collected. The focus is on assessing $\mathbf{r}$ given $\mathbf{d}$, $[\mathbf{r}|\mathbf{d}]$.

*Received by the editors December 2, 2019; accepted for publication (in revised form) January 26, 2021; published electronically April 21, 2021.
https://doi.org/10.1137/19M1302995
Funding: This research is part of the Uncertainty in Reservoir Evaluation (URE) activity at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

†Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim 7491, Norway (henning.omre@ntnu.no, rimstad@gmail.com).

We phrase the assessment in a probabilistic setting by using Bayesian inversion (see [30]):

$$[\mathbf{r}|\mathbf{d}] \to f(\mathbf{r}|\mathbf{d}) = [\int f(\mathbf{d}|\mathbf{r})f(\mathbf{r})d\mathbf{r}]^{-1} \times f(\mathbf{d}|\mathbf{r})f(\mathbf{r}) = \text{const} \times f(\mathbf{d}|\mathbf{r})f(\mathbf{r}), \tag{1.1}$$

where $\mathbf{y} \to f(\mathbf{y})$ reads as the random variable $\mathbf{y}$ is distributed according to the probability density function (pdf) $f(\mathbf{y})$. The $f(\mathbf{r}|\mathbf{d})$ is the posterior pdf being the ultimate solution of Bayesian inversion. The likelihood function $f(\mathbf{d}|\mathbf{r})$, being a function of $\mathbf{r}$, represents the observation acquisition procedure, while the prior pdf $f(\mathbf{r})$ summarizes prior information about the spatial variable of interest. The likelihood and prior models uniquely define the posterior model, although the integral in the normalizing constant usually is complicated to calculate.

Bayesian inversion is defined in a predictive setting with focus on the spatial variable. Classical Bayesian inference (see [13]) has another focus, namely model parameter estimation. The Bayesian inference model is often cast in a conjugate parametric framework for which the posterior model can be determined analytically; see [13]. Hence complicated integral calculations can be avoided. We use this conjugate model construction in the Bayesian inversion model defined in the current study.

In Bayesian spatial inversion, the focus is on the discretized spatial variable represented by the n_r-vector $\mathbf{r}$. The associated observations in the n_d-vector $\mathbf{d}$ are assumed to be acquired by a likelihood function being linear in $\mathbf{r}$ with additive centered Gaussian error term. This likelihood function is in the Gauss-linear class. In subsurface data acquisition, medical imaging, and remote sensing, the observations often appear as contrasts and/or with spatial convolutions which can be represented by Gauss-linear likelihood models. By assigning a prior pdf to $\mathbf{r}$ from the Gaussian class, one can easily demonstrate that the posterior pdf also will be Gaussian. The corresponding model parameters can be analytically calculated from the model parameters of the likelihood and prior models and the actual observed values. Hence the Gaussian class of prior models is conjugate with respect to Gauss-linear likelihood models. The traditional spatial Kriging model (see [15]) can be cast in this Bayesian spatial inversion setting, and its analytical tractability is one major reason for the widespread use of Kriging prediction.

The Kriging model has Gaussian marginal pdfs and can represent spatial variables with approximately unimodal, symmetric spatial histograms. In many spatial applications one observes multimodal spatial histograms often caused by physical effects, categorical latent variables, or extreme spatial heterogeneity. Examples are oil/water saturation in the subsurface with only partial mixing of fluids, density in brain substance with varying latent tissue classes, and pollution monitoring in an area with some emission sources. In the current study we generalize the Gaussian spatial model to define a class of spatial prior models which can represent these kinds of multimodal spatial variables. We demonstrate that for a Gauss-linear likelihood function, the class of Gaussian prior models can be generalized by a selection mechanism, inspired by the model discussed in [2], and still remain a conjugate class. This construction makes it possible to define highly flexible prior spatial models, which may represent spatial variables with multimodal, skewed, and/or peaked spatial histograms.

The developments of the selection Gaussian prior model is inspired by the early work on skewed pdfs in [6] and [8]; see also [20] and [7]. These models are generalized to spatial

settings in [25], [1], and [28]. In the current study we use generalized selection sets as discussed in [4] and [2] to construct spatial prior pdfs with marginal multimodality, skewness, and/or peakedness.

Modeling spatial variables with multimodal marginal characteristics and spatial continuity is challenging. The selection Gaussian spatial model discussed above is not the only model option, however. Trans-Gaussian random fields (see [15]) based on univariate transformations of the marginals may represent multimodality. For these models also the observations must be transformed before conditioning, however, and for likelihood functions with convolutions and observation errors, consistent transformations are not defined. Alternatively, mixture Gaussian random fields containing a latent categorical random field as mode indicator can be used to model multimodality. The latent categorical model can either be a Markov random field (see [9], [27], [10], and [17]) or a level set random field (see [16] and [24]). These categorical spatial models rely heavily on assessment by Markov chain Monte Carlo (McMC) simulation, since the analytical tractability is very limited. For likelihood models including spatial convolutions, the observations usually have wide spatial support and are often highly coupled. In these settings spatial McMC algorithms tend to have poor convergence rates. The selection Gaussian spatial model defined in this study is parametrized differently from the mixture Gaussian spatial model, since it does not explicitly include a latent categorical variable. The former model has more flexibility than the latter, since it can represent spatial variables with multimodal, skewed, and/or peaked spatial histograms. Moreover, the selection Gaussian models define a conjugate class with respect to Gauss-linear likelihood functions, which make the posterior model selection Gaussian with parameters analytically available. The analytical tractability makes it possible to define tailored simulation algorithms.

The selection construction presented above for the class of Gaussian pdfs can be generalized. For a given class of likelihood models, the selection mechanism can be used on the corresponding class of conjugate prior models, if it exists, and the resulting class of selection prior models will remain conjugate. Hence the conjugate characteristic appears as invariant to the selection operator. In the following sections we define the selection concept in a general setting and demonstrate the flexibility for the Gaussian spatial models.

Bayesian predictive inversion and Bayesian parameter inference can be combined to have hierarchical Bayesian inversion. It is, however, complicated to define general conjugate classes of prior models in this setting; see [29] and [11]. We present a brief discussion and evaluation of classical likelihood inference of the model parameters in the selection Gaussian prior pdf, based on the available observations or one training image.

The major contribution of the current paper is, however, the discussion of conjugate prior models in Bayesian spatial inversion and the demonstration that this conjugate characteristic is closed under activation of a selection mechanism. This result holds for all types of conjugate prior models including the prior Gaussian spatial model frequently used in practice. This spatial selection Gaussian model may appear with multimodal, skewed, and/or peaked marginal pdfs as presented in a variety of examples. Model parameter inference based on maximum likelihood is defined and evaluated. Moreover, we present a case study of seismic inversion into subsurface elastic properties being spatial variables with bimodal spatial histograms.

In the presentation $f(\mathbf{y})$ denotes a pdf of the random variable $\mathbf{y}$, while $F(\mathbf{y} \in \mathcal{B})$ denotes the probability that the random variable is in the subset $\mathcal{B}$ of its sample space. For the

Gaussian random n-vector $\mathbf{y}$ we write

$$\begin{aligned}(1.2)\qquad \mathbf{y} \to f(\mathbf{y}) &= \phi_n(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &= [2\pi]^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\Big\{-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^T\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]\Big\}, \\ F(\mathbf{y}\in\mathcal{B}) &= \Phi_n(\mathcal{B};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_{\mathcal{B}}\phi_n(\mathbf{u};\boldsymbol{\mu},\boldsymbol{\Sigma})d\mathbf{u}.\end{aligned}$$

Note that from a computational point of view $|\boldsymbol{\Sigma}|^{-1}$ and $\boldsymbol{\Sigma}^{-1}$ are demanding, and so are simulation from and calculation of the subset $\mathcal{B}$ probability of an arbitrary high-dimensional Gaussian pdf. The former two challenges are widely studied, however, while the latter challenge has drawn much less attention; see one basic and one recent reference, [21] and [19]. Later, we suggest some algorithms for simulation from and calculation of these subset $\mathcal{B}$ probabilities. We also use the notation $\mathbf{i}_n$ for a unit n-vector and $\mathbf{I}_n$ for an identity $(n\times n)$-matrix, while $I(A)$ is an indicator function taking value 1 if A is true and 0 otherwise.

In section 2, Bayesian spatial inversion is discussed, and a selection extended conjugate class of prior pdfs for a given class of likelihood functions is defined. The conjugate class of selection Gaussian prior pdfs is developed and discussed in detail. Section 3 contains a discussion of model parameter inference and the development of a maximum likelihood estimator for the model parameters of the conjugate class of selection Gaussian prior pdfs, based on the actual observations or a training realization of the spatial variable. In section 4, a case study based on real seismic data along a well profile from the Alvheim field in the North Sea is presented. Comparisons with regular Bayesian Gaussian inversion are made. Lastly, in section 5, the conclusions of the study are forwarded. In Appendix A some useful characteristics of the selection Gaussian model are demonstrated.

2. Bayesian spatial inversion.

The focus is on prediction of a spatial variable discretized into the n_r-vector $\mathbf{r}$, based on the observations represented in the n_d-vector $\mathbf{d}$. We phrase the prediction as Bayesian inversion (see (1.1)), which requires that the likelihood function $f(\mathbf{d}|\mathbf{r})$ is given, and that the prior pdf $f(\mathbf{r})$ is specified. Hereby, the corresponding posterior pdf $f(\mathbf{r}|\mathbf{d})$ is defined. Along the lines of the definition of conjugate prior pdfs in traditional Bayesian inference (see [13]), we present the following definition,

Definition 1 (conjugate class of prior pdfs). *Consider Bayesian inversion*

$$f(\mathbf{r}|\mathbf{d};\boldsymbol{\theta}_{r|d}) = const\times f(\mathbf{d}|\mathbf{r};\boldsymbol{\psi}_d)f(\mathbf{r};\boldsymbol{\theta}_r)$$

with likelihood function $f(\mathbf{d}|\mathbf{r};\boldsymbol{\psi}_d)$ in a parametrized pdf class $\mathcal{L}_\psi$ and prior pdf $f(\mathbf{r};\boldsymbol{\theta}_r)$ in a parametrized pdf class $\mathcal{P}_\theta$. If the associated posterior pdf $f(\mathbf{r}|\mathbf{d};\boldsymbol{\theta}_{r|d})$ also is in the pdf class $\mathcal{P}_\theta$, then the pdf class $\mathcal{P}_\theta$ is termed a conjugate class with respect to the likelihood function class $\mathcal{L}_\psi$. The parameters of the posterior model $\boldsymbol{\theta}_{r|d}$ will be a function of $[\boldsymbol{\psi}_d,\boldsymbol{\theta}_r,\mathbf{d}]$.

For continuous spatial variables, the class of Gaussian prior pdfs is known to be a conjugate class with respect to Gauss-linear likelihood functions. Hence, if the observations are collected through a linear forward model with additive Gaussian errors, and the prior pdf is specified to be Gaussian, then the posterior pdf will also be Gaussian. This characteristic is the basis for

Kriging prediction and conditional simulation in geostatistics; see [14]. Moreover, for event spatial variables the Poisson prior pdf is conjugate with respect to thinning likelihood functions (see [23]), while for mosaic spatial variables the Markov prior pdf is conjugate with respect to conditionally independent single-site response likelihood functions; see [22]. These conjugate characteristics are of course the major reason for the frequent use of these spatial models. In the next section we define an extended class of prior pdfs based on a selection concept and demonstrate that this concept can be used to construct an extended conjugate class of prior pdfs.

2.1. Generalization by selection. Consider the previously defined spatial variable represented by the n_r-vector $\mathbf{r}$ with prior pdf $f(\mathbf{r})$. Extend the dimension by an auxiliary random n_ν-vector $\boldsymbol{\nu} \in \mathcal{R}^{n_\nu}$, such that

$$\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \end{bmatrix} \rightarrow f\left(\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \end{bmatrix} \right) = f(\boldsymbol{\nu}|\mathbf{r}) f(\mathbf{r}) \tag{2.1}$$

with arbitrarily chosen pdf $f(\boldsymbol{\nu}|\mathbf{r})$, and denote the pdf $f(\mathbf{r})$ the basis-pdf. Consider an arbitrary subset $\mathcal{A}_\nu \subset \mathcal{R}^{n_\nu}$ and define the associated random selection n_r-vector $\mathbf{r}_A$ by

$$\begin{aligned} \mathbf{r}_A = [\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu] \rightarrow f(\mathbf{r}_A) &= f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu) \\ &= [F(\boldsymbol{\nu} \in \mathcal{A}_\nu)]^{-1} \times F(\boldsymbol{\nu} \in \mathcal{A}_\nu|\mathbf{r}) f(\mathbf{r}), \end{aligned} \tag{2.2}$$

which is constructed by conditioning $[\mathbf{r}, \boldsymbol{\nu}]$ on $\boldsymbol{\nu} \in \mathcal{A}_\nu$ and thereafter marginalizing to $\mathbf{r}$. Note in particular that $f(\mathbf{r}_A) = f(\mathbf{r})$ if we define $f(\boldsymbol{\nu}|\mathbf{r}) = f(\boldsymbol{\nu})$, which of course is the extreme choice of independence between $\mathbf{r}$ and $\boldsymbol{\nu}$. Based on this selection concept we have the following definition.

