Spatio-Temporal Inversion Using the Selection Kalman Model

Maxime Conjard *[†] and Henning Omre[†]

Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

Data assimilation in models representing spatio-temporal phenomena poses a challenge, particularly if the spatial histogram of the variable appears with multiple modes. The traditional Kalman model is based on a Gaussian initial distribution and Gauss-linear forward and observation models. This model is contained in the class of Gaussian distribution and is therefore analytically tractable. It is however unsuitable for representing multimodality. We define the selection Kalman model that is based on a selection-Gaussian initial distribution and Gauss-linear forward and observation models. The selection-Gaussian distribution can be seen as a generalization of the Gaussian distribution and may represent multimodality, skewness and peakedness. This selection Kalman model is contained in the class of selection-Gaussian distributions and therefore it is analytically tractable. An efficient recursive algorithm for assessing the selection Kalman model is specified. The synthetic case study of spatio-temporal inversion of an initial state, inspired by pollution monitoring, suggests that the use of the selection Kalman model offers significant improvements compared to the traditional Kalman model when reconstructing discontinuous initial states.

Keywords: inverse problem, spatio-temporal variables, Kalman model, multimodality, data assimilation

1 INTRODUCTION

Data assimilation in models representing spatio-temporal phenomena is challenging. Most statistical spatio-temporal models are based on assumptions of temporal stationarity, possibly with a parametric, seasonal trend model [1]. We consider spatio-temporal phenomena where the dynamic spatial variables evolve according to a set of differential equations. Such phenomena will, in statistics, normally be modeled as hidden Markov models [2]. The celebrated Kalman model [3] is one of the most frequently used hidden Markov models.

In studies of hidden Markov models, it is natural to distinguish between filtering and smoothing [2]. Filtering entails predicting the spatial variable at a given time with observations up to that point in time. Smoothing entails predicting the spatial variable given observations both at previous and later times. Filtering is naturally based on recursive temporal updating while smoothing appears as more complicated since updating must also be made backwards in time. We focus on a particular smoothing challenge, namely to assess the initial state given observations at later times and we denote the task spatio-temporal inversion.

Spatio-temporal inversion is of interest in many applications. In petroleum engineering, initial water saturation is often unknown. Ensemble smoothing techniques [4–6] are commonly used to evaluate this parameter and improve fluid flow prediction. In air pollution monitoring [7], inverse trajectory methods are used to identify potential source contribution. Source mapping of wildfire

origin from airborne smoke observations is a spectacular example [8]. Evaluation of groundwater pollution mostly focuses on the future pollution of the pollutant, but as emphasized in [9], the identification of the heterogeneous source may be complicated.

We study a continuous spatial variable, a random field (RF), with temporal behavior governed by a set of differential equations. The spatio-temporal variable is discretized in space and time, and the hidden Markov model is cast in a Bayesian framework. The prior model consists of an initial spatial model and a forward spatio-temporal model, representing the evolution of the spatio-temporal phenomenon. The likelihood model represents the observation acquisition procedure. The corresponding posterior model, fully defined by the prior and likelihood models, represents the spatio-temporal phenomenon given the available observations. The traditional Kalman model constitutes a very particular hidden Markov model [3] with a Gaussian initial model and a linear forward function with Gaussian error term (Gauss-linear) forward model, and a Gauss-linear likelihood model. Since the class of Gaussian models is closed under linear operations, the posterior distribution is also Gaussian in the Kalman model, and the posterior model parameters are analytically tractable. Based on this posterior Gaussian model, both filtering and smoothing can easily be performed. In particular, the spatio-temporal inversion can be obtained by integrating out the spatial variables at all time points except the initial one, which is a simple task in Gaussian models. Most spatio-temporal models used in statistical studies are defined in the traditional Kalman model framework [10, 11]. Moreover, most of these models are based on spatial stationarity and consider filtering. Their focus is primarily on computational efficiency, not on model flexibility.

The fundamental Gauss-linear assumptions of the traditional Kalman model are often not suitable in real studies. The initial spatial variable may appear as non-Gaussian and/or the forward and/or the likelihood functions are non-linear. In the control theory community, linearizations such as the extended Kalman filter [12] or quantile-based representation such as the unscented Kalman filter [13] are recommended in these cases. These approaches are suitable for models with mild deviations from Gauss-linearity. Statisticians will more naturally use various Monte-Carlo based approaches such as the particle filter [2] or the ensemble Kalman Filter (EnKF) [14]. The particle filter is a sequential Monte Carlo algorithm with data assimilation made by reweighting the particles. To avoid singular solutions, resampling is usually required during the assimilation. The need for resampling makes the definition of an efficient corresponding particle smoother difficult. The EnKF is also a sequential Monte Carlo algorithm with data assimilation based on linear updates of each ensemble member. The sequential linear updates cause the ensemble distribution to drift toward Gaussianity. A corresponding ensemble smoother [15] is available but the ensemble drift toward Gaussianity makes it difficult to preserve non-Gaussianity in the posterior distribution. The discrete representation of the spatial variable makes the spatio-temporal model high-dimensional, and according to [16, 17], the EnKF is preferable to the particle filter in high dimensional models. Lastly, brute force Markov chain Monte Carlo

(McMC) [18] algorithms may be used for spatio-temporal inversion, but the increasing coupling of the temporal observations makes these algorithms inefficient. Focus in our study is on the spatial initial state of the spatio-temporal phenomenon, and we aim at reproducing clearly non-Gaussian marginal features, such as multi-modality, skewness and peakedness. Several models with such features are presented in the literature.

In [19], a hidden Markov model with a skew-Gaussian initial model is defined, and for Gauss-linear forward and likelihood models, it is demonstrated that the filtering is analytically tractable. The skew-Gaussian model is based on a selection concept, and the current spatio-temporal model will later be defined along these lines.

In [16, 20–24], the initial model in the hidden Markov model is defined to be a Gaussian mixture model representing multimodality. These studies all consider filtering problems and the filter algorithms are based on a combination of clustering/particle filter and Kalman filter/EnKF. The Gaussian mixture model contains a latent categorical mode indicator, which in a spatial setting must have spatial coupling, for example in the form of a Markov RF [25]. Data assimilation in such categorical Markov RFs, either by particle filter or EnKF, appears as very complicated [23, 24], particularly in a smoothing setting.

The spatio-temporal case we consider in the current study has a non-Gaussian spatial initial model, while both the forward and likelihood models are Gauss-linear. We study this special case since it has a particularly elegant analytical solution. Further, we study spatio-temporal inversion, which entails smoothing to assess the initial spatial variable given observations up to current time, as it constitutes a particularly challenging problem. To our knowledge, no reliable methodology exists for solving such a spatio-temporal inverse problem.

In the hidden Markov model considered in this study, the initial spatial model is assigned a selection-Gaussian RF [26], which may capture multi-modality, skewness and/or peakedness in the spatial histogram of the initial spatial variable. Recall that the forward and likelihood models are assumed to be Gauss-linear. Since the class of selection-Gaussian models is closed under linear operations [26] the posterior model will also be selection-Gaussian. The posterior model parameters are then analytically tractable. A general algorithm for assessing this posterior selection-Gaussian model is defined. Based on this posterior model, both smoothing and filtering can be performed. We denote this special hidden Markov model the selection Kalman model. The class of Gaussian models is a central member in the class of selection-Gaussian models [26], hence one may consider the selection Kalman model to be a generalization of the Kalman model. We develop the results presented above and demonstrate the use of the selection Kalman model on a synthetic case study of spatio-temporal inversion. This entails assessing the initial spatial variable in a dynamic model based on a limited set of observations.

The characteristics of the class of selection-Gaussian models are central in the development of the selection Kalman model properties. These characteristics are thoroughly discussed in [26],

FIGURE 1 | Initial state with observation locations (·) and monitoring locations (×).
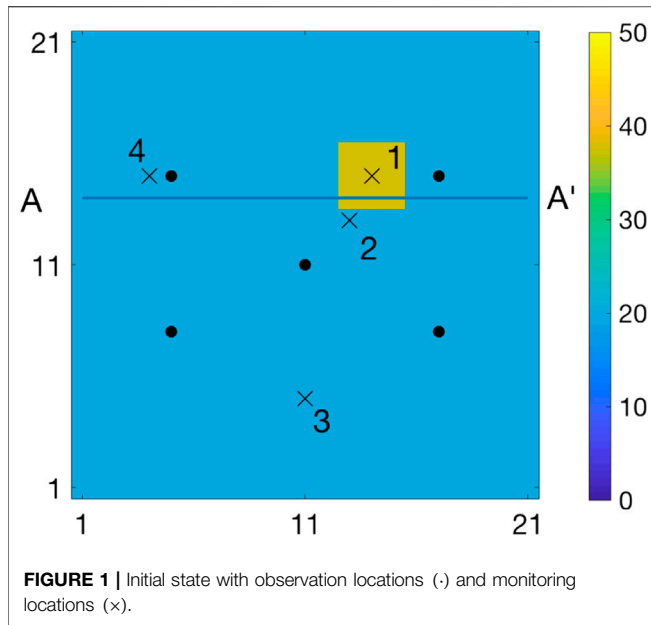
which is inspired by the results presented in [27, 28]. In these papers, the general concept of the selection-Gaussian pdf is defined in a probabilistic setting. In [19], the one-sided selection concept is used to define a skew Kalman filter in a non-spatial setting.

In this paper $y \sim f(y)$ denotes a random variable y distributed according to the probability density function (pdf) $f(y)$, or alternatively according to the corresponding cumulative distribution function (cdf) $F(y)$. Moreover, $\varphi_n(y; \mu, \Sigma)$ denotes the pdf of the Gaussian n-vector y with expectation n-vector μ and covariance $(n \times n)$-matrix Σ. Further $\Phi_n(A; \mu, \Sigma)$ denotes the probability of the aforementioned Gaussian n-vector y to be in $A \subset \mathbb{R}^n$. We also use i_n to denote the all-ones n-vector and I_n to denote the identity $(n \times n)$-matrix.