Definition 2 (selection extension of prior pdfs). *Consider a prior basis-pdf $f(\mathbf{r}; \boldsymbol{\theta}_r)$ in a parametrized pdf class $\mathcal{P}_\theta$, and define auxiliary variable $\boldsymbol{\nu} \in \mathcal{R}^{n_\nu}$ related to $\mathbf{r}$ by pdf $f(\boldsymbol{\nu}|\mathbf{r}; \boldsymbol{\kappa}_\nu)$ in a parametrized pdf class $\mathcal{E}_\kappa$. Specify further a selection set $\mathcal{A}_\nu \subset \mathcal{R}^{n_\nu}$. Define the selection variable,*

$$\begin{aligned} \mathbf{r}_A = [\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu] \rightarrow f(\mathbf{r}_A) &= f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu; \boldsymbol{\theta}_r, \boldsymbol{\kappa}_\nu) \\ &= [F(\boldsymbol{\nu} \in \mathcal{A}_\nu; \boldsymbol{\theta}_r, \boldsymbol{\kappa}_\nu)]^{-1} \times F(\boldsymbol{\nu} \in \mathcal{A}_\nu|\mathbf{r}; \boldsymbol{\kappa}_\nu) f(\mathbf{r}; \boldsymbol{\theta}_r), \end{aligned}$$

with pdf $f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu; \boldsymbol{\theta}_r, \boldsymbol{\kappa}_\nu)$ in the parametrized selection extended pdf class $\mathcal{S}_{\mathcal{A}_\nu}[\mathcal{P}_\theta \times \mathcal{E}_\kappa]$.

The selection extension can be made for any basis-pdf class for arbitrary auxiliary variables with associated selection sets. The class of selection Gaussian pdfs with $f(\mathbf{r})$ from the Gaussian class and $f(\boldsymbol{\nu}|\mathbf{r})$ being Gauss-linear with associated selection sets, and hence $[\mathbf{r}, \boldsymbol{\nu}]$ being jointly Gaussian, is thoroughly discussed in [2]. We define this class of selection Gaussian pdfs by a Gaussian basis-pdf

$$\mathbf{r} \rightarrow f(\mathbf{r}) = \phi_{n_r}(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$$

with the expectation n_r-vector $\boldsymbol{\mu}_r$ and the covariance $(n_r \times n_r)$-matrix $\boldsymbol{\Sigma}_r$ being model parameters. The auxiliary n_ν-vector $\boldsymbol{\nu}$ is defined as

$$[\boldsymbol{\nu}|\mathbf{r}] \rightarrow f(\boldsymbol{\nu}|\mathbf{r}) = \phi_{n_\nu}(\boldsymbol{\nu}; \boldsymbol{\mu}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r})$$

with the conditional expectation n_ν-vector being linear in $\mathbf{r}$, $\boldsymbol{\mu}_{\nu|r} = \boldsymbol{\mu}_\nu + \boldsymbol{\Gamma}_{\nu|r}(\mathbf{r} - \boldsymbol{\mu}_r)$ with expectation n_ν-vector $\boldsymbol{\mu}_\nu$, coupling $(n_\nu \times n_r)$-matrix $\boldsymbol{\Gamma}_{\nu|r}$, and the conditional covariance $(n_r \times n_\nu)$-matrix $\boldsymbol{\Sigma}_{\nu|r}$ all being model parameters of the selection Gaussian pdf. Hence $[\boldsymbol{\nu}|\mathbf{r}]$ is Gauss-linear, and since $\mathbf{r}$ is Gaussian, the joint $(n_r + n_\nu)$-vector $[\mathbf{r}, \boldsymbol{\nu}]$ is Gaussian. By enforcing the selection $\boldsymbol{\nu} \in \mathcal{A}_\nu \subset \mathcal{R}^{n_\nu}$, we obtain the selection Gaussian n_r-vector $\mathbf{r}_A$,

$$\begin{aligned}(2.3)\quad \mathbf{r}_A = [\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu] \to f(\mathbf{r}_A) &= f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu) \\ &= [\Phi_{n_\nu}(\mathcal{A}_\nu; \boldsymbol{\mu}_\nu, \boldsymbol{\Sigma}_\nu)]^{-1} \times \Phi_{n_\nu}(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r}) \phi_{n_r}(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r),\end{aligned}$$

where the marginal covariance $(n_\nu \times n_\nu)$-matrix is $\boldsymbol{\Sigma}_\nu = \boldsymbol{\Gamma}_{\nu|r} \boldsymbol{\Sigma}_r \boldsymbol{\Gamma}_{\nu|r}^T + \boldsymbol{\Sigma}_{\nu|r}$. All valid sets of model parameters $(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r, \boldsymbol{\mu}_\nu, \boldsymbol{\Gamma}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r}, \mathcal{A}_\nu)$ define the class of selection Gaussian pdfs. Note that by assigning $\boldsymbol{\Gamma}_{\nu|r}$ a null-matrix or setting $\mathcal{A}_\nu = \mathcal{R}^{n_\nu}$, the selection Gaussian class of pdfs is identical to the Gaussian one.

In Figure 1 several examples of selection Gaussian pdfs are displayed. The top row contains univariate selection Gaussian pdfs with $n_r = n_\nu = 1$. The displays present the auxiliary $[r, \nu]$-variable, being standard bi-Gaussian with correlation 0.9, and varying selection sets $\mathcal{A}_\nu$. The corresponding selection Gaussian pdfs is presented inverted on the horizontal axis. The left display contains a one-sided selection, providing a selection Gaussian pdf that is skewed. The selection set in the middle display consists of two subsets, resulting in a bimodal pdf. Similarly, the right display contains three selection subsets providing a trimodal pdf. The asymmetry of the pdfs is governed by the relative shifts of the selection sets $\mathcal{A}_\nu$ and $\mu_\nu = 0$. The bottom row contains bivariate selection Gaussian pdfs with $n_r = n_\nu = 2$ and identical marginal pdfs from the corresponding top row display. The $[r_1, r_2]$ correlation is 0.1. The left display contains a selection Gaussian pdf that is skewed. The pdfs in the two other displays are multimodal, and the number of modes is the number in their marginal pdfs squared. Figure 1 demonstrates the flexibility of the selection Gaussian class of pdfs, which can represent multimodal, skewed, and/or peaked variables. For a spatial variable represented by the n_r-vector $\mathbf{r}$ having multimodal marginals, the number of modes in its selection Gaussian model will obviously increase quickly with increasing grid-size n_r.

The likelihood function of the actual observations $\mathbf{d}^o$, $f(\mathbf{d}^o|\mathbf{r})$, is a function of $\mathbf{r}$ and only dependent on the observation acquisition procedure, independent of the choice of prior pdf. Hence, the posterior pdf based on a prior selection pdf can be expressed as

$$\begin{aligned}(2.4)\qquad [\mathbf{r}_A|\mathbf{d}^o] \to f(\mathbf{r}_A|\mathbf{d}^o) &= \text{const} \times f(\mathbf{d}^o|\mathbf{r}_A) f(\mathbf{r}_A) \\ &= \text{const}_1 \times f(\mathbf{d}^o|\mathbf{r}) f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu) \\ &= \text{const}_2 \times f(\mathbf{d}^o|\mathbf{r}) F(\boldsymbol{\nu} \in \mathcal{A}_\nu|\mathbf{r}) f(\mathbf{r}) \\ &= \text{const}_3 \times F(\boldsymbol{\nu} \in \mathcal{A}_\nu|\mathbf{r}, \mathbf{d}^o) f(\mathbf{r}|\mathbf{d}^o) \\ &= [F(\boldsymbol{\nu} \in \mathcal{A}_\nu|\mathbf{d}^o)]^{-1} \times F(\boldsymbol{\nu} \in \mathcal{A}_\nu|\mathbf{r}, \mathbf{d}^o) f(\mathbf{r}|\mathbf{d}^o),\end{aligned}$$

which relies on the conditional independence relation $f(\boldsymbol{\nu}, \mathbf{d}|\mathbf{r}) = f(\boldsymbol{\nu}|\mathbf{r}) f(\mathbf{d}|\mathbf{r})$. Note that the posterior pdf corresponds to the selection pdf with basis-pdf being the conditional pdf $f(\mathbf{r}|\mathbf{d})$, which provides the following theorem.

Theorem 1 (selection extended conjugate class of prior pdfs). *Consider a likelihood function in parametrized pdf class $\mathcal{L}_\psi$ and a prior pdf in the parametrized pdf class $\mathcal{P}_\theta$. According to*

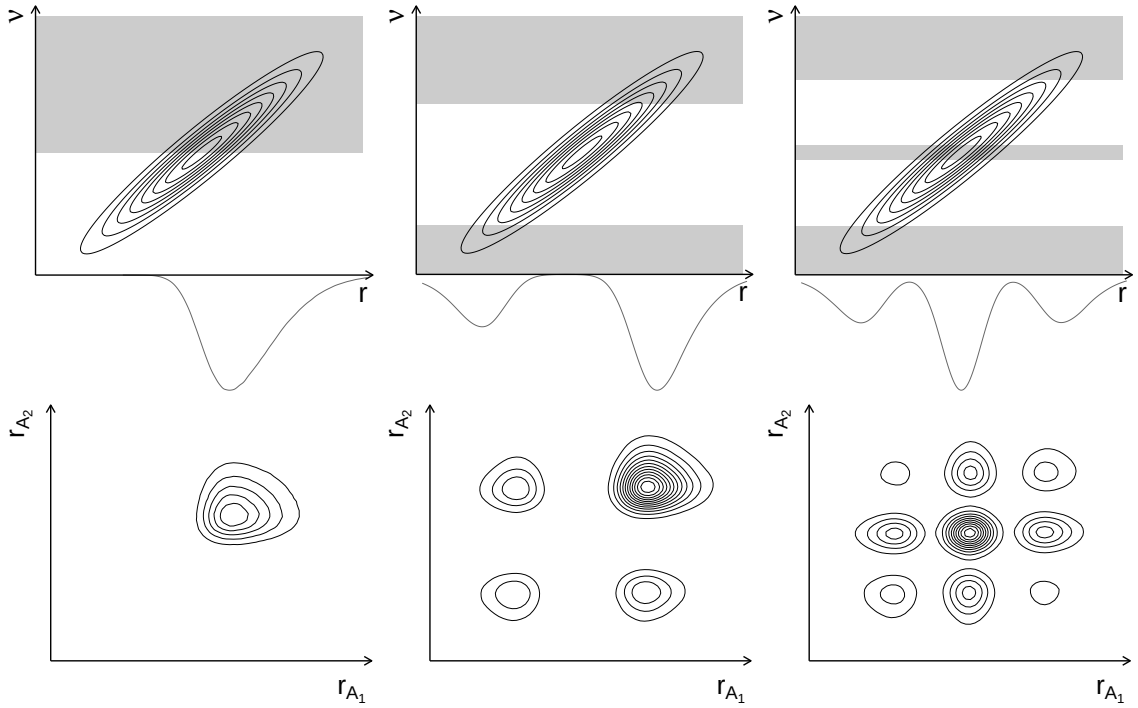

Figure 1. *Examples of univariate (top row) and bivariate (bottom row) selection Gaussian pdfs with varying selection sets (gray).*

Definition 1*, let the pdf class* $\mathcal{P}_\theta$ *be a conjugate class with respect to likelihood function class* $\mathcal{L}_\psi$*. According to Definition* 2*, define the associated selection extended pdf class* $\mathcal{S}_{\mathcal{A}_\nu}[\mathcal{P}_\theta \times \mathcal{E}_\kappa]$ *based on auxiliary pdf class* $\mathcal{E}_\kappa$ *and selection set* $\mathcal{A}_\nu$*. Then the pdf class* $\mathcal{S}_{\mathcal{A}_\nu}[\mathcal{P}_\theta \times \mathcal{E}_\kappa]$ *is a conjugate class with respect to the likelihood function class* $\mathcal{L}_\psi$ *for all pdf classes* $\mathcal{E}_\kappa$ *and selection sets* $\mathcal{A}_\nu$*. The conjugate characteristics of a prior pdf class* $\mathcal{P}_\theta$ *are closed under selection extension.*

This closedness property for conjugate pdf classes is very general and applies to continuous, event and mosaic spatial variables. Moreover, it may be used in traditional Bayesian inference. For continuous spatial variables with a Gauss-linear likelihood function, prior pdfs from the class of Gaussian pdfs are known to be conjugate. According to the results above, the selection Gaussian pdf for any arbitrary auxiliary extension and selection set will define a conjugate class of pdfs with respect to a Gauss-linear likelihood function. In the following subsections we will explore this opportunity to define more flexible prior pdfs in Bayesian spatial inversion with observations collected through a Gauss-likelihood function.

2.2. Likelihood model. We limit the likelihood function to be from the Gauss-linear class,

$$[\mathbf{d}|\mathbf{r}] = \mathbf{H}\mathbf{r} + \mathbf{e}_{d|r} \rightarrow f(\mathbf{d}|\mathbf{r}) = \phi_{n_d}(\mathbf{d}; \mathbf{H}\mathbf{r}, \boldsymbol{\Sigma}_{d|r}), \tag{2.5}$$

where $\mathbf{H}$ is an observation acquisition $(n_d \times n_r)$-matrix and $\mathbf{e}_{d|r}$ is a centered Gaussian error n_d-vector with covariance $(n_d \times n_d)$-matrix $\boldsymbol{\Sigma}_{d|r}$ independent of $\mathbf{r}$. Hence the model parameters are $\boldsymbol{\theta}_l = (\mathbf{H}, \boldsymbol{\Sigma}_{d|r})$. There are no constraints on the matrix $\mathbf{H}$, and it may contain features of

the observation acquisition procedure that need to be estimated. It may be binary, selecting only a subset of elements in the vector $\mathbf{r}$, or it could represent convolution by averaging over elements in the vector $\mathbf{r}$. Also numerical differentiation or integration can be captured by $\mathbf{H}$. In the case study we use a convolved, contrast-linearized approximation to the wave equation to model acquisition of seismic data; see [12]. The seismic convolution kernel can often be complicated to infer since it depends on the elastic properties of the geological overburden.

2.3. Prior model. The spatial variable of interest is represented in the n_r-vector $\mathbf{r}$. We define the prior basis-pdf $f(\mathbf{r})$ to be a spatially stationary Gaussian pdf,

$$\mathbf{r} \to f(\mathbf{r}) = \phi_{n_r}(\mathbf{r}; \mu \mathbf{i}_{n_r}, \sigma^2 \mathbf{C}), \tag{2.6}$$

where the scalars (μ, σ^2) are the stationary expectation and variance, respectively, while the spatial correlation $(n_r \times n_r)$-matrix $\mathbf{C}$ is defined by the spatial translation invariant positive definite correlation function $\rho(\boldsymbol{\tau}); \boldsymbol{\tau} \in \mathcal{R}^m$, which is assumed to tend towards 0 as $|\boldsymbol{\tau}|$ increases. This pdf is spatially stationary in the sense that the marginal pdfs $f(r_i) = \phi_1(r_i; \mu, \sigma^2)$, $i = 1, \ldots, n_r$, are all identical. Moreover, it exhibits ergodicity since $f(r_i, r_j) \to f(r_i) f(r_j)$ as $|\mathbf{x}_i - \mathbf{x}_j| \to \infty$ since $\rho(\boldsymbol{\tau})$ tends towards 0 for increasing $|\boldsymbol{\tau}|$, which entails that consistent estimates of the model parameters can be obtained.

We use this Gaussian pdf as basis-pdf to define a selection pdf, with the auxiliary n_r-vector $\boldsymbol{\nu}$ extension,

$$\begin{aligned}[\boldsymbol{\nu}|\mathbf{r}] = \gamma\sigma^{-1}[\mathbf{r} - \mu\mathbf{i}_{n_r}] + \mathbf{e}_{\nu|r} \to f(\boldsymbol{\nu}|\mathbf{r}) &= \phi_{n_r}(\boldsymbol{\nu}; \gamma\sigma^{-1}[\mathbf{r} - \mu\mathbf{i}_{n_r}], [1-\gamma^2]\mathbf{I}_{n_r}) \\ &= \prod_{i=1}^{n_r} \phi_1(\nu_i; \gamma\sigma^{-1}[r_i - \mu], [1-\gamma^2]),\end{aligned}$$

where $\gamma \in [-1, 1] \subset \mathcal{R}$ is a coupling parameter while $\mathbf{e}_{\nu|r}$ is a centered Gaussian n_r-vector with independent elements with variance $[1-\gamma^2]$, independent of $\mathbf{r}$. The extended variable becomes jointly Gaussian,

$$\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \end{bmatrix} \to f\left(\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \end{bmatrix}\right) = \phi_{2n_r}\left(\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \end{bmatrix}; \begin{bmatrix} \mu\mathbf{i}_{n_r} \\ 0\mathbf{i}_{n_r} \end{bmatrix}, \begin{bmatrix} \sigma^2\mathbf{C} & \gamma\sigma\mathbf{C} \\ \gamma\sigma\mathbf{C} & \gamma^2\mathbf{C} + [1-\gamma^2]\mathbf{I}_{n_r} \end{bmatrix}\right), \tag{2.7}$$

and consists of two grid-discretized spatial variables with variances σ^2 and 1, respectively, with intercorrelation $\gamma\sigma$ and with identical spatial correlation function $\rho(\boldsymbol{\tau})$.

Define the selection set $\mathcal{A}^{n_r} \subset \mathcal{R}^{n_r}$ with $\mathcal{A} \subset \mathcal{R}$, and hence identical selection sets for each component in $\mathbf{r}$. The corresponding spatial selection Gaussian pdf, which belongs to the class of selection Gaussian pdfs (see (2.3)), is defined as

$$\begin{aligned}\mathbf{r}_A = [\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}] \to f(\mathbf{r}_A) &= f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}) \\ &= [F(\boldsymbol{\nu} \in \mathcal{A}^{n_r})]^{-1} \times F(\boldsymbol{\nu} \in \mathcal{A}^{n_r}|\mathbf{r}) f(\mathbf{r}) \\ &= [\Phi_{n_r}(\mathcal{A}^{n_r}; 0\mathbf{i}_{n_r}, \gamma^2\mathbf{C} + [1-\gamma^2]\mathbf{I}_{n_r})]^{-1} \\ &\times \prod_{i=1}^{n_r} \Phi_1(\mathcal{A}; \gamma\sigma^{-1}[r_i - \mu], [1-\gamma^2])\ \phi_{n_r}(\mathbf{r}; \mu\mathbf{i}_{n_r}, \sigma^2\mathbf{C})\end{aligned} \tag{2.8}$$

with the actual model parameters $\boldsymbol{\theta}_p = (\mu, \sigma^2, \gamma, \rho(\boldsymbol{\tau}), \mathcal{A})$. The spatial selection Gaussian prior model, represented by the n_r-vector $\mathbf{r}_A$, is closely related to a spatial Gaussian model, with (μ, σ^2) being parameters related to marginal centering and variability, respectively, and $\rho(\boldsymbol{\tau})$ being related to spatial smoothness. The model parameters γ and $\mathcal{A}$ define the deviations from Gaussianity of the marginal pdfs. If $\gamma = 0$ and/or $\mathcal{A}$ is identical to $\mathcal{R}$, the selection Gaussian model will be identical to the Gaussian prior basis-model. Hence the selection Gaussian prior model appears as a generalization of the Gaussian prior model.

Note in particular that the prior model for the spatial variable of interest, represented by the selection Gaussian pdf on the vector $\mathbf{r}$, is subject to the grid $\mathcal{L}_\mathcal{D}$. The selection Gaussian concept breaks down when the grid size tends to zero by infilling of the grid; see [26] and [28]. This lack of generality limits the use of the model, of course, but in many applications like geophysics, image analysis, and remote sensing, there is a natural choice of grid $\mathcal{L}_\mathcal{D}$ due to the observation acquisition procedure. Moreover, this limitation is shared by the categorical Markov random field model (see [9]), which has proven immensely useful in many segmentation applications. One consequence of this grid dependency is that the model parameter values will depend on the actual spatial discretization as it is for categorical Markov random fields. In section 3 we discuss this subject in greater detail.

The selection Gaussian class of pdfs can be shown to be closed under marginalization (see [2]), and the univariate marginal pdf is

$$
\begin{aligned}
r_{Ai} = [r_i|\boldsymbol{\nu} \in \mathcal{A}^{n_r}] \rightarrow & f(r_{Ai}) = f(r_i|\boldsymbol{\nu} \in \mathcal{A}^{n_r}) \\
& = [F(\boldsymbol{\nu} \in \mathcal{A}^{n_r})]^{-1} \times F(\boldsymbol{\nu} \in \mathcal{A}^{n_r}|r_i) f(r_i) \\
& = [\Phi_{n_r}(\mathcal{A}^{n_r}; 0\mathbf{i}_{n_r}, \gamma^2\mathbf{C} + [1-\gamma^2]\mathbf{I}_{n_r})]^{-1} \\
& \times \Phi_{n_r}(\mathcal{A}^{n_r}; \gamma\sigma^{-1}\mathbf{c}_i(r_i - \mu), \gamma^2[\mathbf{C} - \mathbf{c}_i\mathbf{c}_i^T] + [1-\gamma^2]\mathbf{I}_{n_r}) \\
& \times \phi_1(r_i; \mu, \sigma^2),
\end{aligned}
\tag{2.9}
$$

which is a selection Gaussian pdf with n_r-vector $\mathbf{c}_i$ being the ith column of the correlation matrix $\mathbf{C}$. Note, however, that the two first moments of the marginal pdf, $\mathrm{E}(r_{Ai})$ and $\mathrm{Var}(r_{Ai})$, do not have nice closed-form expressions. All lower dimensional marginal pdfs will also be selection Gaussian pdfs.

All marginal selection Gaussian pdfs $f(r_{Ai}) = f(r_i|\boldsymbol{\nu} \in \mathcal{A}^{n_r})$, $i = 1, \ldots, n_r$, are conditioned on all n_r elements of the auxiliary variable $\boldsymbol{\nu}$—not only the respective ν_i. This conditioning causes the pdfs to be defined relative to the grid $\mathcal{L}_\mathcal{D}$. This coupling is contrary to marginal pdfs of discretized Gaussian random fields where all dependence of other dimensions are integrated out. Recall that the joint pdf $f(\mathbf{r}, \boldsymbol{\nu})$ is a discretized bivariate stationary Gaussian random field with spatial correlation function $\rho(\boldsymbol{\tau})$, and therefore $[r_i, \nu_i]$ and $[r_j, \nu_j]$ tend towards independence as their interdistance $|\boldsymbol{\tau}_{ij}|$ increases. Hence $f(r_{Ai})$ is primarily dependent on ν_j with $|\boldsymbol{\tau}_{ij}|$ being small. The prior selection Gaussian model is discretized on a regular grid $\mathcal{L}_\mathcal{D}$ with identical selection sets $\mathcal{A}$ for each node. Hence the marginal selection Gaussian pdfs $f(r_{Ai})$, $i = 1, \ldots, n_r$, appear as spatially stationary, except for grid nodes close to the boundary of $\mathcal{D}$. Moreover, the bivariate $f(r_{Ai}, r_{Aj})$ tends towards $f(r_{Ai})f(r_{Aj})$ as the interdistance $|\boldsymbol{\tau}_{ij}|$ increases since the coupling through $\boldsymbol{\nu}$ decreases with increasing $|\boldsymbol{\tau}_{ij}|$. Consequently, by introducing the auxiliary variable $\boldsymbol{\nu}$ in a spatial setting on the grid $\mathcal{L}_\mathcal{D}$, the prior selection

Table 1
Model parameters for six cases, with $\mu = 0$ and $\sigma^2 = 1$ for all cases.

Case	γ	d_h	d_v	A	Description
1	0.8000	2.0	2.0	$(\infty, -0.3] \cup [0.3, \infty)$	sym. bimodal iso.
2	0.6500	6.0	0.85	$(\infty, -0.3] \cup [0.3, \infty)$	sym. bimodal aniso.
3	0.9250	2.0	0.60	$(\infty, -0.85] \cup [0.8, \infty)$	asym. bimodal aniso.
4	0.9995	3.0	3.0	$[-0.45, -0.2] \cup [-0.1, 0.1] \cup [0.2, 0.45]$	sym. trimodal iso.
5	0.7000	2.0	2.0	$(\infty, -0.7] \cup [-0.1, 2.5]$	asym. unimodal iso.
6	0.7000	2.0	2.0	$(\infty, -1.75] \cup [-0.5, 0.5] \cup [1.75, \infty)$	sym. heavy tailed iso.

Gaussian pdf exhibits approximate stationarity and ergodicity in the sense defined above. These characteristics make the model suitable as a prior pdf in Bayesian spatial inversion. Moreover, by using the auxilary n_r-vector $\boldsymbol{\nu}$ flexibility is introduced in the prior model such that the posterior model can capture more of the information in the observations.

The prior selection Gaussian pdf will naturally be inspected by simulation. Simulation is performed sequentially by first generating a realization of the auxiliary variable $\boldsymbol{\nu}^s \in \mathcal{A}^{n_r}$ and thereafter generating the realization $\mathbf{r}_A^s = [\mathbf{r}|\boldsymbol{\nu}^s]^s$. Since the joint variable $[\mathbf{r}, \boldsymbol{\nu}]$ is Gaussian, many efficient algorithms are available; see section SM1 in the supplemental material.

The variety of the prior selection Gaussian model is exhibited in Figure 2, which is based on the model parameters listed in Table 1 with an anisotropic second-order exponential correlation function with anisotropy factor (d_h, d_v). A more detailed discussion of the example is given in section SM3 of the supplemental material. The spatial variable is represented on a (64×64)-grid. The anisotropy factors vary, and so do the correlation and the selection sets for the auxiliary variable. The selection set, for each case, defining the shape of the marginal pdfs of the prior model is marked by light gray bars on the horizontal axis. We observe a large variety of prior spatial models, all of them approximately stationary and ergodic in the sense discussed above. The prior models have marginal distributions that can be multimodal, skewed, peaked, or a combination of these features. The computer demand for generating one such realization is typically a couple of minutes on a regular laptop computer.