In **Section 2**, the problem is set. In **Section 3**, the traditional Kalman model is cast in a Bayesian hidden Markov model framework. The generalization to the selection Kalman model is then defined, and the analytical tractability is investigated. Further a general recursive algorithm for assessing the posterior distribution is specified. In **Section 3**, a synthetic case study of the advection-diffusion equation is discussed to showcase the ability of the selection Kalman model to solve the spatio-temporal inversion problem. The goal is to reconstruct the initial state. Results from the selection Kalman model and the traditional Kalman model are compared. In **section 4**, conclusions are presented.

2 PROBLEM SETTING

The case is defined in a spatio-temporal setting. Consider the variable $\{r_t(x); x \in \mathcal{L}_r, t \in \mathcal{T}\}; r.(\cdot) \in \mathbb{R}$, with $\mathcal{L}_r$ a grid of size n over a two-dimensional spatial area of interest while $\mathcal{T}$: $\{0, 1, \ldots, T\}$ is a regular discretization in time. Let $t = T$ represent current time while $t = 0$ represents the initial time. The spatial variable $\{r_0(x); x \in \mathcal{L}_r\}$ is a discretized representation of the initial state which later will be assumed to be unknown. **Figure 1**

displays the initial state that we evaluate in the case study. It is a spatial field with two areas: the blue area is at low value and the yellow area is at much higher value. One may consider the yellow area as the release of a pollutant at time $t = 0$.

The spatio-temporal variable evolves in time, $\{r_{t+1}(x); x \in \mathcal{L}_r\} = \omega_t[\{r_t(x); x \in \mathcal{L}_r\}]$ where $\omega_t(\cdot)$ is a dynamic function usually represented by a set of discretized differential equations. **Figure 2** shows the temporal evolution of the field presented in **Figure 1** according to a set of differential equations. The spatio-temporal variable is not fully observable, it can only be measured at a number of observation sites. The observations at the observation sites are collected with some measurement error, they appear as time series denoted $\{d_t = (d_t^1, \ldots, d_t^m), t \in \mathcal{T}\}$ where m is the number of observation sites. The five observation locations in the case study are represented by dots in **Figure 1**. **Figure 3** displays the observations collected at these observation locations. The typical challenge is to infer the spatio-temporal variable $\{r_t(x); x \in \mathcal{L}_r, t \in \mathcal{T}\}$ based on the observed time series $\{d_t; t \in \mathcal{T}\}$. It constitutes a complex spatio-temporal inverse problem. In the current study we focus on assessing the initial spatial variable $\{r_0(x); x \in \mathcal{L}_r\}$ from the observed time series $\{d_t; t \in \mathcal{T}\}$.

3 MODEL DEFINITION

Consider the unknown temporal n-vector r_t, representing the discretized spatial variable $\{r_t(x); x \in \mathcal{L}_r\}$, for $t \in \mathcal{T}_r$: $\{0, 1, \ldots, T, T+1\}$. Define the variable $r = \{r_0, r_1, \ldots, r_T, r_{T+1}\}$ and let $r_{i:j}$ denote $\{r_i, r_{i+1}, \ldots, r_j\}, \forall (i, j) \in \mathcal{T}_r^2, i \le j$. Moreover assume that the temporal m-vectors of observations d_t for $t \in \mathcal{T}_d$: $\{0, 1, \ldots, T\}$ are available, and define $d = \{d_0, d_1, \ldots, d_T\}$ and $d_{i:j} = \{d_i, \ldots, d_j\}$ accordingly. The objective of this study is to assess $[r_0|d]$. To that end, we define a Kalman type model, represented as a hidden Markov model in a Bayesian inversion framework.

3.1 Bayesian Inversion
The Kalman type model, phrased as Bayesian inversion, requires the specification of a prior model for r and a likelihood model for $[d|r]$. The model specified below defines a hidden Markov model as displayed in **Figure 4**.

3.1.1 Prior Model
The prior model on r synthesizes the knowledge and experience with the spatial variable of interest, and it consists of an initial distribution and a forward model:

Initial Distribution
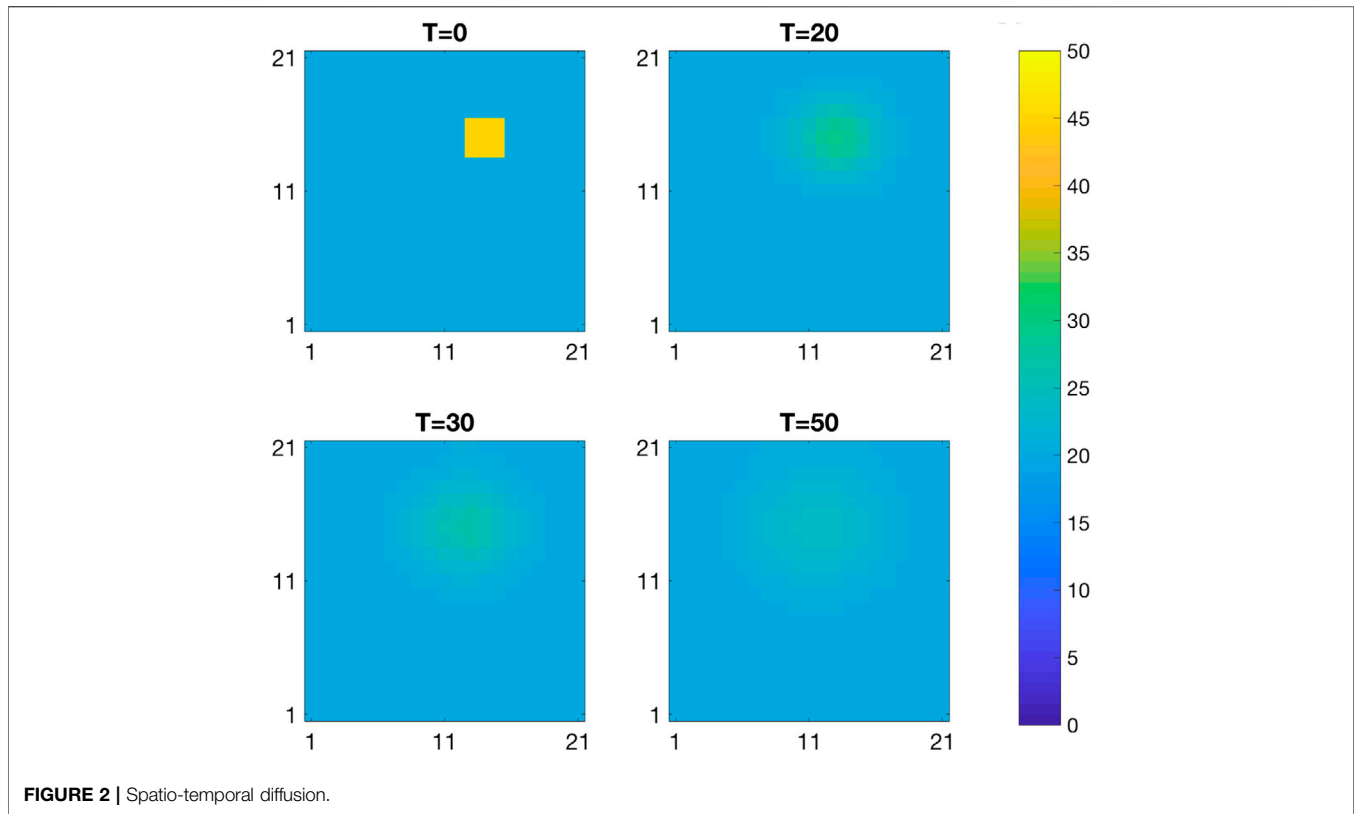   The prior distribution for the initial state r_0 is denoted $f(r_0)$.

Forward Model
    The forward model given the initial state $[r_{1:T+1}|r_0]$ is defined as,

$$f(r_{1:T+1}|r_0) = \prod_{t=0}^{T} f(r_{t+1}|r_t) \tag{1}$$

with,

$$[r_{t+1}|r_t] = \omega_t(r_t, \varepsilon_t^r) \sim f(r_{t+1}|r_t) \tag{2}$$

FIGURE 2 | Spatio-temporal diffusion.

where $\omega_t(\cdot,\cdot) \in \mathbb{R}^n$ is the function that propagates r_t forward in time, with ε_t^r a random component. Since $\omega_t(\cdot,\cdot)$ only involves the variable at the previous time step, r_t, the forward model is a Markov chain.

3.1.2 Likelihood Model

The likelihood model on $[d|r]$ provides a link between the variable of interest r and the observations d and is defined as,

$$f(d|r) = \prod_{t=0}^{T} f(d_t|r_t) \tag{3}$$

with,

$$[d_t|r_t] = \psi_t\left(r_t, \varepsilon_t^d\right) \sim f(d_t|r_t) \tag{4}$$

where $\psi_t(\cdot,\cdot) \in \mathbb{R}^m$ is the likelihood function with ε_t^d a random component. The likelihood model is defined assuming conditional independence and single state response and is thus in factored form.

3.1.3 Posterior Model

Bayesian inversion endeavors to assess the posterior distribution of $[r|d]$,

$$\begin{aligned} f(r|d) &= \left[\int f(d|r)f(r)dr\right]^{-1} \times f(d|r)f(r) \\ &= const \times f(d_0|r_0)f(r_0) \\ &\quad \times \prod_{t=1}^{T} f(d_t|r_t)f(r_t|r_{t-1})f(r_{T+1}|r_T) \\ &= f(r_0|d) \prod_{t=1}^{T} f(r_t|r_{t-1}, d_{t:T})f(r_{T+1}|r_T) \end{aligned} \tag{5}$$

which is a non-stationary Markov chain for the hidden Markov model with a likelihood model in factored form as defined above [29]. Assessing such a posterior distribution is usually difficult as both the normalizing constant and the conditional transition matrices are challenging to calculate.

3.2 Kalman Type Models

The current study is limited to Kalman type models. They comprise an initial and a process part.

Initial Distribution

The initial distribution is identical to the initial distribution of the prior model $f(r_0)$, and as such captures the characteristics of the initial state of the process. Two model classes are later discussed: the Gaussian and the selection-Gaussian classes.

Process Model

The process model includes the forward model and likelihood models defined in **Section 3.1**. It thus characterizes the process dynamics and the observation acquisition procedure. The forward model is defined by,

$$\begin{aligned} [r_{t+1}|r_t] &= A_t r_t + \varepsilon_t^r \\ f(r_{t+1}|r_t) &= \varphi_n\left(r_{t+1}; A_t r_t, \Sigma_t^{r|r}\right) \end{aligned} \tag{6}$$

with forward $(n \times n)$-matrix A_t and n-vector error term ε_t^r defined as centered Gaussian with covariance $(n \times n)$-matrix $\Sigma_t^{r|r}$. The forward model is therefore Gauss-linear. The likelihood component is defined by,
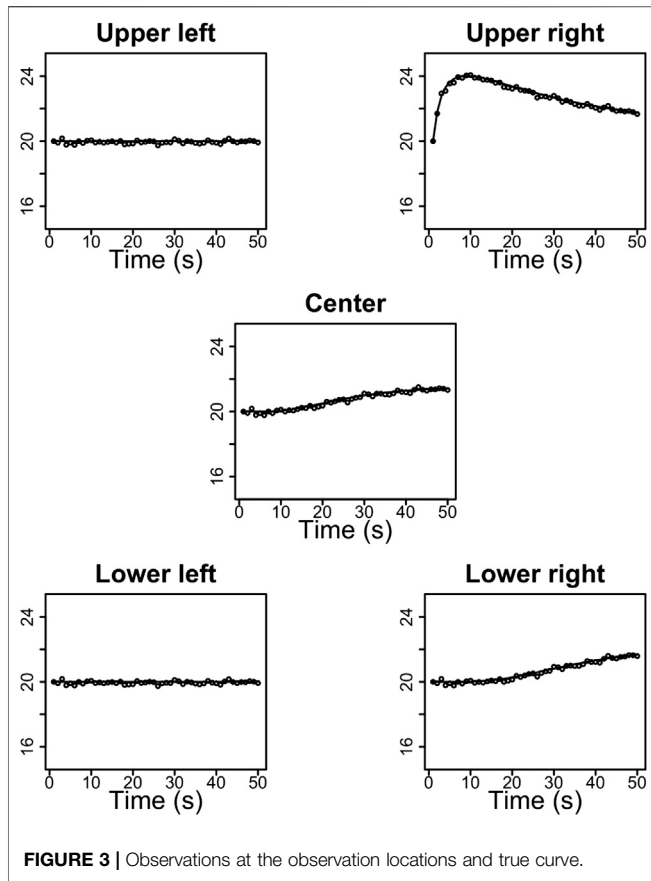
FIGURE 3 | Observations at the observation locations and true curve.

$$[d_t|r_t] = Hr_t + \varepsilon_t^d$$
$$f(d_t|r_t) = \varphi_p\left(d_t; Hr_t, \Sigma_t^{d|r}\right) \qquad (7)$$

with the observation $(m \times n)$-matrix H and the m-vector error term ε_t^d defined as centered Gaussian with covariance $(m \times m)$-matrix $\Sigma_t^{d|r}$. The likelihood model is also Gauss-linear. This process model coincides with the frequently used traditional Kalman model [3].

3.3 Traditional Kalman Model

The traditional Kalman model is defined by letting the initial distribution be in the class of Gaussian pdfs,

$$r_0 \sim f(r_0) = \varphi_n\left(r_0; \mu_0^r, \Sigma_0^r\right) \qquad (8)$$

with initial expectation n-vector μ_0^r and positive definite covariance $(n \times n)$-matrix Σ_0^r. The Gaussian initial distribution is parametrized by $\Theta^G = (\mu_0^r, \Sigma_0^r)$. In our spatial study, this initial distribution will be a discretized stationary Gaussian RF. The process model is Gauss-linear and identical to the traditional Kalman type.

This traditional Kalman model is analytically tractable. The posterior distribution $f(r|d)$ is Gaussian and the posterior distribution parameters can be calculated by algebraic operations on the parameters of the initial distribution, process model and the observed data. Therefore the assessment of the posterior distribution does not require computationally demanding integrals. The analytical tractability follows from the recursive reproduction of Gaussian pdfs:

- The initial model $f(r_0)$ is Gaussian and the likelihood model $f(d_0|r_0)$ is Gauss-linear, hence the joint model $f(r_0, d_0)$ is Gaussian. Consequently, the conditional model $f(r_0|d_0)$ is Gaussian.
- The conditional model $f(r_0|d_0)$ is Gaussian and the dynamic model $f(r_1|r_0)$ is Gauss-linear, hence the joint conditional model $f(r_1, r_0|d_0)$ is Gaussian.

By recursion, we obtain that $f(r|d) = f(r_0, \ldots, r_{T+1}|d_0, \ldots, d_T)$ is Gaussian. Note in particular that since $f(r|d)$ is Gaussian, so is $f(r_0|d)$. This pdf is obtained by marginalization of $f(r|d)$ which, for the Gaussian case, amounts to removing rows from the expectation vector and rows and columns from the covariance matrix. Additionally, the joint pdf $f(r, d)$ can be assessed using a simple recursive algorithm, see **Supplementary Appendix Algorithm A1** in **Supplementary Appendix A**.

From the joint Gaussian pdf $f(r, d)$, the posterior distribution $f(r|d)$ can be analytically assessed. In spatial models, the grid dimension n may be large while the number of data collection sites m usually is small. **Supplementary Appendix Algorithm A1** requires storing the covariance $[n(T+2) + m(T+1)] \times [n(T+2) + m(T+1)]$-matrix of the Gaussian vector $[r, d]$ which is hardly ever necessary in practice where the target distribution is clearly identified. Only the spatial variables of interest need to be stored, which entails that only the covariance $[n + m(T+1)] \times [n + m(T+1)]$-matrix of $[r_0, d]$ need to be stored in our spatio-temporal inversion study.
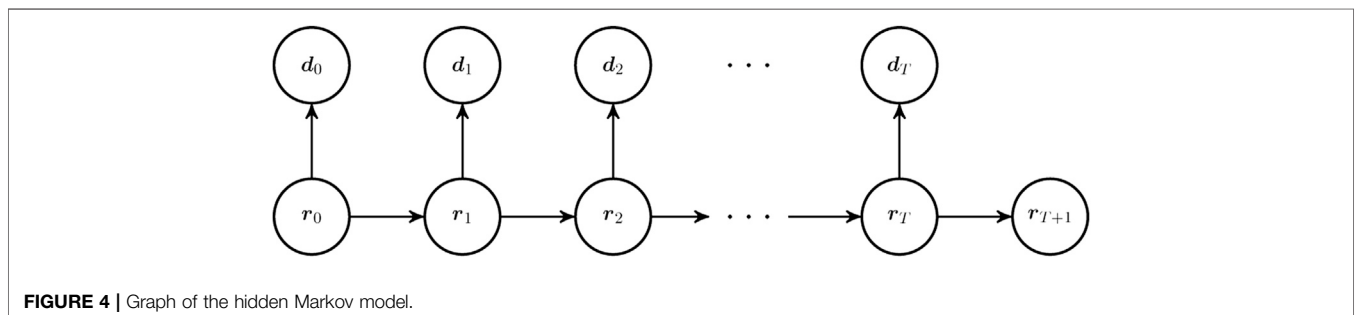

FIGURE 4 | Graph of the hidden Markov model.

3.4 Selection Kalman Model

The selection Kalman model is defined by letting the initial distribution be in the class of selection-Gaussian pdfs [27, 28]. This class is defined by considering an auxilary n-vector $\tilde{r}$ with pdf from the Gaussian class,

$$f(\tilde{r}) = \varphi_n(\tilde{r}; \mu_{\tilde{r}}, \Sigma_{\tilde{r}}) \qquad (9)$$

with expectation n-vector $\mu_{\tilde{r}}$ and covariance $(n \times n)$-matrix $\Sigma_{\tilde{r}}$. In our spatial study this pdf will represent a discretized stationary Gaussian RF. Define further an auxiliary q-vector ν by a Gauss-linear extension,

$$[\nu|\tilde{r}] = \mu_\nu + \Gamma_{\nu|\tilde{r}}(\tilde{r} - \mu_{\tilde{r}}) + \varepsilon_{\nu|\tilde{r}} \qquad (10)$$

with the expectation q-vector μ_ν, and the regression $(q \times n)$-matrix $\Gamma_{\nu|\tilde{r}}$ and the centered Gaussian q-vector $\varepsilon_{\nu|\tilde{r}}$, independent of $\tilde{r}$, with covariance $(q \times q)$-matrix $\Sigma_{\nu|\tilde{r}}$. In the current spatial study the dimension of $\tilde{r}$ and ν will be identical. Generally, we have,

$$f(\nu|\tilde{r}) = \varphi_q(\nu; \mu_{\nu|\tilde{r}}, \Sigma_{\nu|\tilde{r}}) \qquad (11)$$

with $\mu_{\nu|\tilde{r}} = \mu_\nu + \Gamma_{\nu|\tilde{r}}(\tilde{r} - \mu_{\tilde{r}})$. As a consequence, $[\tilde{r}, \nu]$ is jointly Gaussian,

$$\begin{bmatrix} \tilde{r} \\ \nu \end{bmatrix} \sim f(\tilde{r}, \nu) = \varphi_{n+q}\left(\begin{bmatrix} \tilde{r} \\ \nu \end{bmatrix}; \begin{bmatrix} \mu_{\tilde{r}} \\ \mu_\nu \end{bmatrix}, \begin{bmatrix} \Sigma_{\tilde{r}} & \Sigma_{\tilde{r}}\Gamma_{\nu|\tilde{r}}^T \\ \Gamma_{\nu|\tilde{r}}\Sigma_{\tilde{r}} & \Sigma_\nu \end{bmatrix} \right) \qquad (12)$$

with the covariance $(q \times q)$-matrix $\Sigma_\nu = \Gamma_{\nu|\tilde{r}}\Sigma_{\tilde{r}}\Gamma_{\nu|\tilde{r}}^T + \Sigma_{\nu|\tilde{r}}$. Define a selection subset $A \subset \mathbb{R}^q$, and define the class of selection-Gaussian pdfs by $r_A = [\tilde{r}|\nu \in A]$. In the current spatial study the set A will be separable in $\mathbb{R}^q$. Generally, it follows that,

$$\begin{aligned} f(r_A) &= f(\tilde{r}|\nu \in A) \\ &= \left[\Phi_q(A; \mu_\nu, \Sigma_\nu)\right]^{-1} \\ &\quad \times \Phi_q(A; \mu_{\nu|\tilde{r}}, \Sigma_{\nu|\tilde{r}}) \times \varphi_n(\tilde{r}; \mu_{\tilde{r}}, \Sigma_{\tilde{r}}) \end{aligned} \qquad (13)$$