2.4. Posterior model. The posterior pdf is uniquely defined by the likelihood function and the prior pdf. With a likelihood function from the Gauss-linear class and a prior pdf from the selection Gaussian class, the posterior pdf, due to Theorem 1, will also be from the selection Gaussian class. The selection Gaussian class of prior pdfs is conjugate with respect to Gauss-linear likelihood functions. Hence the model parameters of the posterior pdf are analytically tractable based on the model parameters of the likelihood and prior models and the actual observations. The joint pdf is

(2.10)
$$\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \\ \mathbf{d} \end{bmatrix} \to f\left(\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \\ \mathbf{d} \end{bmatrix}\right)$$
$$= \phi_{2n_r+n_d}\left(\begin{bmatrix} \mathbf{r} \\ \boldsymbol{\nu} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \mu\mathbf{i}_{n_r} \\ 0\mathbf{i}_{n_r} \\ \mu\mathbf{H}\mathbf{i}_{n_r} \end{bmatrix}, \begin{bmatrix} \sigma^2\mathbf{C} & \gamma\sigma\mathbf{C} & \sigma^2\mathbf{C}\mathbf{H}^T \\ \gamma\sigma\mathbf{C} & \gamma^2\mathbf{C} + [1-\gamma^2]\mathbf{I}_{n_r} & \gamma\sigma\mathbf{C}\mathbf{H}^T \\ \sigma^2\mathbf{H}\mathbf{C} & \gamma\sigma\mathbf{H}\mathbf{C} & \sigma^2\mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r} \end{bmatrix}\right),$$

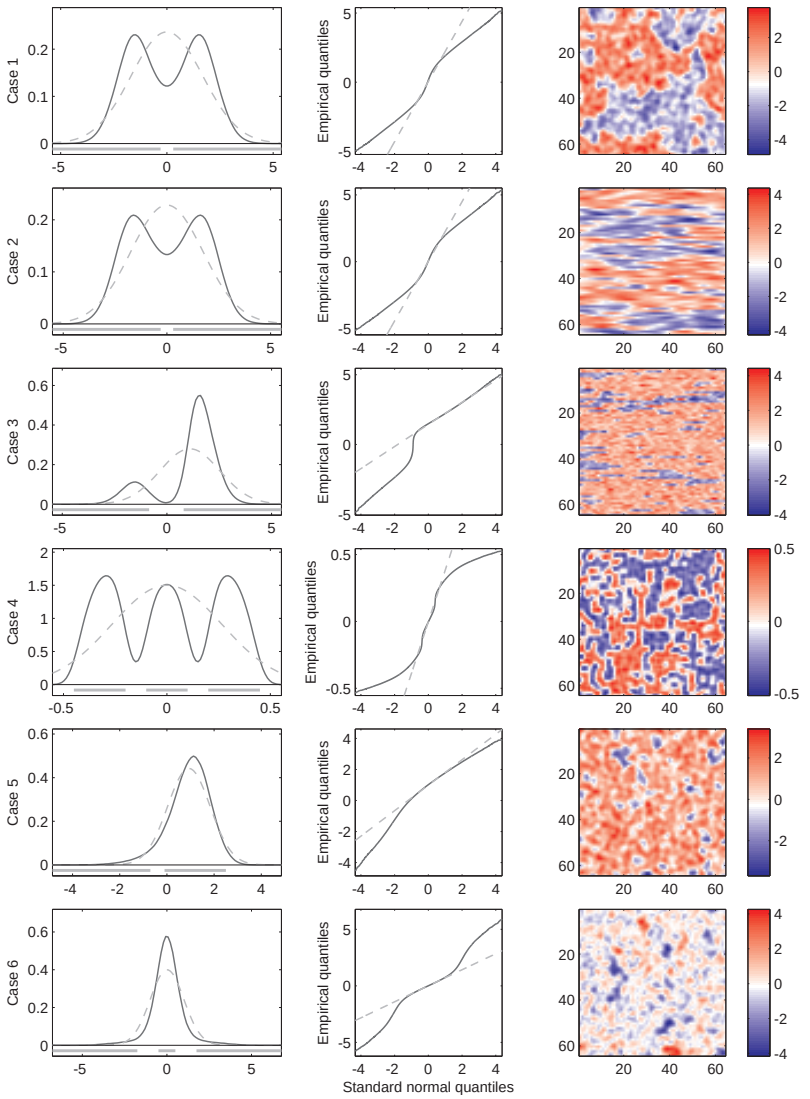

Figure 2. *First column: Marginal distribution of selection Gaussian random field (solid black), standard normal distribution (dashed gray), and selection sets on auxiliary random field on axis (solid gray). Second column: Quantile-quantile plot of marginal selection Gaussian random field versus theoretical quantiles from the Gaussian distribution. Third column: Realization from selection Gaussian random field.*

and one may demonstrate that $[\boldsymbol{\nu}, \mathbf{d}|\mathbf{r}]$ are conditionally independent. Note also that the joint $[\mathbf{r}, \mathbf{d}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}]$ will be selection Gaussian, and so will the two marginals $[\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}]$ and $[\mathbf{d}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}]$. Hence the marginal pdf of the observations will be dependent on the actual prior model, which the likelihood model will not. The focus of the study is on the posterior $[\mathbf{r}|\mathbf{d}, \boldsymbol{\nu} \in \mathcal{A}^{n_r}]$ which will be selection Gaussian as well; see Appendix A.

From Theorem 1 and standard Gaussian theory one has

$$
\begin{aligned}
(2.11)\qquad [\mathbf{r}_A|\mathbf{d}] = [\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}, \mathbf{d}] \rightarrow & f(\mathbf{r}_A|\mathbf{d}) = f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}, \mathbf{d}) \\
&= [F(\boldsymbol{\nu} \in \mathcal{A}^{n_r}|\mathbf{d})]^{-1} \times F(\boldsymbol{\nu} \in \mathcal{A}^{n_r}|\mathbf{r}, \mathbf{d}) f(\mathbf{r}|\mathbf{d}) \\
&= [\Phi_{n_r}(\mathcal{A}^{n_r}; \boldsymbol{\mu}_{\nu|d}, \boldsymbol{\Sigma}_{\nu|d})]^{-1} \\
&\times \Phi_{n_r}(\mathcal{A}^{n_r}; \boldsymbol{\mu}_{\nu|rd}, \boldsymbol{\Sigma}_{\nu|rd}) \\
&\times \phi_{n_r}(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d})
\end{aligned}
$$

with

$$
\begin{bmatrix} \boldsymbol{\mu}_{r|d} \\ \boldsymbol{\mu}_{\nu|d} \end{bmatrix} = \begin{bmatrix} \mu \mathbf{i}_{n_r} \\ 0 \mathbf{i}_{n_r} \end{bmatrix} + \begin{bmatrix} \sigma^2 \mathbf{C}\mathbf{H}^T \\ \gamma\sigma \mathbf{C}\mathbf{H}^T \end{bmatrix} \left[\sigma^2 \mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}\right]^{-1} [\mathbf{d} - \mu \mathbf{H}\mathbf{i}_{n_r}],
$$

$$
\begin{aligned}
\begin{bmatrix} \boldsymbol{\Sigma}_{r|d} & \boldsymbol{\Gamma}_{r\nu|d} \\ \boldsymbol{\Gamma}_{\nu r|d} & \boldsymbol{\Sigma}_{\nu|d} \end{bmatrix} = & \begin{bmatrix} \sigma^2 \mathbf{C} & \gamma\sigma \mathbf{C} \\ \gamma\sigma \mathbf{C} & \gamma^2 \mathbf{C} + [1 - \gamma^2] \mathbf{I}_{n_r} \end{bmatrix} \\
& - \begin{bmatrix} \sigma^2 \mathbf{C}\mathbf{H}^T \\ \gamma\sigma \mathbf{C}\mathbf{H}^T \end{bmatrix} \left[\sigma^2 \mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}\right]^{-1} \begin{bmatrix} \sigma^2 \mathbf{H}\mathbf{C} & \gamma\sigma \mathbf{H}\mathbf{C} \end{bmatrix},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\mu}_{\nu|rd} &= \boldsymbol{\mu}_{\nu|d} + \boldsymbol{\Gamma}_{vr|d} \boldsymbol{\Sigma}_{r|d}^{-1} [\mathbf{r} - \boldsymbol{\mu}_{r|d}], \\
\boldsymbol{\Sigma}_{\nu|rd} &= \boldsymbol{\Sigma}_{\nu|d} - \boldsymbol{\Gamma}_{\nu r|d} \boldsymbol{\Sigma}_{r|d}^{-1} \boldsymbol{\Gamma}_{r\nu|d}.
\end{aligned}
$$

This posterior pdf will of course be spatially nonstationary due to conditioning on the observations $\mathbf{d}$. The pdf will, however, be in the class of selection Gaussian pdfs (see (2.3)) and hence will be closed under marginalization and conditioning, with the corresponding model parameters analytically tractable. Assessment of the posterior pdf is usually made by simulation of realizations and locationwise prediction with associated precision intervals.

Simulation of realizations from the posterior pdf is made sequentially by first generating a realization $[\boldsymbol{\nu}^s \in \mathcal{A}^{n_r}|\mathbf{d}]$ and thereafter generating a realization $[\mathbf{r}_A|\mathbf{d}]^s = [\mathbf{r}|\boldsymbol{\nu}^s, \mathbf{d}]^s$. Since the joint variable $[\mathbf{r}, \boldsymbol{\nu}|\mathbf{d}]$ is Gaussian, many efficient algorithms are available, and the algorithm actually used in the current study is specified in section SM1 of the supplemental material with

$$
\begin{aligned}
(2.12)\qquad \boldsymbol{\nu}^s = [\boldsymbol{\nu}|\boldsymbol{\nu} \in \mathcal{A}^{n_r}, \mathbf{d}] &\rightarrow [\Phi_{n_r}(\mathcal{A}^{n_r}; \boldsymbol{\mu}_{\nu|d}, \boldsymbol{\Sigma}_{\nu|d})]^{-1} \times \phi_{n_r}(\boldsymbol{\nu}; \boldsymbol{\mu}_{\nu|d}, \boldsymbol{\Sigma}_{\nu|d}) \times I[\boldsymbol{\nu} \in \mathcal{A}^{n_r}], \\
[\mathbf{r}_A|\mathbf{d}]^s = [\mathbf{r}|\boldsymbol{\nu}^s, \mathbf{d}] &\rightarrow \phi_{n_r}(\mathbf{r}; \boldsymbol{\mu}_{r|\nu^s d}, \boldsymbol{\Sigma}_{r|\nu^s d}),
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\mu}_{r|\nu^s d} &= \boldsymbol{\mu}_{r|d} + \boldsymbol{\Gamma}_{r\nu|d} \boldsymbol{\Sigma}_{\nu|d}^{-1} [\boldsymbol{\nu}^s - \boldsymbol{\mu}_{\nu|d}], \\
\boldsymbol{\Sigma}_{r|\nu^s d} &= \boldsymbol{\Sigma}_{r|d} - \boldsymbol{\Gamma}_{r\nu|d} \boldsymbol{\Sigma}_{\nu|d}^{-1} \boldsymbol{\Gamma}_{\nu r|d}.
\end{aligned}
$$

Prediction of $[\mathbf{r}_A|\mathbf{d}]$ needs to be carefully designed since we often define selection Gaussian prior models with multiple modes, and so will the posterior pdf be. The traditional expectation (E) predictor based on a minimum locationwise squared error loss, denoted $\hat{\mathbf{r}}_E = \mathrm{E}\{\mathbf{r}_A|\mathbf{d}\}$,

will often appear in low-probability regions in between modes of the posterior pdf. The median (M) predictor based on a minimum locationwise absolute error criterion, denoted $\hat{\mathbf{r}}_M = MED\{\mathbf{r}_A|\mathbf{d}\}$, shares the same tendency to appear in low-probability regions. The preferred predictor is the global maximum posterior predictor, but it is usually too computer-demanding to determine since it requires optimization of an n_r-dimensional multimodal function. Therefore, we recommend the maximum posterior (MAP) predictor based on a maximum locationwise posterior criterion,

$$\begin{aligned}\hat{\mathbf{r}}_{MAP} &= \text{MAP}\{\mathbf{r}_A|\mathbf{d}\} \\ &= \{\text{MAP}\{r_{Aj}|\mathbf{d}\} = \underset{r_j}{\text{argmax}}\{f(r_j|\mathbf{d})\}; j = 1,\dots,n_r\},\end{aligned} \tag{2.13}$$

with

$$\begin{aligned}[r_{Ai}|\mathbf{d}] = [r_i|\boldsymbol{\nu} \in \mathcal{A}^{n_r}, \mathbf{d}] \rightarrow & f(r_{Ai}|\mathbf{d}) = f(r_i|\boldsymbol{\nu} \in \mathcal{A}^{n_r}, \mathbf{d}) \\ &= [F(\boldsymbol{\nu} \in \mathcal{A}^{n_r}|\mathbf{d})]^{-1} F(\boldsymbol{\nu} \in \mathcal{A}^{n_r}|r_i, \mathbf{d}) f(r_i|\mathbf{d}) \\ &= [\Phi_{n_r}(\mathcal{A}^{n_r}; \boldsymbol{\mu}_{\nu|d}, \boldsymbol{\Sigma}_{\nu|d})]^{-1} \\ &\times \Phi_{n_r}(\mathcal{A}^{n_r}; \boldsymbol{\mu}_{\nu|r_i d}, \boldsymbol{\Sigma}_{\nu|r_i d}) \\ &\times \phi_1(r_i; \mu_{r_i|d}, \sigma^2_{r_i|d}),\end{aligned}$$

with

$$\begin{bmatrix} \mu_{r_i|d} \\ \boldsymbol{\mu}_{\nu|d} \end{bmatrix} = \begin{bmatrix} \mu \\ 0\mathbf{i}_{n_r} \end{bmatrix} + \begin{bmatrix} \sigma^2 \mathbf{c}_i^T \mathbf{H}^T \\ \gamma\sigma \mathbf{C}\mathbf{H}^T \end{bmatrix} \left[\sigma^2 \mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}\right]^{-1} [\mathbf{d} - \mu \mathbf{H}\mathbf{i}_{n_r}],$$