This class of pdfs is parametrized by $\Theta^{SG} = (\mu_{\tilde{r}}, \Sigma_{\tilde{r}}, \mu_\nu, \Gamma_{\nu|\tilde{r}}, \Sigma_{\nu|\tilde{r}}, A)$ for all valid parameter sets. The class of selection-Gaussian pdfs is very flexible and may represent multi-modality, skewness and peakedness [26].

Four one-dimensional selection-Gaussian pdfs are displayed in **Figure 5** in order to demonstrate the influence of the selection set $A \subset \mathbb{R}$. The bivariate variable $[\tilde{r}, \nu]$ is bi-Gaussian and identical in all displays, while the selection sets are marked as solid gray bars along the vertical ν-axis. **Figure 5A** contains a selection set comprised of two segments symmetric about the expectation of ν, making the selection-Gaussian pdf along the horizontal axis bimodal and symmetric. **Figure 5B** contains a selection set of two asymmetric segments, making the selection-Gaussian pdf bimodal and asymmetric. **Figure 5C** contains a selection set of three segments symmetric about the expectation of ν, making the selection-Gaussian pdf trimodal and symmetric.
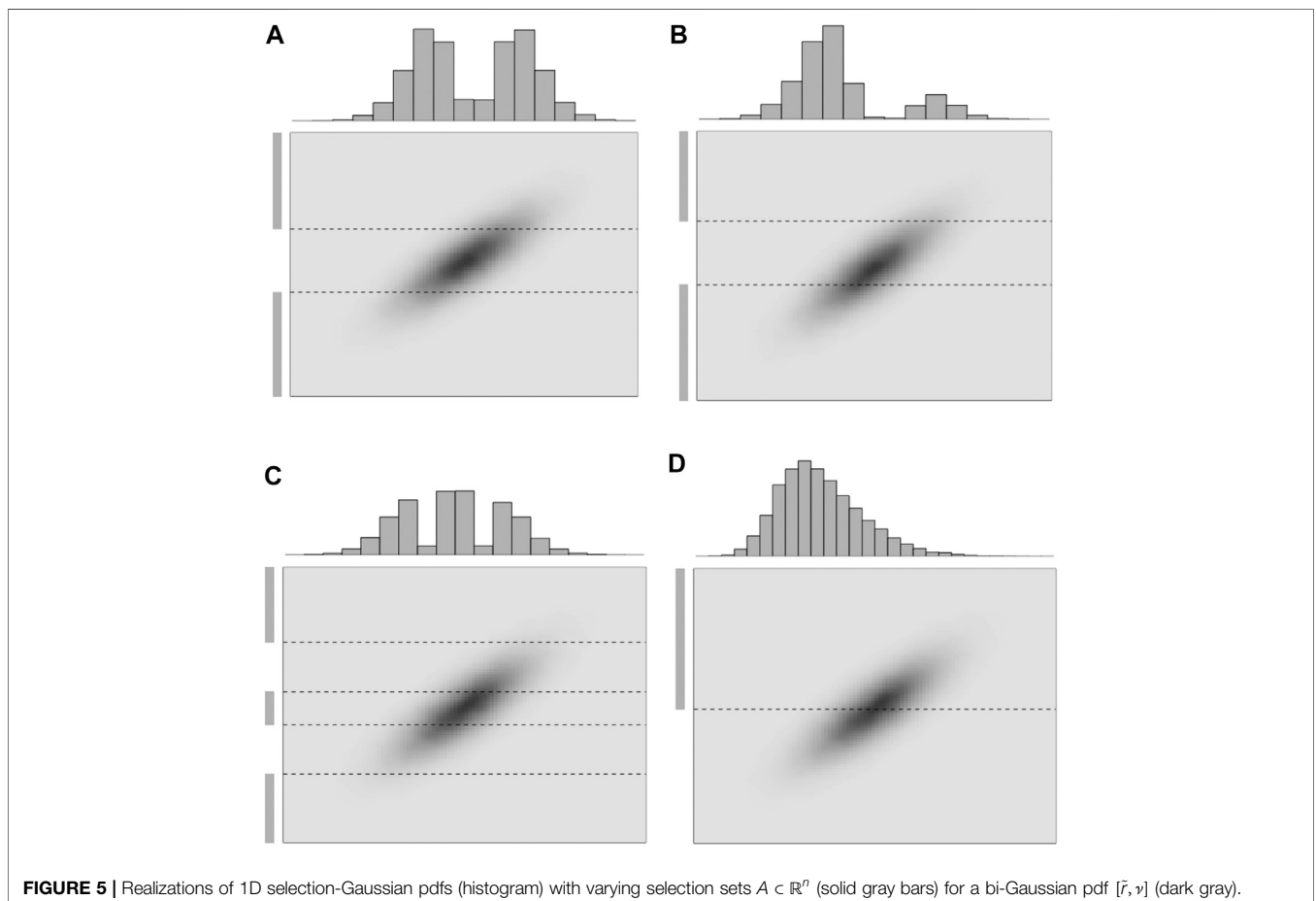
FIGURE 5 | Realizations of 1D selection-Gaussian pdfs (histogram) with varying selection sets $A \subset \mathbb{R}^n$ (solid gray bars) for a bi-Gaussian pdf $[\tilde{r}, \nu]$ (dark gray).
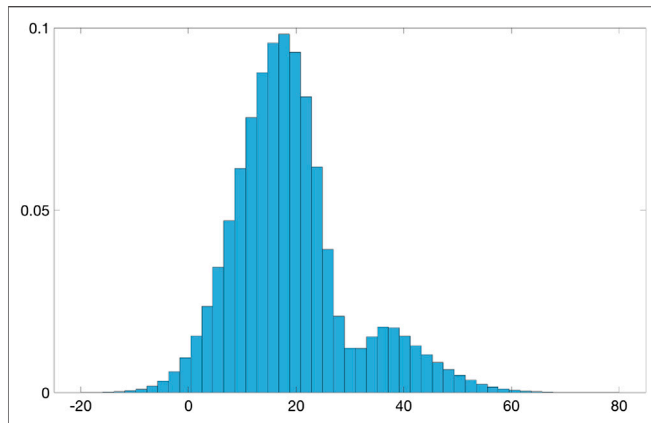
FIGURE 6 | Prior marginal distribution of the initial temperature field for the selection Kalman model.

Lastly, **Figure 5D** contains a selection set comprised of only one segment, making the selection-Gaussian pdf skewed. This selection concept can be extended to higher dimensions and even to discretized spatial models.

Note that assigning a null-matrix to $\Gamma_{\nu|\tilde{r}}$ entails that $f(\tilde{r}, \nu) = f(\tilde{r})f(\nu)$ and selection on ν does not influence $\tilde{r}$. It follows that $f(r_A) = f(\tilde{r})$ is Gaussian. The selection-Gaussian model can therefore be seen as a generalization of the Gaussian one. Assume that the conditional independence, $f(d_t, \nu|\tilde{r}_t) = f(d_t|\tilde{r}_t)f(\nu|\tilde{r}_t)$, holds for all t, it can then be demonstrated [26] that the following recursive reproduction of selection-Gaussian pdfs holds:

- The initial model $f(r_{A,0})$ is selection-Gaussian and the likelihood model $f(d_0|r_{A,0})$ is Gauss-linear, hence the joint model $f(r_{A,0}, d_0)$ is selection-Gaussian. Moreover, the conditional model $f(r_{A,0}|d_0)$ is selection-Gaussian.
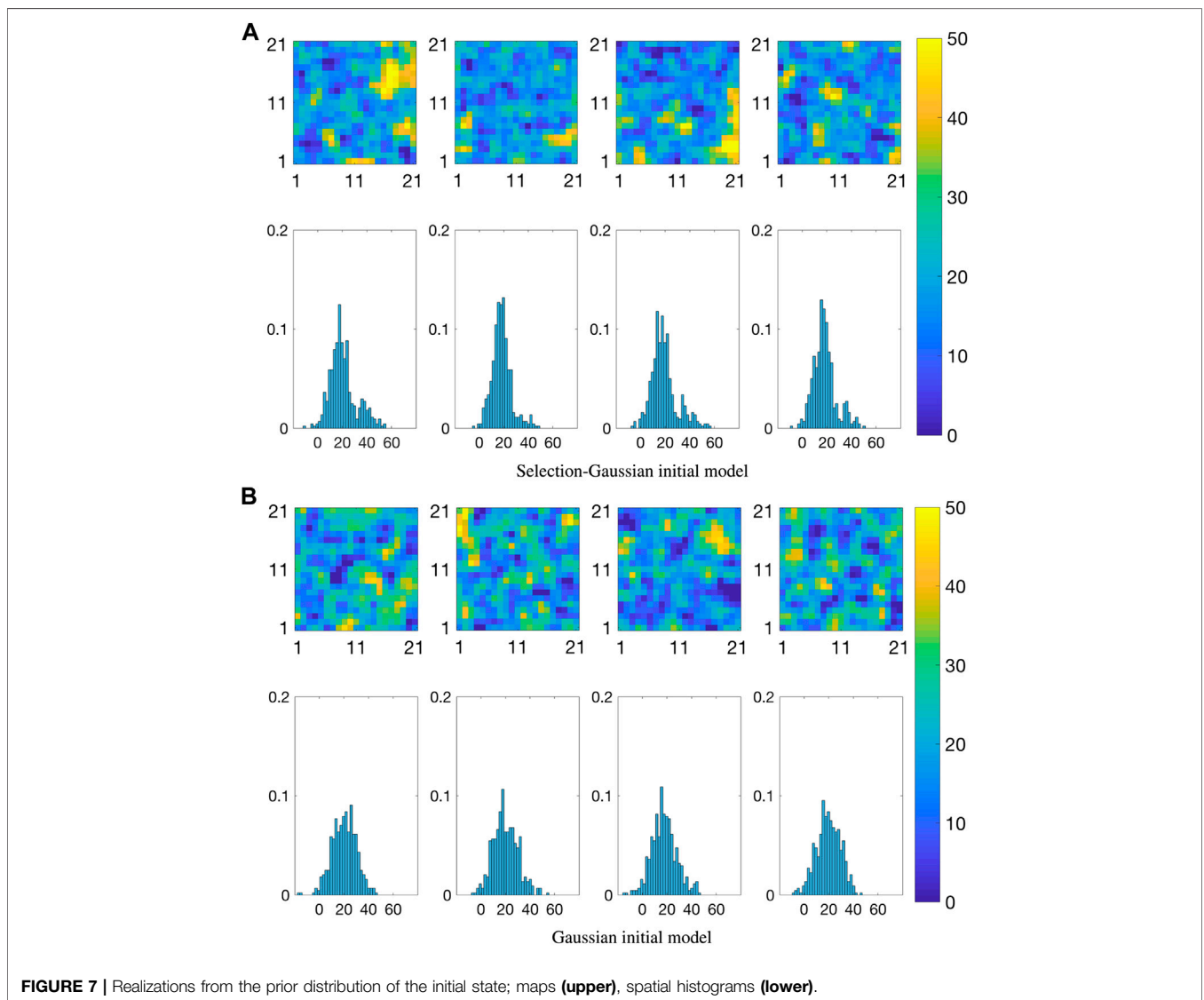
FIGURE 7 | Realizations from the prior distribution of the initial state; maps **(upper)**, spatial histograms **(lower)**.

- The conditional model $f(r_{A,0}|d_0)$ is selection-Gaussian and the dynamic model $f(r_{A,1}|r_{A,0})$ is Gauss-linear, the joint conditional model $f(r_{A,1}, r_{0,A}|d_0)$ is therefore selection-Gaussian.

By recursion, we obtain that $f(r_A|d) = f(r_{A,0}, \ldots, r_{A,T+1}|d_0, \ldots, d_T)$ is selection-Gaussian. Recall that these characteristics are similar to those of the class of Gaussian pdfs that make the traditional Kalman model analytically tractable. The selection Kalman model is defined with an initial distribution from the class of selection-Gaussian pdfs and a process model which is Gauss-linear and identical to the traditional Kalman type. From the characteristics of the class of selection-Gaussian distributions, it

follows that the posterior distribution $f(r_A|d)$ is in the class of selection-Gaussian distributions and so is $f(r_{A,0}|d)$.

Consider the augmented $(n + q)$-vector $[\tilde{r}_0, \nu]$ which together with the selection set $A \in \mathbb{R}^q$ defines the initial state $r_{A,0} = [\tilde{r}_0||\nu \in A]$. The recursive algorithm, see **Algorithm 1**, is initiated with this augmented vector which is Gaussian. The conditional independence $f(d_t, \nu|\tilde{r}_t) = f(d_t|\tilde{r}_t)f(\nu|\tilde{r}_t)$ entails that $f(d_t|r_{A,t}) = f(d_t|r_t)$, which is Gauss-linear for all t.

Algorithm 1 provides the Gaussian pdf of the $[n(T+2) + q + m(T+1)]$-vector $[\tilde{r}, \nu, d]$. From the joint Gaussian pdf $f(\tilde{r}, \nu, d)$, the pdf $f(r_{A,0}|d) = f(\tilde{r}_0|\nu \in A, d)$ can be assessed by first marginalizing $\tilde{r}$ and thereafter sequentially

ALGORITHM 1 | *Joint Selection Kalman Model*

- *Define*

$$\boldsymbol{\mu}_t^{\tilde{r}} = \mathbb{E}[\tilde{\boldsymbol{r}}_t]$$
$$\boldsymbol{\mu}_0^{\nu} = \mathbb{E}[\boldsymbol{\nu}]$$
$$\boldsymbol{\mu}_t^{d} = \mathbb{E}[\boldsymbol{d}_t]$$
$$\boldsymbol{\Sigma}_{ts}^{\tilde{r}\tilde{r}} = \text{Cov}(\tilde{\boldsymbol{r}}_t, \tilde{\boldsymbol{r}}_s) = \boldsymbol{\Sigma}_{st}^{\tilde{r}\tilde{r}T}$$
$$\boldsymbol{\Sigma}_{ts}^{d\tilde{d}} = \text{Cov}(\boldsymbol{d}_t, \boldsymbol{d}_s) = \boldsymbol{\Sigma}_{st}^{dd^T}$$
$$\boldsymbol{\Sigma}_{00}^{\nu\nu} = \text{Cov}(\boldsymbol{\nu}, \boldsymbol{\nu})$$
$$\boldsymbol{\Gamma}_{ts}^{\tilde{r}d} = \text{Cov}(\tilde{\boldsymbol{r}}_t, \boldsymbol{d}_s) = \boldsymbol{\Gamma}_{st}^{d\tilde{r}T}$$
$$\boldsymbol{\Gamma}_{t0}^{\tilde{r}\nu} = \text{Cov}(\tilde{\boldsymbol{r}}_t, \boldsymbol{\nu}) = \boldsymbol{\Gamma}_{0t}^{\nu\tilde{r}T}$$
$$\boldsymbol{\Gamma}_{t0}^{d\nu} = \text{Cov}(\boldsymbol{d}_t, \boldsymbol{\nu}) = \boldsymbol{\Gamma}_{0t}^{\nu d^T}$$

- *Initiate*

$$\begin{bmatrix} \tilde{\boldsymbol{r}}_0 \\ \boldsymbol{\nu} \end{bmatrix} \sim \varphi_{n+q} \left(\begin{bmatrix} \tilde{\boldsymbol{r}} \\ \boldsymbol{\nu} \end{bmatrix} ; \begin{bmatrix} \boldsymbol{\mu}_{\tilde{r}} \\ \boldsymbol{\mu}_{\nu} \end{bmatrix} , \begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{r}} & \boldsymbol{\Sigma}_{\tilde{r}}\boldsymbol{\Gamma}_{\nu|\tilde{r}}^T \\ \boldsymbol{\Gamma}_{\nu|\tilde{r}}\boldsymbol{\Sigma}_{\tilde{r}} & \boldsymbol{\Sigma}_{\nu} \end{bmatrix} \right)$$

$$\boldsymbol{\mu}_0^{\tilde{r}} = \boldsymbol{\mu}_{\tilde{r}}$$
$$\boldsymbol{\mu}_0^{\nu} = \boldsymbol{\mu}_{\nu}$$
$$\boldsymbol{\Sigma}_{00}^{\tilde{r}\tilde{r}} = \boldsymbol{\Sigma}_{\tilde{r}}$$
$$\boldsymbol{\Sigma}_{00}^{\tilde{r}\nu} = \boldsymbol{\Sigma}_{\tilde{r}}\boldsymbol{\Gamma}_{\nu|\tilde{r}}^T$$
$$\boldsymbol{\Sigma}_{00}^{\nu\nu} = \boldsymbol{\Sigma}_{\nu}$$

- *Iterate* $t = 0, \ldots, T$

 Likelihood model:

$$\boldsymbol{\mu}_t^{d} = \boldsymbol{H}\boldsymbol{\mu}_t^{\tilde{r}}$$
$$\boldsymbol{\Sigma}_{tt}^{dd} = \boldsymbol{H}\boldsymbol{\Sigma}_{tt}^{\tilde{r}\tilde{r}}\boldsymbol{H}^T + \boldsymbol{\Sigma}_t^{d|r}$$
$$\boldsymbol{\Gamma}_{t0}^{d\nu} = \boldsymbol{H}\boldsymbol{\Gamma}_{t0}^{\tilde{r}\nu}$$

 Iterate $s = 0, \ldots, t$

$$\boldsymbol{\Gamma}_{ts}^{\tilde{r}d} = \boldsymbol{\Sigma}_{ts}^{\tilde{r}\tilde{r}}\boldsymbol{H}^T$$

 End iterate s

 If $t > 0$: *Iterate* $s = 0, \ldots, t-1$

$$\boldsymbol{\Sigma}_{t,s}^{dd} = \boldsymbol{H}\boldsymbol{\Gamma}_{ts}^{\tilde{r}d}$$

 End iterate s

 Forwarding model:

$$\boldsymbol{\mu}_{t+1}^{\tilde{r}} = \boldsymbol{A}_t\boldsymbol{\mu}_t^{\tilde{r}}$$
$$\boldsymbol{\Sigma}_{(t+1)(t+1)}^{\tilde{r}\tilde{r}} = \boldsymbol{A}_t\boldsymbol{\Sigma}_{tt}^{\tilde{r}\tilde{r}}\boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t^{r|r}$$
$$\boldsymbol{\Gamma}_{t+1,0}^{\tilde{r}\nu} = \boldsymbol{A}_t\boldsymbol{\Gamma}_{t0}^{\tilde{r}\nu}$$