$$\begin{bmatrix} \sigma^2_{r_i|d} & \boldsymbol{\gamma}_{r_i\nu|d} \\ \boldsymbol{\gamma}_{\nu r_i|d} & \boldsymbol{\Sigma}_{\nu|d} \end{bmatrix} = \begin{bmatrix} \sigma^2 & \gamma\sigma\mathbf{c}_i^T \\ \gamma\sigma\mathbf{c}_i & \gamma^2\mathbf{C} + [1-\gamma^2]\mathbf{I}_{n_r} \end{bmatrix} - \begin{bmatrix} \sigma^2 \mathbf{c}_i^T \mathbf{H}^T \\ \gamma\sigma \mathbf{C}\mathbf{H}^T \end{bmatrix} \left[\sigma^2 \mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}\right]^{-1} \begin{bmatrix} \sigma^2 \mathbf{H}\mathbf{c}_i & \gamma\sigma\mathbf{H}\mathbf{C} \end{bmatrix},$$

$$\begin{aligned}\boldsymbol{\mu}_{\nu|r_i d} &= \boldsymbol{\mu}_{\nu|d} + \boldsymbol{\gamma}_{\nu r_i|d} \sigma^{-1}_{r_i|d} [r_i - \mu_{r_i|d}], \\ \boldsymbol{\Sigma}_{\nu|r_i d} &= \boldsymbol{\Sigma}_{\nu|d} - \boldsymbol{\gamma}_{\nu r_i|d} \sigma^{-1}_{r_i|d} \boldsymbol{\gamma}_{r_i \nu|d},\end{aligned}$$

which normally appears close to the dominant mode of the posterior pdf. These locationwise predictors can be identified from the marginal posterior pdfs which are known to be selection Gaussian pdfs with analytically assessable parameter values. The associated locationwise prediction α-intervals will naturally be the interval between the upper/lower $\alpha/2$-quantiles of the marginal posterior pdfs, which usually must be assessed by simulation-based inference. Note also that these predictors and prediction intervals will correspond to Kriging if the prior pdf is from the pure Gaussian class since MAP and E predictors coincide for unimodal symmetrical pdfs.

The characteristics of the posterior selection Gaussian spatial model are exhibited in Figure 3, which is based on the parameter sets listed in Table 2. A more detailed discussion of the

Table 2

Model parameters for four posterior cases, with $\mu = 0$ and $\sigma^2 = 1$ for all cases.

Case	γ	d_h	A	Description	Cond. values
1	0.900	4	$(\infty, -0.4] \cup [0.4, \infty)$	sym. bimodal	$2.5, -2.5$
2	0.999	4	$[-0.65, -0.4] \cup [0.12, 0.12] \cup [0.40.65]$	sym. trimodal	$0.55, -0.55$
3	0.600	4	$(\infty, -1.5] \cup [-0.5, 0.5)$	asym. unimodal	$1.0, -3.0$
4	0.700	4	$(\infty, -1.75] \cup [-0.5, 0.5] \cup [1.75, \infty)$	sym. heavy tailed	$3.0, -3.0$

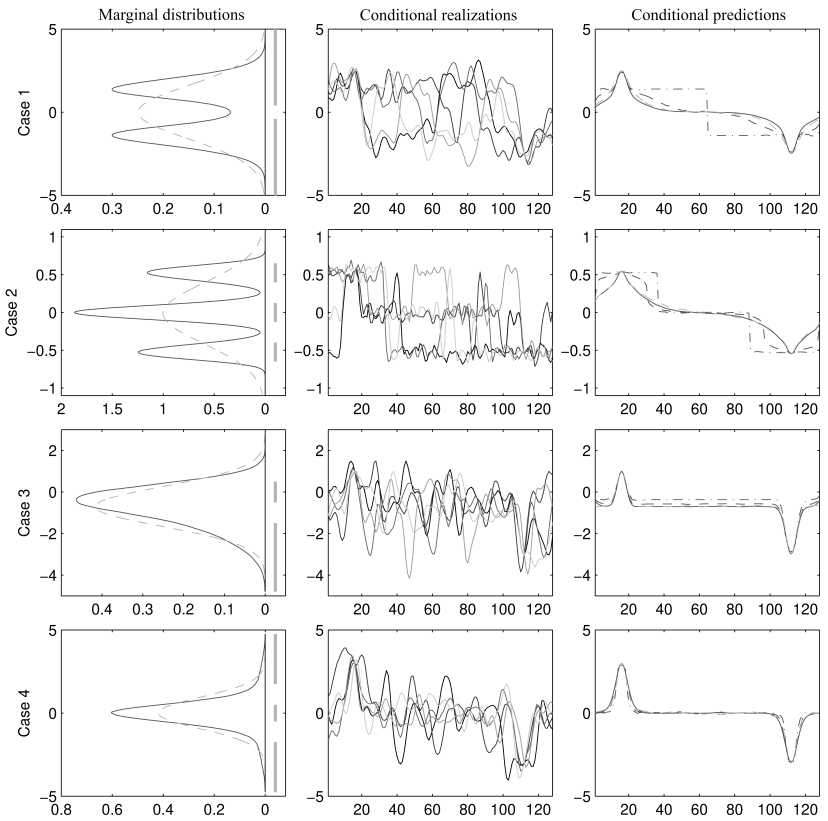

Figure 3. *First column: Marginal distribution of prior selection Gaussian model (solid black) and corresponding Gaussian model (dashed gray), and selection set on auxiliary random field on axis (solid gray). Second column: Five realizations of the posterior selection Gaussian random field. Third column: Posterior selection Gaussian model predictions, with E-prediction (solid black), MED-prediction (dashed black), and MAP-prediction (dashed-dotted black). The corresponding Gaussian model prediction (E/MED/MAP) (dashed gray).*

examples is presented in section SM4 of the supplementary material. The spatial variable is represented on a 128-grid with exact observations in grid nodes 16 and 112. The different prior models produce very different posterior realizations and predictions, all of them exactly honoring the observations of course. The MAP-predictor is particularly sensitive to multimodal marginals in the prior model. The computer demand for this simple example is very modest since the posterior model is analytically tractable.

3. Model parameter inference. One challenge with this class of selection Gaussian pdfs is the lack of clear interpretation of the model parameters, even in the reduced parametrization used as spatial stationary prior pdf in this study. The fact that the model parameter values are dependent on the actual grid design $\mathcal{L}_{\mathcal{D}}$ complicates matters even more. In this section we discuss model parameter inference in greater detail.

In order to perform Bayesian inversion, all model parameters of both the likelihood function and the prior pdf must be assigned values. The likelihood parameters $\boldsymbol{\theta}_l$ are assumed to be known through studies of the observation acquisition procedure. The model parameters of the prior pdf $\boldsymbol{\theta}_p$ are more complicated to elicit.

One may consider a hierarchical Bayesian inversion model, combining Bayesian inversion and Bayesian inference, and consider $\boldsymbol{\theta}_p$ as a random variable with a suitable prior model $f(\boldsymbol{\theta}_p)$. Then, in principle, the posterior model for $\boldsymbol{\theta}_p$ is available,

$$[\boldsymbol{\theta}_p|\mathbf{d}] \rightarrow f(\boldsymbol{\theta}_p|\mathbf{d};\boldsymbol{\theta}_l) = \text{const} \times \int f(\mathbf{d}|\mathbf{r}_A;\boldsymbol{\theta}_l) f(\mathbf{r}_A|\boldsymbol{\theta}_p) f(\boldsymbol{\theta}_p) d\mathbf{r}_A.$$

Remember that $\boldsymbol{\theta}_p = (\mu, \sigma^2, \gamma, \rho(\boldsymbol{\tau}), \mathcal{A})$, where $\mu \in \mathcal{R}$, $\sigma^2 \in \mathcal{R}_+$, and $\gamma \in [-1, 1] \subset \mathcal{R}$, while $\rho(0) = 1$, and $\rho(\boldsymbol{\tau}); \boldsymbol{\tau} \in \mathcal{R}^m_{\oplus}$ is a positive definite function, and $\mathcal{A} \subset \mathcal{R}$. Hence, both assigning suitable prior models to $\boldsymbol{\theta}_p$ and calculating the normalizing constant appear to be very complicated.

We recommend using a maximum likelihood criterion when estimating the prior model parameters $\boldsymbol{\theta}_p$ given the likelihood model parameters $\boldsymbol{\theta}_l$,

$$\hat{\boldsymbol{\theta}}_p = \underset{\boldsymbol{\theta}_p}{\text{argmax}}\{\log f(\mathbf{d};\boldsymbol{\theta}_p)\},$$

with

$$f(\mathbf{d};\boldsymbol{\theta}_p) = \int f(\mathbf{d}|\mathbf{r}_A;\boldsymbol{\theta}_l) f(\mathbf{r}_A;\boldsymbol{\theta}_p) d\mathbf{r}_A.$$

This estimate, $\hat{\boldsymbol{\theta}}_p$, will depend on the actual grid design $\mathcal{L}_{\mathcal{D}}$ of $\mathbf{r}_A$. Note that this expression is identical to the inverse of the normalizing constant of the posterior pdf in Bayesian inversion; see (1.1). In the general case this constant is unfeasible to compute even for given model parameters $(\boldsymbol{\theta}_l, \boldsymbol{\theta}_p)$ and hence even more challenging to optimize with respect to $\boldsymbol{\theta}_p$ given $\boldsymbol{\theta}_l$.

For Bayesian inversion in a conjugate setting, however, the class of posterior pdfs is known, and therefore analytical expressions for the normalizing constant are also known. This advantage of using conjugate models in Bayesian inversion is seldom recognized. For the selection Gaussian class of conjugate models with respect to Gauss-linear likelihood models, the parameter likelihood function can be demonstrated to have the form of a selection Gaussian pdf (see Appendix A),

$$f(\mathbf{d};\boldsymbol{\theta}_p) = [\Phi_{n_r}(\mathcal{A}^{n_r};\boldsymbol{\mu}_\nu,\boldsymbol{\Sigma}_\nu)]^{-1} \times \Phi_{n_r}(\mathcal{A}^{n_r};\boldsymbol{\mu}_{\nu|d},\boldsymbol{\Sigma}_{\nu|d})\ \phi_{n_d}(\mathbf{d};\boldsymbol{\mu}_d,\boldsymbol{\Sigma}_d),$$

with parameters $(\boldsymbol{\mu}_\nu, \boldsymbol{\Sigma}_\nu)$ and $(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)$ defined in (2.10), and

$$\begin{aligned}
\boldsymbol{\mu}_{\nu|d} &= \gamma\sigma\mathbf{C}\mathbf{H}^T[\sigma^2\mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]^{-1}(\mathbf{d} - \mu\mathbf{H}\mathbf{i}_{n_r}),\\
\boldsymbol{\Sigma}_{\nu|d} &= [\gamma^2\mathbf{C} + [1-\gamma^2]\mathbf{I}_{n_r}] - \gamma\sigma\mathbf{C}\mathbf{H}^T[\sigma^2\mathbf{H}\mathbf{C}\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r}]^{-1}\gamma\sigma\mathbf{H}\mathbf{C}.
\end{aligned}$$

In theory we may then define the maximum likelihood estimator for $\boldsymbol{\theta}_p$ by

$$\hat{\boldsymbol{\theta}}_p = \underset{\boldsymbol{\theta}_p}{\text{argmax}}\{\log f(\mathbf{d};\boldsymbol{\theta}_p)\}, \tag{3.1}$$

but in practice we need to parametrize also $\rho(\boldsymbol{\tau})$ and A in order to perform the optimization, and this parametrization will be problem-specific. One major challenge in the optimization is that the probability $\Phi_{n_r}(\cdot)$ needs to be recalculated for varying $\boldsymbol{\theta}_p$, which may be extremely computer-demanding. We calculate this probability by an importance blocking rejection algorithm; see section SM2 of the supplementary material. Lastly, there is no guarantee that the object function $\log f(\mathbf{d};\boldsymbol{\theta}_p)$ may not be multimodal, which makes optimization notoriously complicated.

Alternatively one may use training images of the spatial variable of interest and discretize them to a grid design corresponding to $\mathcal{L}_{\mathcal{D}}$, i.e., the same grid spacing along all dimensions. Denote one such discretized training image by n_r^o-vector $\mathbf{r}_A^o$. The corresponding Bayesian inference expression is

$$[\boldsymbol{\theta}_p|\mathbf{r}_A^o] \to f(\boldsymbol{\theta}_p|\mathbf{r}_A^o) = \text{const} \times f(\mathbf{r}_A^o|\boldsymbol{\theta}_p) f(\boldsymbol{\theta}_p),$$

which also will be very complicated to assess for the full model parameter vector $\boldsymbol{\theta}_p$. In [3] and [11] this posterior model for the model parameter γ given the other parameters in $\boldsymbol{\theta}_p$ and with $\mathcal{A} = \mathcal{R}_{\oplus}$ is discussed. The authors of the former reference also provide guidelines for obtaining conjugate prior models for γ. The generalization of these results to cover the full prior model parameter vector $\boldsymbol{\theta}_p$ appears to be very complicated.