 Iterate $s = 0, \ldots, t$

$$\boldsymbol{\Sigma}_{(t+1)s}^{\tilde{r}\tilde{r}} = \boldsymbol{A}_t\boldsymbol{\Sigma}_{ts}^{\tilde{r}\tilde{r}}$$
$$\boldsymbol{\Gamma}_{(t+1)s}^{\tilde{r}d} = \boldsymbol{A}_t\boldsymbol{\Gamma}_{ts}^{\tilde{r}d}$$

 End iterate s

- *End iterate* t :

The Gaussian $(n(T+2) + q + m(T+1))$-*vector* $[\tilde{\boldsymbol{r}}, \boldsymbol{\nu}, \boldsymbol{d}]$ *is then fully assessed by the algorithm.*
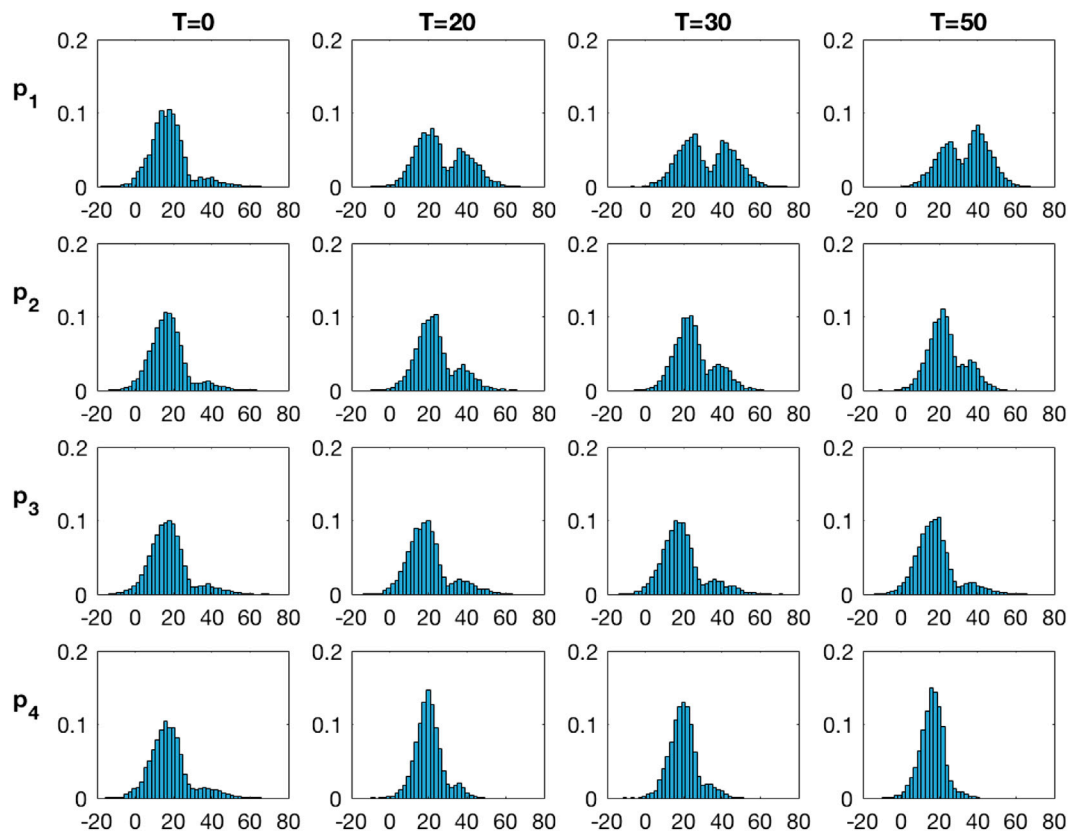
FIGURE 8 | Marginal pdfs at monitoring locations for increasing current time T from the inversion with the selection Kalman model.

conditioning on d and then on v. The final step, conditioning on $v \in A$, is computer demanding even though v has only dimension q. Rejection sampling is only possible for very low values of q. We therefore use the Metropolis-Hastings algorithm, a McMC method, detailed in [26] and extended from [30]. **Algorithm 1** requires storing the covariance $[n(T+2)+q+m(T+1)] \times [n(T+2)+q+m(T+1)]$ -matrix of the augmented Gaussian vector $[\tilde{r}, v, d]$ which can usually be avoided in practice. Only the spatial variables of interest need to be stored, which entails that only the covariance $[n+q+m(T+1)] \times [n+q+m(T+1)]$-matrix of $[\tilde{r}_0, v, d]$ need to be stored in our spatio-temporal inversion study.

4 SYNTHETIC STUDY

The synthetic study is introduced in **Section 2**, and we discuss it in larger detail in this section.

4.1 Model

Consider a discretized spatio-temporal continuous RF representing the evolution of a temperature field $\{r_t(x), x \in \mathcal{L}_r\}$, $t \in \mathcal{T}_r : \{0, 1, \ldots, T, T+1\}$; $r_t(x) \in \mathbb{R}$, as defined in **Section 2**. The number of spatial grid nodes is $n = 21 \times 21$, while temporal reference T is the current time up to $T = 50$. The discretized spatial field at time t is represented by the n-vector r_t.

The initial temperature field r_0, given in **Figure 1**, is assumed to be unknown. It is divided into two distinct areas: the blue area where the temperature is set at $20°C$ and the yellow area where the temperature is set at $45°C$. Assume that, given the initial temperature field, the field evolves according to the advection-diffusion equation, a linear partial differential equation,

$$\frac{\partial r_t(x)}{\partial t} - \lambda \nabla^2 r_t(x) + c \cdot \nabla r_t(x) = 0 \qquad (14)$$

$$\nabla r_t(x) \cdot n = 0 \qquad (15)$$

with $\lambda \in \mathbb{R}_+$ the known diffusivity coefficient, n the outer normal to the domain and $c = [c_1, c_2]$ the known velocity field. The forward model is defined as,

$$[r_{t+1}|r_t] = A r_t + \varepsilon_t^r \qquad (16)$$

$$f(r_{t+1}|r_t) = \varphi_n(r_{t+1}; A r_t, \Sigma_t^{r|r}) \qquad (17)$$

where the $(n \times n)$-matrix A is obtained by discretizing the advection-diffusion equation using finite differences, see **Supplementary Appendix B** for finite differences scheme and parameter values, while the centered Gaussian n-vector ε_t^r, with covariance $(n \times n)$-matrix $\Sigma_t^{r|r} = 0 \times I_n$ represents model error. Under these assumptions, the forward model is exact which constitutes a limiting case to Gauss-linear models. The evolution of the temperature field is described in **Figure 2**.
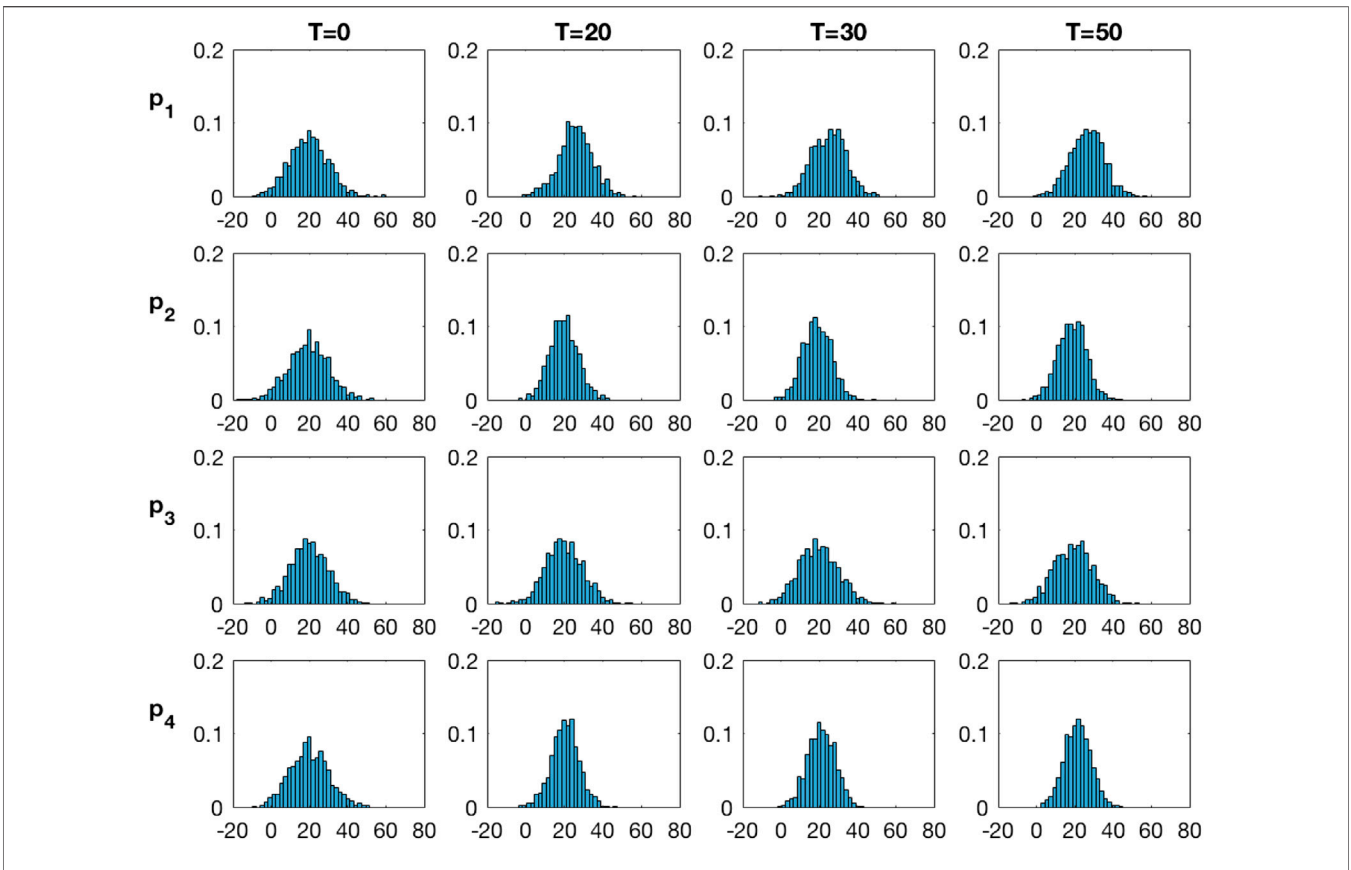
FIGURE 9 | Marginal pdfs at monitoring locations for increasing current time *T* from the inversion with the traditional Kalman model.
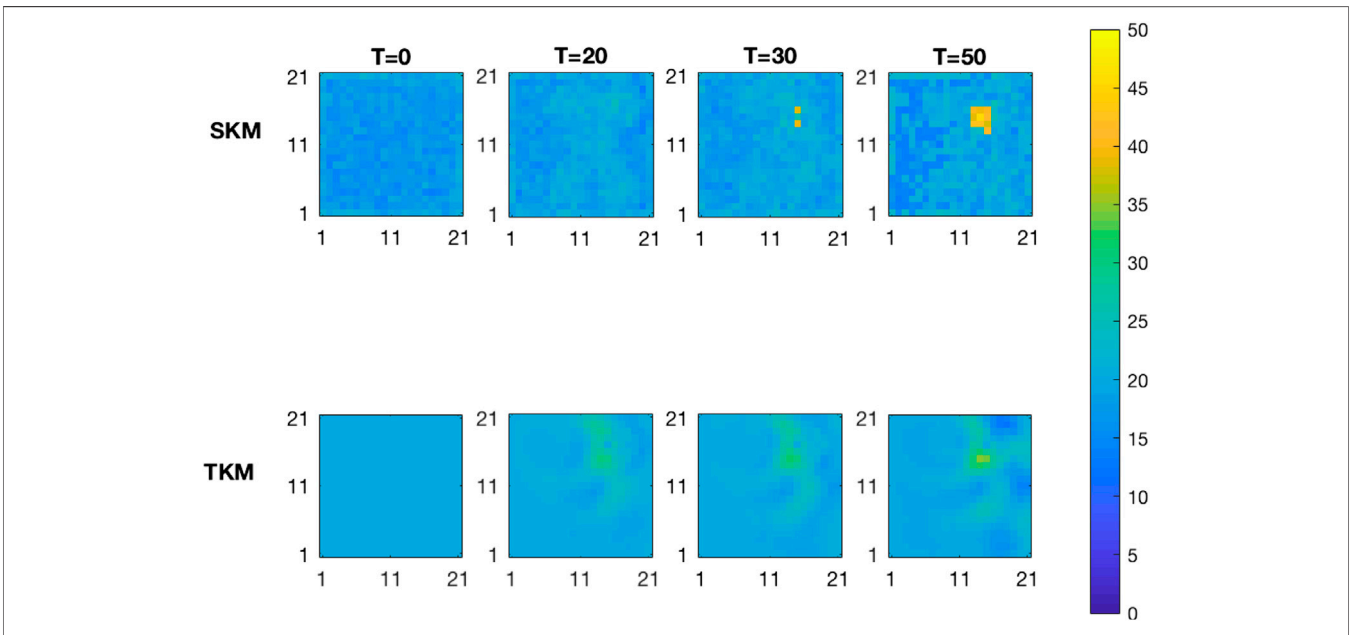
FIGURE 10 | MMAP predictions of the initial state for increasing current time *T* from the inversion with the selection Kalman model (SKM-upper) and with the traditional Kalman model (TKM-lower).
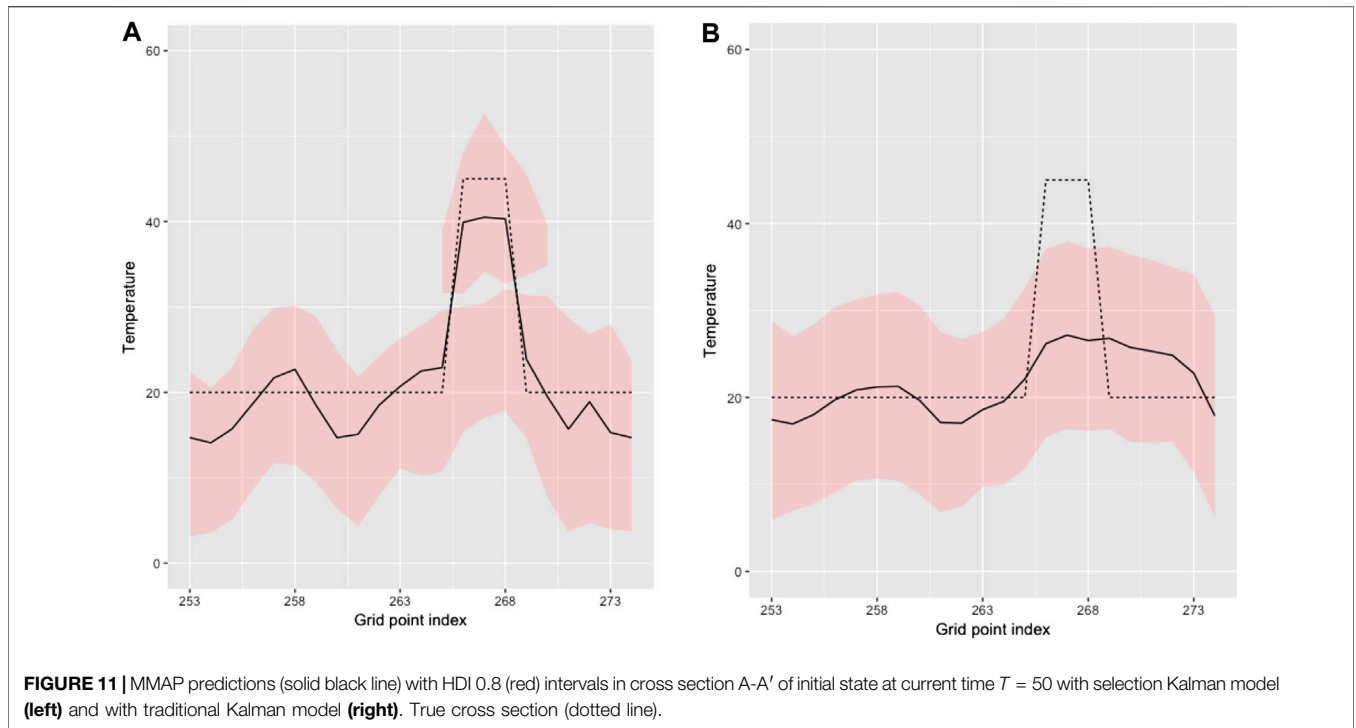
FIGURE 11 | MMAP predictions (solid black line) with HDI 0.8 (red) intervals in cross section A-A′ of initial state at current time $T = 50$ with selection Kalman model **(left)** and with traditional Kalman model **(right)**. True cross section (dotted line).

The observations are acquired at $m = 5$ different locations on the spatial grid $\mathcal{L}_r$ at each temporal node in $\mathcal{T}_d$ providing the set of m-vectors $\{d_t, t \in \mathcal{T}_d\}$. The observation locations are represented with dots in **Figure 1**. The corresponding likelihood model is defined as,

$$[d_t | r_t] = Hr_t + \varepsilon_t^d \tag{18}$$

$$f(d_t | r_t) = \varphi_m\big(d_t; Hr_t, \Sigma_t^{d|r}\big) \tag{19}$$

where the observation $(m \times n)$-matrix H is a binary selection matrix, see **Supplementary Appendix B**, while the centered Gaussian m-vector ε_t^d with covariance $(m \times m)$-matrix $\Sigma_t^{d|r} = \sigma_{d|r}^2 \times I_m$ with $\sigma_{d|r} = 0.1$, represents independent observation errors. Under these assumptions, the likelihood model is Gauss-linear. The observations are displayed as time series in **Figure 3**. Note that $\Sigma_t^{r|r}$ and $\Sigma_t^{d|r}$ are in this example constant through time.

The forward and likelihood models are Gauss-linear. In order to fully defined the selection Kalman model and traditional Kalman model, we must specify the prior distribution for the initial temperature field for both approaches.

We assume we know the initial temperature field has large areas with low temperatures in the range $[5^\circ\mathrm{C}, 25^\circ\mathrm{C}]$ and possibly, smaller areas with high temperatures in the range $[40^\circ\mathrm{C}, 55^\circ\mathrm{C}]$. The exact location, extent and temperature of these smaller areas are unknown. The prior is therefore spatially stationary in both models.