If we use the training image $\mathbf{r}_A^o$ in a likelihood setting, it corresponds to using the parameters $\boldsymbol{\theta}_l$ as $n_r = n_d = n_r^o$, $\mathbf{H} = \mathbf{I}_{n_d}$, and $\boldsymbol{\Sigma}_{d|r} = 0 \times \mathbf{I}_{n_d}$ in the likelihood expression above,

$$f(\mathbf{r}_A^o;\boldsymbol{\theta}_p) = [\Phi_{n_r^o}(\mathcal{A}^{n_r};\boldsymbol{\mu}_\nu,\boldsymbol{\Sigma}_\nu]^{-1} \times \prod_{i=1}^{n_r^o} \Phi_1(\mathcal{A};\mu_{\nu_i|r_{Ai}^o},\sigma^2_{\nu_i|r_{Ai}^o})\ \phi_{n_r^o}(\mathbf{r}_A^o;\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r),$$

and estimating $\boldsymbol{\theta}_p$ by

$$\hat{\boldsymbol{\theta}}_p = \underset{\boldsymbol{\theta}_p}{\text{argmax}}\{\log f(\mathbf{r}_A^o;\boldsymbol{\theta}_p)\}.$$

In practice one encounters the same challenges in the optimization as those discussed above. It will be unfair to say that not many unresolved issues remain, but we present one encouraging example of prior model parameter elicitation based on training images below.

We evaluate the characteristics of the maximum likelihood estimator for $\boldsymbol{\theta}_p = (\mu, \sigma^2, d, \gamma, a)$, where $d = d_h = d_v$ and $\mathcal{A} : (-\infty, -a] \cup [a, \infty)$ for case 1 in Figure 2 and Table 1. The results are exhibited in Figure 4, and a more detailed discussion is presented in section SM5 of the supplementary material. The training images are subsets of the realization in Figure 2 of sizes $[8 \times 8], [16 \times 16], [24 \times 24]$, and $[32 \times 32]$. By repeating this inference on 1000 realizations from the prior model we can assess the accuracy and precision of the estimator. We observe from Figure 4 that the estimator appears to be biased, but consistent as the training image increases, which is as expected for maximum likelihood estimators for ergodic spatial models. Moreover, it appears that relatively reliable estimates can be obtained even for training images of size $[24 \times 24]$. The computer demand for estimating $\boldsymbol{\theta}_p$ for one training image of size $[32 \times 32]$ is typically one minute on a regular laptop computer.

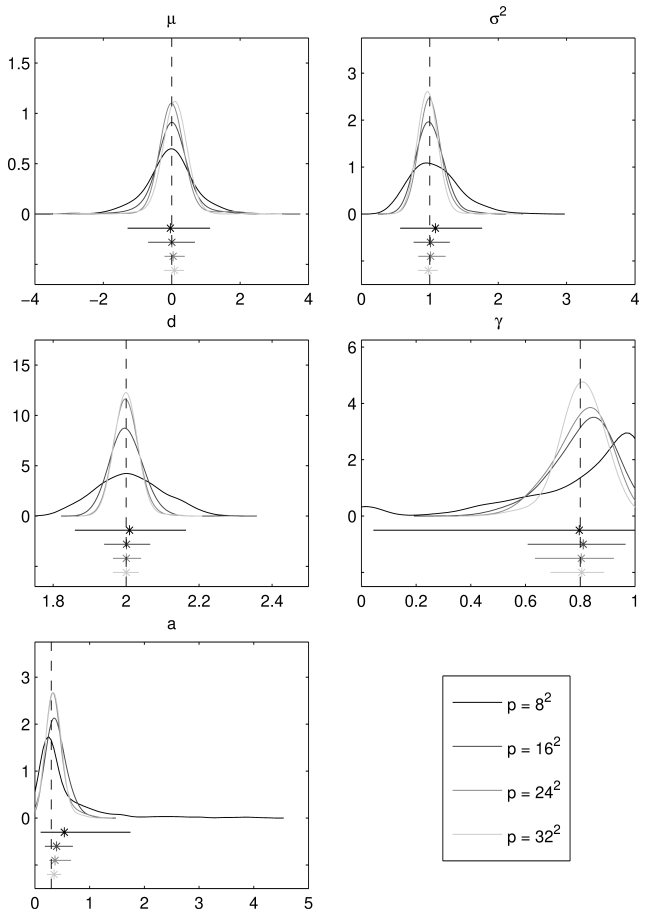

Figure 4. *Density plots of $\hat{\boldsymbol{\theta}}_p$ based on repeated estimates, evaluated for increasing size of the training image $\mathbf{r}^o$. Below the axis are means and empirical 90% confidence intervals. True values (vertical dashed lines) are marked.*

4. Case study: Seismic inversion. The objective of seismic inversion is to predict the elastic material properties—pressure-wave velocity, shear-wave velocity, and density—in the subsurface based on observed amplitude-versus-offset (AVO) seismic data collected at the surface. The data appear as time-lagged, angle-dependent reflection intensities from the subsurface created by an air pulse generated at the surface. We model the log-transformed properties in the subsurface in order to have a linear likelihood function (see [12]) $\mathbf{r} = (\log \mathbf{v}_p, \log \mathbf{v}_s, \log \boldsymbol{\rho}) \in \mathcal{R}^{3n_r}$ and denote the seismic data by $\mathbf{d} = (\mathbf{d}^1, \mathbf{d}^2, \mathbf{d}^3) \in \mathcal{R}^{3n_d}$ with upper-index representing three angles. Hence the objective is to assess $[\mathbf{r}|\mathbf{d}]$, and we phrase the inversion in a Bayesian setting. In petroleum exploration, seismic AVO data is used to map the elastic material properties in the reservoir zone of the subsurface. Based on these elastic properties, rock physics relations are then used to predict the porosity/permeability characteristics and oil/gas/water saturations in the zone in order to evaluate the petroleum recovery potential.

The case study is based on data from the Alvheim field in the North Sea; see [5] and [27].

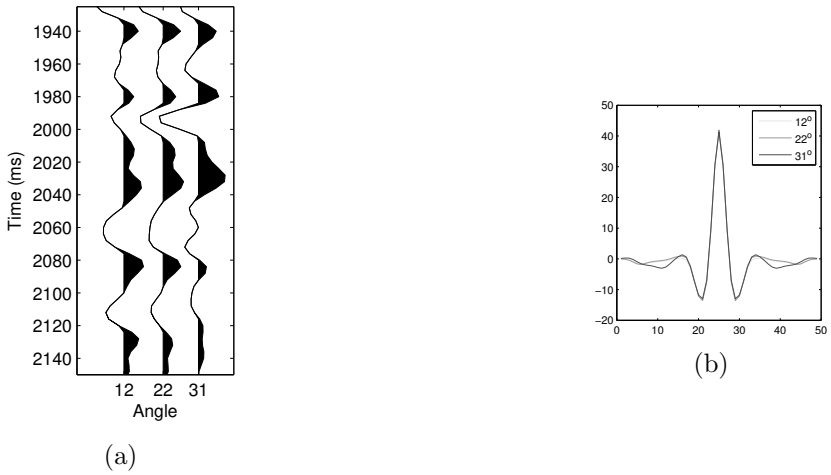

Figure 5. *Seismic AVO data in the well trace for reflection angles* 12°, 22°, *and* 31°, *with depth in seismic two-way travel time* (a) *and seismic wavelets shape for reflection angles* 12°, 22°, *and* 31° (b).

The subsurface contains a turbiditic oil and gas reservoir at about 2000 meters depth, but we use reflection time as depth reference with one meter (m) corresponding to approximately one millisecond (ms). We consider one vertical profile at the depth range $\mathcal{D} : [1935 - 2145]$ ms discretized to $\mathcal{L}_{\mathcal{D}}$ with $n = n_r = n_d = 55$ grid nodes, where both AVO seismic data $\mathbf{d}$ and exact observations of the elastic material properties $\mathbf{r}^o$ in a well are available; see Figures 5a and 6. Both $\mathbf{d}$ and $\mathbf{r}^o$ are used to infer the likelihood and prior model parameters $\boldsymbol{\theta}_l$ and $\boldsymbol{\theta}_p$, by considering $[\mathbf{d}|\mathbf{r}^o]$ and $\mathbf{r}^o$ as training images, respectively. Since both the likelihood and prior models are spatially stationary with a relative small number of parameters, overfitting does not appear to be a problem. In the Bayesian spatial inversion we consider $[\mathbf{r}|\mathbf{d}]$; hence the posterior model is only conditioned on $\mathbf{d}$. The training image $\mathbf{r}^o$ is used to validate the results. In practical use, seismic inversion of profiles in the neighborhood of the well trace, without well observations, will be made (see [18]), but then model validation is complicated. We perform Bayesian inversion based on two alternative prior models, one selection Gaussian and one traditional Gaussian prior model, and we compare the corresponding posterior models.

The likelihood model $f(\mathbf{d}|\mathbf{r})$ links the seismic data $\mathbf{d}$ and the elastic material properties of interest $\mathbf{r}$. The model is based on a linearization of the wave equation as defined in [12],

$$
\begin{aligned}
[\mathbf{d}|\mathbf{r}] &= \mathbf{WADr} + \boldsymbol{\epsilon}_{d|r} \\
&\rightarrow f(\mathbf{d}|\mathbf{r}) = \phi_{3n_r}(\mathbf{d}; \mathbf{WADr}, \sigma^2_{d|r}\boldsymbol{\Sigma}^o_{d|r}),
\end{aligned} \tag{4.1}
$$

where $\mathbf{W}$ is a convolution matrix defined by the kernels in Figure 5b; matrix $\mathbf{A}$ represents the angle-dependent linearized wave equation; $\mathbf{D}$ is a differentiation matrix; and $\boldsymbol{\epsilon}_{d|r}$ is a centered Gaussian vector with covariance matrix $\boldsymbol{\Sigma}_{d|r} = \sigma^2_{d|r}\boldsymbol{\Sigma}^o_{d|r}$. The correlation matrix $\boldsymbol{\Sigma}^o_{d|r}$ is defined by exponential correlation functions in angle and time with ranges d^a and d^t, respectively. Hence the likelihood model is Gauss-linear with parameters $\boldsymbol{\theta}_l = (\sigma^2_{d|r}, d^a, d^t)$.

The prior model $f(\mathbf{r})$ represents the general characteristics of the elastic material properties of interest. Figure 6 contains a display of the exact observations of the properties $\mathbf{r}^o$ along the

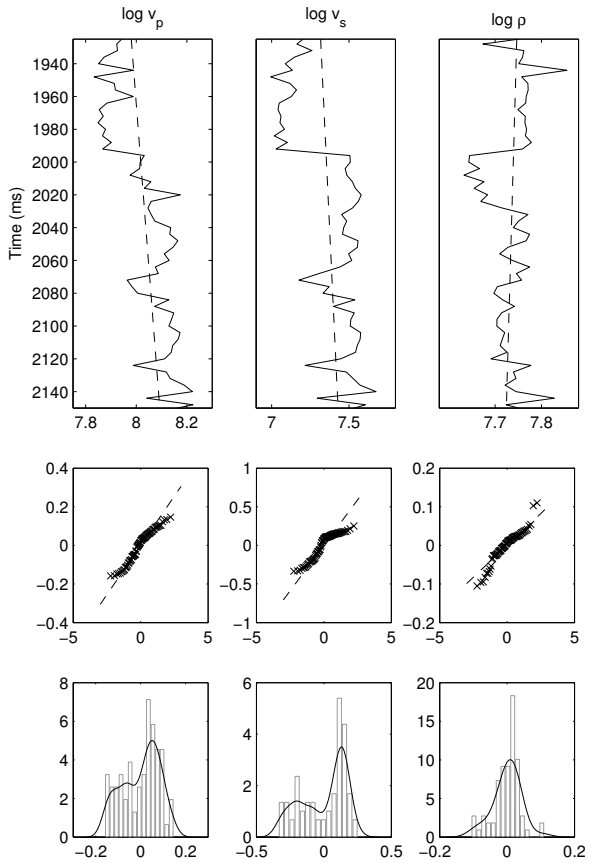

Figure 6. *Well observations of logarithm of pressure-wave velocity v_p, share-wave velocity v_s, and density ρ. Top: Elastic properties in the well with estimated linear trend (dashed black). Middle: Quantile-quantile plot of residual elastic properties. Bottom: Histograms and density estimates of residual elastic properties.*

profile and Gaussian quantile-quantile plots and spatial histograms of residuals after removing the linear vertical trend. The bimodality of the histograms of $\log v_p$ and $\log v_s$ are caused by vertically varying rock types in the subsurface. By using a selection Gaussian prior model, this bimodality in the marginal pdfs can be captured. The model as defined in section 2.3 must be extended to represent the trivariate $\mathbf{r}$, and the parametrization becomes $\boldsymbol{\theta}_p = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}, d^r, \mathbf{a})$. The spatial exponential correlation function with range d^r is common for all three variables, and the selection sets are parametrized as $\mathcal{A} : \{(-\infty, a], [a, \infty)\}$ with specific a values for each variable. The alternative traditional Gaussian prior model $f^G(\mathbf{r})$ has parametrization $\boldsymbol{\theta}_p^G = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, d^r)$ and will not capture the bimodality in the variables.

We infer the likelihood parameters $\boldsymbol{\theta}_l$ from the available seismic data $\mathbf{d}$ and the exact observations of the elastic material properties $\mathbf{r}^o$ by using a maximum likelihood criterion,

$$\hat{\boldsymbol{\theta}}_l = \underset{\boldsymbol{\theta}_l}{\operatorname{argmax}} \{ f(\mathbf{d}|\mathbf{r}^o; \boldsymbol{\theta}_l) \}.$$

Likewise we infer the model parameters for the two alternative prior models, $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_p^G$, from

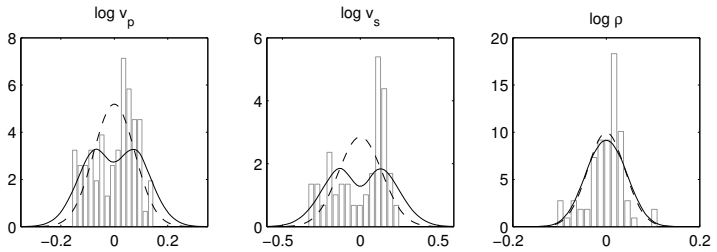

Figure 7. *Estimated prior marginal models. Marginal distributions of estimated selection Gaussian random field (solid black), marginal distributions of estimated Gaussian random field (dashed black), and histograms of well observations.*

$\mathbf{r}^o$. We set the location parameter $\boldsymbol{\mu}$, for both models, equal to the vertical linear trend for each of the three variables, and estimate the remaining parameters by a maximum likelihood criterion,

$$\hat{\boldsymbol{\theta}}_p = \underset{\boldsymbol{\theta}_p|\boldsymbol{\mu}-\text{trend}\,\mathbf{r}^o}{\text{argmax}} \{f(\mathbf{r}^o;\boldsymbol{\theta}_p)\},$$
$$\hat{\boldsymbol{\theta}}_p^G = \underset{\boldsymbol{\theta}_p^G|\boldsymbol{\mu}-\text{trend}\,\mathbf{r}^o}{\text{argmax}} \{f^G(\mathbf{r}^o;\boldsymbol{\theta}_p^G)\}.$$

The optimizations of the likelihood functions all appear to converge to unique optima with computer demands of a few minutes on a regular laptop computer. The actual estimates for the likelihood model parameters are

$$\hat{\boldsymbol{\theta}}_l : \sigma^2_{d|r} = 0.402,\ d^a = 7.3,\ d^t = 11.1,$$

while the estimates for the two alternative prior model parameters are

$$\hat{\boldsymbol{\theta}}_p : \boldsymbol{\Sigma} = \begin{bmatrix} 0.0073 & 0.0126 & -0.0013 \\ 0.0126 & 0.0250 & -0.0039 \\ -0.0013 & -0.0039 & 0.0018 \end{bmatrix},$$
$$\boldsymbol{\gamma} = \begin{bmatrix} 0.8656 \\ 0.9061 \\ 0.3331 \end{bmatrix}, d^r = 1.61, \mathbf{a} = \begin{bmatrix} 0.1110 \\ 0.2619 \\ 0.1151 \end{bmatrix},$$

$$\hat{\boldsymbol{\theta}}_p^G : \boldsymbol{\Sigma} = \begin{bmatrix} 0.0059 & 0.0093 & -0.0007 \\ 0.0093 & 0.0195 & -0.0025 \\ -0.0007 & -0.0025 & 0.0016 \end{bmatrix}, d^r = 1.53.$$

The parameter estimates for the two alternative prior models appear to be consistent with comparable range lengths and dependence structures between the three variables. In Figure 7 the marginal pdfs of the two prior models are displayed together with the histograms of

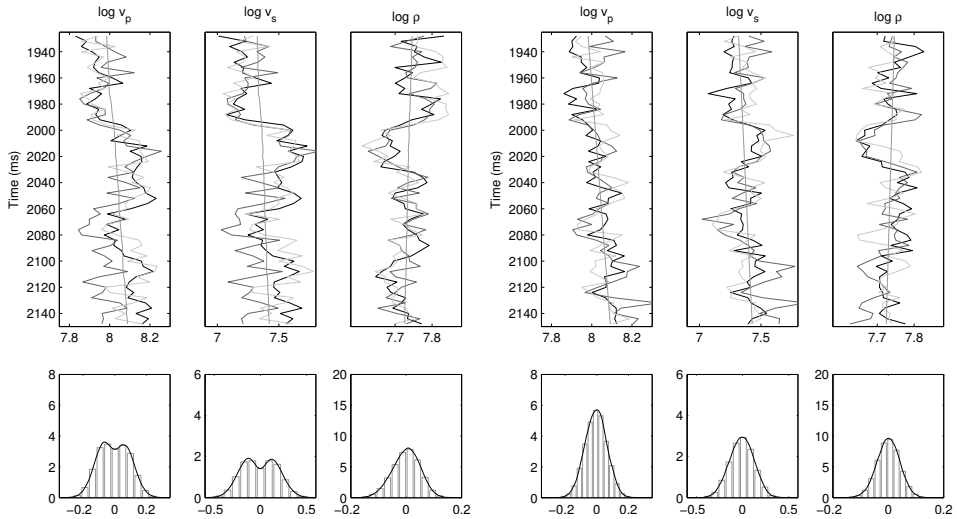

Figure 8. *Three simulated realizations from posterior random fields, and realizations integrated over time. Left: Selection Gaussian model. Right: Traditional Gaussian model.*

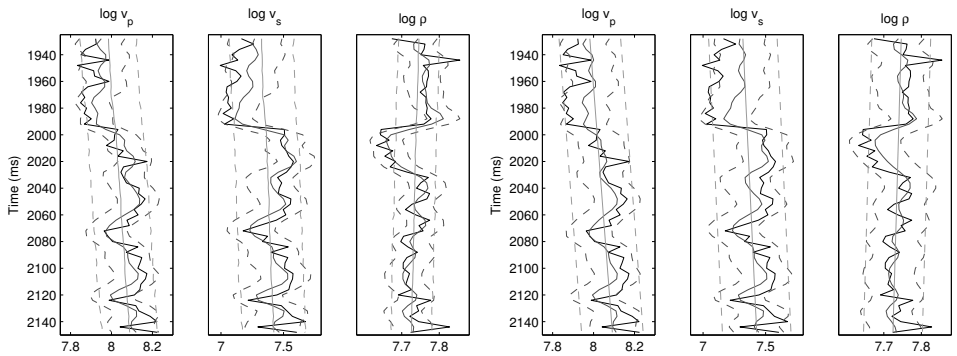

Figure 9. *Well predictions. Left: Selection Gaussian model. Right: Traditional Gaussian model. Well observations (solid black), posterior mean (solid dark gray), posterior 80% prediction interval (dashed dark gray), prior mean (solid light gray), and prior 80% prediction interval (dashed light gray).*

$\mathbf{r}^o$. The selection Gaussian prior model captures the bimodality of the histograms without overfitting to the available well observations.

Based on the Gauss-linear likelihood model $f(\mathbf{d}|\mathbf{r})$ with parameter values $\hat{\boldsymbol{\theta}}_l$ and the selection Gaussian prior model $f(\mathbf{r})$ with parameter values $\hat{\boldsymbol{\theta}}_p$, we use Bayesian spatial inversion to assess the posterior model $f(\mathbf{r}|\mathbf{d})$ which also will be selection Gaussian. By using the alternative traditional Gaussian prior model with associated parameter values we obtain a Gaussian posterior model $f^G(\mathbf{r}|\mathbf{d})$.

Realizations from the two alternative posterior models, $f(\mathbf{r}|\mathbf{d})$ and $f^G(\mathbf{r}|\mathbf{d})$, are displayed in Figure 8. The realizations from the former can be generated sequentially as outlined in section 2.4. The realizations from the selection Gaussian posterior model appear with abrupt changes between two levels defined by the two modes in the prior model; hence the

Table 3

Summary of well predictions for the selection Gaussian and Gaussian models. Mean square error (MSE) of predictions and posterior and prior 80% coverage of prediction intervals.

	MSE		Prior 80% coverage		Posterior 80% coverage	
	Selection	Gaussian	Selection	Gaussian	Selection	Gaussian
$\log v_p$	0.0034	0.0050	0.84	0.88	0.85	0.96
$\log v_s$	0.0112	0.0191	0.82	0.89	0.84	0.87
$\log \rho$	0.0009	0.0011	0.82	0.95	0.83	0.89

corresponding spatial histograms are bimodal. The realizations from the Gaussian posterior model are smoother, and the corresponding spatial histograms are unimodal.

Predictions of $[\mathbf{r}|\mathbf{d}]$ based on the two alternative posterior models with associated 0.8-prediction intervals are displayed in Figure 9. Also the correct elastic material property profiles $\mathbf{r}^o$ are presented. Moreover, predictions and 0.8-prediction intervals for the two alternative prior models are displayed. We use E-predictors for both models. In this case study, contrary to the example in section 2.4, we have densely sampled data. The data consists of convolved, gradient observations in every node of $\mathcal{L}_{\mathcal{D}}$, and the marginal pdf of the selection Gaussian posterior model will appear as almost unimodal, although flipping between modes vertically. Consequently the E, MED, and MAP predictors will be almost identical, and the former is used for computational convenience. The predictions based on both posterior models do reproduce the correct profiles relatively well with large improvements of the prior predictions. The predictions from the selection Gaussian posterior model appear with more abrupt changes than the Gaussian one whenever the correct profiles have large steps. Moreover, the 0.8-prediction intervals are narrower for the former model than for the latter.

Table 3 contains summary statistics for Figure 9. The mean-square-error (MSE) of the predictions relative to the correct profiles for each variable for both alternative models is listed. Moreover, the coverage of the 0.8-prediction intervals for the correct profiles for the prior and posterior models is specified. The predictions from the selection Gaussian model appear to be clearly superior to the predictions from the Gaussian one. The improvements in MSE are in the range of 20–40 %. The coverage values for the selection Gaussian model are close to 0.8, as they should be, while the coverages for the Gaussian model are far too large and more variable.

5. Concluding remarks. We study Bayesian spatial inversion and introduce the concept of conjugate classes of prior parametrized pdfs with respect to given classes of likelihood functions. For this class of prior pdfs the associated posterior pdfs will be in the same class. Such conjugate classes exist for continuous, event, and mosaic spatial variables for frequently used likelihood functions. The conjugate class of prior models can be selection extended without loss of the conjugate characteristic.

We demonstrate the potential of the selection extension by introducing the class of selection Gaussian prior pdfs which is conjugate with respect to Gauss-linear likelihood functions. The flexibility of this selection Gaussian class is displayed in a variety of examples which represent multimodality, skewness, and peakedness in the marginal distributions.

By using a prior model from a conjugate class for a given likelihood function, the associated posterior model can be assessed exactly based only on the model parameters of the prior and likelihood models—and the actual observations, of course. The normalizing constant,

which usually complicates Bayesian spatial inversion, will be available in parametric form. We demonstrate this favorable characteristic for the class of selection Gaussian prior pdfs and show that the posterior selection Gaussian pdf is analytically tractable, which is used to make efficient algorithms for simulation and prediction. Several examples presenting conditional simulations and predictions exposing multimodality, skewness, and peakedness are displayed.

The class of selection Gaussian prior pdfs is parametrized by a number of model parameters which are not easily interpretable. Moreover, the model parameter values will be dependent on the spatial discretization of the spatial variable under study. Based on either the available observations or one training image of the spatial variable, we define maximum likelihood estimators for the model parameters. A limited simulation study is conducted, and we conclude that the estimators appear to be consistent and that even for relatively small training images reliable estimates can be obtained. These results are encouraging.

Lastly, a case study using the selection Gaussian prior model on seismic inversion of real data is presented. We demonstrate 20–40% improvement in the MSE of predictions compared to the traditional Gaussian inversion. Also the prediction intervals of the former model appear to be more reliable than those of the latter.

The selection extension of conjugate classes of prior pdfs in Bayesian spatial inversion appears to have a large potential. We have to some extent explored this potential for continuous spatial variables and the class of selection Gaussian prior pdfs. The challenge for these models appears in sampling from and calculation of subset probabilities in high dimensional Gaussian pdfs. We have presented some relatively efficient algorithms for these purposes. Many improvements of these algorithms are definitely possible. For event and mosaic spatial variables, the class of Poisson and Markov pdfs are conjugate with respect to certain likelihood functions. Also, these classes can be selection extended and still remain conjugate. We have not yet explored these possibilities.

Appendix A. Selection Gaussian model.

The closedness properties of the selection Gaussian model is demonstrated.

Definition 3 (selection Gaussian pdf). *Consider the n-vector Gaussian basis-pdf,*

$$\mathbf{r} \to f(\mathbf{r}) = \phi_n(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r),$$

and Gauss-linear auxiliary q-vector variable,

$$[\boldsymbol{\nu}|\mathbf{r}] \to f(\boldsymbol{\nu}|\mathbf{r}) = \phi_q(\boldsymbol{\nu}; \boldsymbol{\mu}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r}),$$

with $\boldsymbol{\mu}_{\nu|r} = \boldsymbol{\mu}_\nu + \boldsymbol{\Gamma}_{\nu|r}(\mathbf{r} - \boldsymbol{\mu}_r)$, where $\boldsymbol{\Gamma}_{\nu|r}$ is denoted as the coupling $(q \times n)$-matrix.

Define a selection set $\mathcal{A}_\nu \subset \mathcal{R}^q$ and the corresponding n-vector selection Gaussian pdf,

$$\begin{aligned}\mathbf{r}_A &= [\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu] \to f(\mathbf{r}_A) = f(\mathbf{r}|\boldsymbol{\nu} \in \mathcal{A}_\nu)\\ &= [\Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_\nu, \boldsymbol{\Sigma}_\nu)]^{-1} \times \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r}) \times \phi_n(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)\\ &= const \times \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r}) \times \phi_n(\mathbf{r}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r),\end{aligned}$$

with the covariance $(q \times q)$-matrix $\boldsymbol{\Sigma}_\nu = \boldsymbol{\Gamma}_{\nu|r}\boldsymbol{\Sigma}_r\boldsymbol{\Gamma}_{\nu|r}^T + \boldsymbol{\Sigma}_{\nu|r}$.

The class of selection Gaussian pdfs is defined by all valid sets of parameters $(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r, \boldsymbol{\mu}_\nu, \boldsymbol{\Gamma}_{\nu|r}, \boldsymbol{\Sigma}_{\nu|r}, \mathcal{A}_\nu)$.

The following results are useful for later proofs.

Result 1 (conditional probabilities). *Consider the joint $(n+m)$-vectorial variable $(\mathbf{x},\mathbf{y})$ with joint pdf $f(\mathbf{x},\mathbf{y})$; then,*

$$R1.\ f(\mathbf{x}) = \int_{\Omega_y} f(\mathbf{x},\mathbf{y})d\mathbf{y} = \int_{\Omega_y} f(\mathbf{x}|\mathbf{y})f(\mathbf{y})d\mathbf{y} = \mathrm{E}_y\{f(\mathbf{x}|\mathbf{y})\},$$

and also for arbitrary subset $\mathcal{A}_x \subset \Omega_x$,

$$R2.\ F(\mathbf{x}\in\mathcal{A}_x) = \int_{\mathcal{A}_x} f(\mathbf{x})d\mathbf{x} = \int_{\mathcal{A}_x} \mathrm{E}_y\{f(\mathbf{x}|\mathbf{y})\}d\mathbf{x} = \mathrm{E}_y\{F(\mathbf{x}\in\mathcal{A}_x|\mathbf{y})\}.$$

For $f(\mathbf{x},\mathbf{y})$ being a Gaussian pdf,

$$R1G.\ \phi_n(\mathbf{x};\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x) = \mathrm{E}_y\{\phi_n(\mathbf{x};\boldsymbol{\mu}_{x|y},\boldsymbol{\Sigma}_{x|y})\},$$
$$R2G.\ \Phi_n(\mathcal{A}_x;\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x) = \mathrm{E}_y\{\Phi_n(\mathcal{A}_x;\boldsymbol{\mu}_{x|y},\boldsymbol{\Sigma}_{x|y})\}.$$

The major statements are captured in the following proposition.

Proposition 1 (selection Gaussian models). *Consider the selection Gaussian prior model,*

$$\mathbf{r}_A \to f(\mathbf{r}_A) = f(\mathbf{r}|\boldsymbol{\nu}\in\mathcal{A}_\nu) = const \times \Phi_q(\mathcal{A}_\nu;\boldsymbol{\mu}_{\nu|r},\boldsymbol{\Sigma}_{\nu|r}) \times \phi_n(\mathbf{r};\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r),$$

and Gauss-linear m-vector likelihood model,

$$[\mathbf{d}|\mathbf{r}_A] \to f(\mathbf{d}|\mathbf{r}_A) = \phi_m(\mathbf{d};\boldsymbol{\mu}_{d|r},\boldsymbol{\Sigma}_{d|r}),$$

with conditional expectation $\boldsymbol{\mu}_{d|r} = \mathbf{H}\mathbf{r}$, where $\mathbf{H}$ is an observation $(m\times n)$-matrix. Moreover, assume conditional independence of $[\boldsymbol{\nu},\mathbf{d}|\mathbf{r}]$.
Then the following hold:

A. $[\mathbf{r}_A,\mathbf{d}]$ *is selection Gaussian.*
B. $\mathbf{d}$ *is selection Gaussian.*
C. $[\mathbf{r}_A|\mathbf{d}]$ *is selection Gaussian.*

The proposition is justified by the following proof.

Proof 1. *The three proposition items are demonstrated sequentially.*

The joint pdf in **A** *is*

$$\begin{aligned}
[\mathbf{r}_A,\mathbf{d}] \to f(\mathbf{r}_A,\mathbf{d}) &= f(\mathbf{d}|\mathbf{r}_A)f(\mathbf{r}_A)\\
&= \phi_m(\mathbf{d};\boldsymbol{\mu}_{d|r},\boldsymbol{\Sigma}_{d|r}) \times const \times \Phi_q(\mathcal{A}_\nu;\boldsymbol{\mu}_{\nu|r},\boldsymbol{\Sigma}_{\nu|r}) \times \phi_n(\mathbf{r};\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r)\\
&= const \times \Phi_q(\mathcal{A}_\nu;\boldsymbol{\mu}_{\nu|r},\boldsymbol{\Sigma}_{\nu|r}) \times \phi_{n+m}\left(\begin{bmatrix}\mathbf{r}\\ \mathbf{d}\end{bmatrix};\begin{bmatrix}\boldsymbol{\mu}_r\\ \mathbf{H}\boldsymbol{\mu}_r\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{\Sigma}_r\mathbf{H}^T\\ \mathbf{H}\boldsymbol{\Sigma}_r & \mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^T+\boldsymbol{\Sigma}_{d|r}\end{bmatrix}\right)\\
&= const \times \Phi_q(\mathcal{A}_\nu;\boldsymbol{\mu}_{\nu|rd},\boldsymbol{\Sigma}_{\nu|rd}) \times \phi_{n+m}\left(\begin{bmatrix}\mathbf{r}\\ \mathbf{d}\end{bmatrix};\begin{bmatrix}\boldsymbol{\mu}_r\\ \mathbf{H}\boldsymbol{\mu}_r\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{\Sigma}_r\mathbf{H}^T\\ \mathbf{H}\boldsymbol{\Sigma}_r & \mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^T+\boldsymbol{\Sigma}_{d|r}\end{bmatrix}\right)
\end{aligned}$$

with the last identity from conditional independence of $[\boldsymbol{\nu},\mathbf{d}|\mathbf{r}]$.

Hence from Definition 3, the joint $(n+m)$-vector $[\mathbf{r}, \mathbf{d}]$ is selection Gaussian.
The marginal pdf in $\mathbf{B}$ is

$$\begin{aligned}
\mathbf{d} \rightarrow f(\mathbf{d}) &= \int f(\mathbf{r}_A, \mathbf{d}) d\mathbf{r} \\
&= const \int \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|rd}, \boldsymbol{\Sigma}_{\nu|rd}) \times \phi_{n+m}\left(\begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_r \\ \mathbf{H}\boldsymbol{\mu}_r \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{\Sigma}_r\mathbf{H}^T \\ \mathbf{H}\boldsymbol{\Sigma}_r & \mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r} \end{bmatrix} \right) d\mathbf{r} \\
&= const \times \int \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|rd}, \boldsymbol{\Sigma}_{\nu|rd}) \times \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d}) d\mathbf{r} \times \phi_m(\mathbf{d}; \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d) \\
&= const \times \mathrm{E}_{r|d}\{\Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|rd}, \boldsymbol{\Sigma}_{\nu|rd})\} \times \phi_m(\mathbf{d}; \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d) \\
&= const \times \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|d}, \boldsymbol{\Sigma}_{\nu|d}) \times \phi_m(\mathbf{d}; \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)
\end{aligned}$$

with the last identity from result R2G.
Hence from Definition 3, the marginal m-vector $\mathbf{d}$ is selection Gaussian.
The conditional pdf in $\mathbf{C}$ is

$$\begin{aligned}
[\mathbf{r}_A|\mathbf{d}] \rightarrow f(\mathbf{r}_A|\mathbf{d}) &= const \times \frac{f(\mathbf{r}_A, \mathbf{d})}{f(\mathbf{d})} \\
&= \frac{const \times \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|rd}, \boldsymbol{\Sigma}_{\nu|rd}) \times \phi_{n+m}\left(\begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_r \\ \mathbf{H}\boldsymbol{\mu}_r \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{\Sigma}_r\mathbf{H}^T \\ \mathbf{H}\boldsymbol{\Sigma}_r & \mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^T + \boldsymbol{\Sigma}_{d|r} \end{bmatrix} \right)}{const \times \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|d}, \boldsymbol{\Sigma}_{\nu|d}) \times \phi_m(\mathbf{d}; \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)} \\
&= const \times \Phi_q(\mathcal{A}_\nu; \boldsymbol{\mu}_{\nu|rd}, \boldsymbol{\Sigma}_{\nu|rd}) \times \phi_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d}).
\end{aligned}$$

Hence from Definition 3, the conditional n-vector $[\mathbf{r}_A|\mathbf{d}]$ is selection Gaussian.

REFERENCES

[1] D. Allard and P. Naveau, *A new spatial skew-normal random field model*, Comm. Statist. Theory Methods, 36 (2007), pp. 1821–1834.
[2] R. Arellano-Valle, M. Branco, and M. Genton, *A unified view on skewed distributions arising from selections*, Canad. J. Statist., 34 (2006), pp. 581–601.
[3] R. Arellano-Valle, M. Genton, and R. Loschi, *Shape mixtures of multivariate skew-normal distributions*, J. Multivariate Anal., 100 (2009), pp. 91–101.
[4] R. B. Arellano-Valle and G. E. del Pino, *From symmetric to asymmetric distributions: A unified approach*, in Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, M. G. Genton, ed., Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 113–130.
[5] P. Avseth, A. Dræge, A.-J. van Wijngaarden, T. A. Johansen, and A. Jørstad, *Shale rock physics and implications for AVO analysis: A North Sea demonstration*, The Leading Edge, 27 (2008), pp. 788–797, https://doi.org/10.1190/1.2944166.
[6] A. Azzalini, *A class of distributions which includes the normal ones*, Scand. J. Statist., 12 (1985), pp. 171–178.
[7] A. Azzalini, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, UK, 2013.
[8] A. Azzalini and A. Dalla Valle, *The multivariate skew-normal distribution*, Biometrika, 83 (1996), pp. 715–726.
[9] J. Besag, *Spatial interaction and the statistical analysis of lattice systems*, J. Roy. Statist. Soc. Ser. B, 36 (1974), pp. 192–236.
[10] D. Bolin, J. Wallin, and F. Lindgren, *Latent Gaussian random field mixture models*, Comput. Statist. Data Anal., 130 (2019), pp. 80–93.

[11] M. Branco, M. Genton, and B. Liseo, *Objective Bayesian analysis of skew-t distributions*, Scand. J. Statist., 40 (2013), pp. 63–85.
[12] A. Buland and H. Omre, *Bayesian linearized AVO inversion*, Geophysics, 68 (2003), pp. 185–198, https://doi.org/10.1190/1.1543206.
[13] G. Casella and R. L. Berger, *Statistical Inference*, Duxbury, Pacific Grove, CA, 2002.
[14] J.-P. Chiles and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed., Wiley, Hoboken, NJ, 2012.
[15] N. Cressie, *Statistics for Spatial Data*, revised ed., Wiley Ser. Probab. Math. Statist. Appl. Probab. Statist., John Wiley & Sons, New York, 1993.
[16] M. Dunlop, M. Inglesias, and A. Stuart, *Hierarchical Bayesian level set inversion*, Statist. Comput., 27 (2016), pp. 1555–1584.
[17] T. Fjeldstad and H. Omre, *Bayesian inversion of convolved hidden Markov models with applications in reservoir prediction*, IEEE Trans. Geoscience Remote Sensing, 58 (2020), pp. 1957–1968.
[18] O. B. Forberg, D. Grana, and H. Omre, *Bayesian inversion of time-lapse seismic AVO data for multimodal reservoir properties*, IEEE Trans. Geoscience Remote Sensing, (2021), pp. 1–16, https://doi.org/10.1109/TGRS.2020.3046102.
[19] M. Genton, D. Keyes, and G. Turkiyyah, *Hierarchical decompositions for the computation of high-dimensional multivariate normal probabilities*, J. Comput. Graph. Statist., 27 (2018), pp. 268–277.
[20] M. G. Genton, ed., *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall/CRC, Boca Raton, FL, 2004.
[21] A. Genz and F. Bretz, *Computation of Multivariate Normal and t Probabilities*, Springer-Verlag, New York, 2009.
[22] M. Hurn, O. Husby, and H. Rue, *A Tutorial on Image Analysis*, in Spatial Statistics and Computational Methods, J. Moller, ed., Lect. Notes Stat. 173, Springer, New York, 2003, pp. 87–141.
[23] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & Sons, Chichester, UK, 2008, https://doi.org/10.1002/9780470725160.
[24] M. Inglesias, Y. Lu, and A. Stuart, *A Bayesian level set method for geometric inverse problems*, Interfaces Free Bound., 18 (2016), pp. 181–217.
[25] H.-M. Kim and B. K. Mallick, *A Bayesian prediction using the skew Gaussian distribution*, J. Stat. Plann. Inference, 120 (2004), pp. 85–101, https://doi.org/DOI:10.1016/S0378-3758(02)00501-3.
[26] M. Minozzo and L. Ferracuti, *On the existence of some skew-normal stationary processes*, Chil. J. Stat., 3 (2012), pp. 157–170.
[27] K. Rimstad, P. Avseth, and H. Omre, *Hierarchical Bayesian lithology/fluid prediction: A North Sea case study*, Geophysics, 77 (2012), pp. B69–B85, https://doi.org/10.1190/geo2011-0202.1.
[28] K. Rimstad and H. Omre, *Skew-Gaussian random fields*, Spat. Stat., 10 (2014), pp. 43–62, https://doi.org/10.1016/j.spasta.2014.08.001.
[29] J. Røislien and H. Omre, *T-distributed random fields: A parametric model for heavy-tailed well-log data*, Math. Geology, 38 (2006), pp. 821–849.
[30] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005, https://doi.org/10.1137/1.9780898717921.