The prior distribution is set to be selection-Gaussian for the selection Kalman model. Such a prior model can represent multimodality. The model is constructed according to [26] and is defined considering an auxiliary discretized stationary Gaussian RF,

$$f\left(\tilde{r}_0\right) = \varphi_n\big(\tilde{r}_0; \mu_{\tilde{r}} i_n, \sigma_{\tilde{r}}^2 \Sigma_{\tilde{r}}^\rho\big) \tag{20}$$

with expectation and variance levels, $\mu_{\tilde{r}}$ and $\sigma_{\tilde{r}}^2$ respectively. The spatial correlation $(n \times n)$-matrix $\Sigma_{\tilde{r}}^\rho$ is defined by an isotropic second order exponential spatial correlation function $\rho_{\tilde{r}}(\tau) = \exp(-\tau^2/\delta^2); \tau \in \mathbb{R}_+$. Define the auxiliary n-vector ν given $\tilde{r}_0$,

$$[\nu | \tilde{r}_0] = \gamma(\tilde{r}_0 - \mu_{\tilde{r}} i_n) + \varepsilon^\gamma \tag{21}$$

$$f(\nu | \tilde{r}_0) = \varphi_n\big[\nu; \gamma(\tilde{r}_0 - \mu_{\tilde{r}} i_n), (1 - \gamma^2) I_n\big] \tag{22}$$

$$= \prod_{i=1}^n \varphi_1\big[\nu_i; \gamma(\tilde{r}_{0,i} - \mu_{\tilde{r}}), (1 - \gamma^2)\big] \tag{23}$$

with coupling parameter $\gamma \in \mathbb{R}_{[-1,1]}$ and centered Gaussian independent n-vector ε^γ with variance $(1 - \gamma^2)$. Note that this pdf is in factored form. Consequently the joint pdf of $[\tilde{r}_0, \nu]$ is,

$$\begin{bmatrix} \tilde{r}_0 \\ \nu \end{bmatrix} \sim \varphi_{2n}\left(\begin{bmatrix} \tilde{r}_0 \\ \nu \end{bmatrix}; \begin{bmatrix} \mu_{\tilde{r}} i_n \\ 0 i_n \end{bmatrix}, \begin{bmatrix} \sigma_{\tilde{r}}^2 \Sigma_{\tilde{r}}^\rho & \sigma_{\tilde{r}}^2 \gamma \Sigma_{\tilde{r}}^\rho \\ \sigma_{\tilde{r}}^2 \gamma \Sigma_{\tilde{r}}^\rho & \sigma_{\tilde{r}}^2 \gamma^2 \Sigma_{\tilde{r}}^\rho + (1 - \gamma^2) I_n \end{bmatrix}\right) \tag{24}$$

Define a separable selection set $A \subset \mathbb{R}^n$ such that $A = B^n, B \subset \mathbb{R}$. Therefore, the prior distribution is represented by the discretized selection-Gaussian RF $r_{A,0}$ defined as,

$$r_{A,0} = [\tilde{r}_0 | \nu \in A] \tag{25}$$

$$f(r_{A,0}) = \big[\Phi_n\big(A; 0 i_n, \sigma_{\tilde{r}}^2 \gamma^2 \Sigma_{\tilde{r}}^\rho + (1 - \gamma^2) I_n\big)\big]^{-1}$$
$$\times \prod_{i=1}^n \Phi_1\big[A_i; \gamma(\tilde{r}_i - \mu_{\tilde{r}}), (1 - \gamma^2)\big] \tag{26}$$
$$\times \varphi_n\big(r_{A,0}; \mu_{\tilde{r}} i_n, \sigma_{\tilde{r}}^2 \Sigma_{\tilde{r}}^\rho\big)$$

Note that after selection on the auxiliary variable ν is made, the expectation and variance of the resulting $r_{A,0}$ will no longer be $\mu_{\tilde{r}} i_n$ and $\sigma_{\tilde{r}}^2 \Sigma_{\tilde{r}}^\rho$.
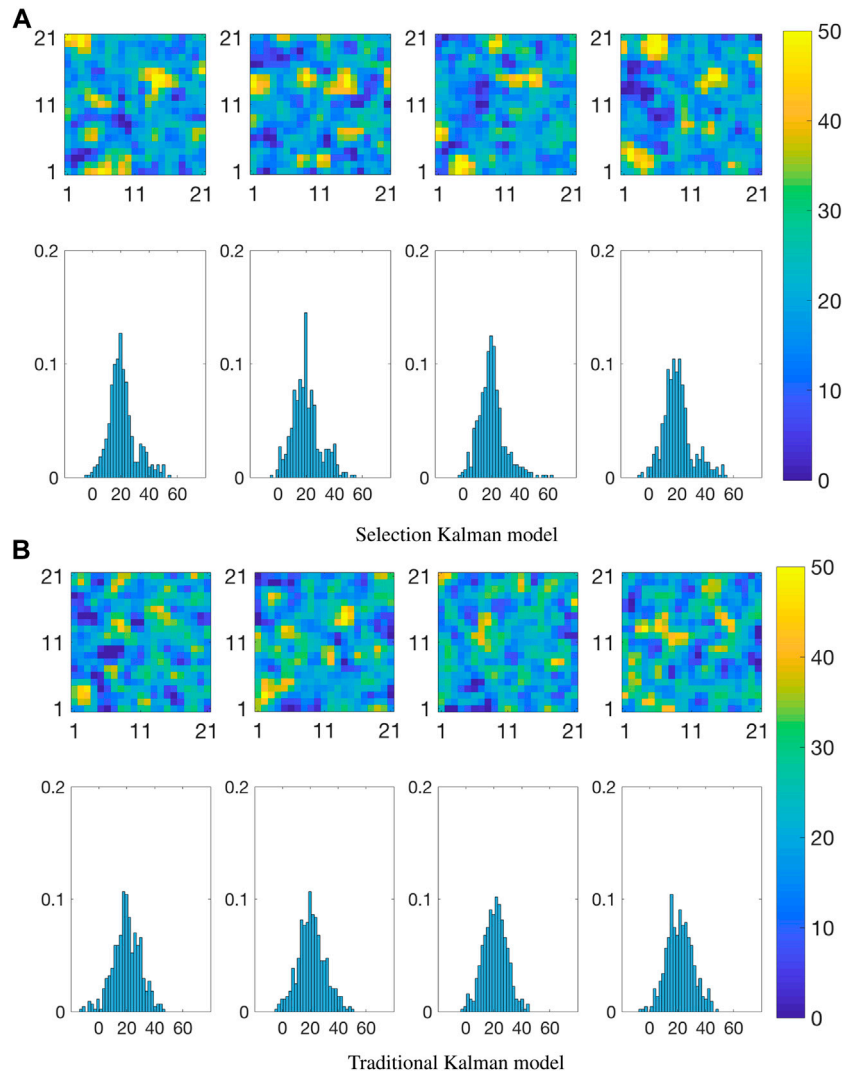
FIGURE 12 | Realizations from the posterior distribution of the initial state at current time $T = 50$.

The parameters values are listed in **Table 1** and they are chosen to reflect what is assumed about the initial temperature field. The prior marginal distribution is bimodal, the dominant mode is centered about 18°C while the smaller mode is centered about 40°C as shown in **Figure 6**. The spread of the dominant mode covers the assumed temperature range for the low temperature areas while the spread of the smaller mode covers the assumed range for the high temperature areas. Realizations from the prior distribution and associated spatial histograms are shown in **Figure 7A**. They exhibit large areas at low temperatures and smaller areas at higher temperatures. Similarly to the marginal distribution, the spatial histograms cover the assumed range for high and low temperature areas.

The prior distribution for the traditional Kalman model is Gaussian and is defined as,

$$f(r_0) = \varphi_n\left(r_0; \mu_r i_n, \sigma_r^2 \Sigma_r^\rho\right) \tag{27}$$

with expectation and variance levels, μ_r and σ_r^2, respectively and spatial correlation $(n \times n)$-matrix Σ_r^ρ defined by a second order spatial correlation function $\rho_r(\tau) = \exp\left(-\tau^2/\delta^2\right); \tau \in \mathbb{R}_+$. The parameter values are listed in **Table 1**.

Figure 7B displays four realizations with associated spatial histograms from the prior distribution for the traditional Kalman model. The mean and variance levels are chosen so that the prior covers the assumed range for the high and low temperature areas as can be seen in the spatial histograms.

TABLE 1 | Parameter values for the selection Gaussian initial model (SKM) and the Gaussian initial model (TKM).

SKM	$\mu_{\tilde{r}}$	$\sigma_{\tilde{r}}$	δ	Γ	A
	28.75	10	0.15	0.95	$[[-\infty, -0.2)\cup[0.5, +\infty)]^n$
TKM	μ_r	σ_r	δ		
	20	10	0.15		

Figure 7B can be compared to **Figure 7A**, and one observes that only the selection-Gaussian distribution can capture bimodality in the spatial histograms. In studies with real data, the prior model specification must of course be based on experience with the phenomenon under study.

In the next section, we demonstrate the effect of specifying different initial models in the spatio-temporal inversion model.

4.2 Results

The challenge is to restore r_0 based on the observations $d = \{d_0, \ldots, d_T\}$ by evaluating the posterior distribution in the selection Kalman model $f(r_{A,0}|d_0, \ldots, d_T)$ and in the traditional Kalman filter $f(r_0|d_0, \ldots, d_T)$. We compare the results from these two models that have been specified in the previous section. The posterior distributions are analytically tractable for both the selection Kalman model and the traditional Kalman model. They are calculated using **Algorithm 1** and **Supplementary Appendix Algorithm A1** respectively. In order to evaluate the results, we present various characteristics of the posterior distributions for increasing values of current time T:

1. Marginal pdfs at four monitoring locations represented by crosses and numbered $1, 2, 3, 4$ in **Figure 1**,

$$f\left(r_{A,0,i}|d_{0:T}\right) = \int f\left(r_{A,0}|d_{0:T}\right) dr_{A,0,-i} \quad i = 1, \ldots, 4 \quad (28)$$

and similarly for $f(r_{0,i}|d_{0:T})$ based on $f(r_0|d_{0:T})$. The index $-i$ stands for all the indices in $1, \ldots, n$ but the ith index.

2. Spatial prediction based on a marginal maximum a posteriori (MMAP) criterion,

$$\hat{r}_{A,0} = MMAP\{r_{A,0}|d_{0:T}\} = \{MAP\{r_{A,0,j}|d_{0:T}\}; j = 1, 2, \ldots, n\}$$
$$= \{\arg\max\{f\left(r_{A,0,j}|d_{0:T}\right)\}, j = 1, 2, \ldots, n\} \quad (29)$$

and similarly for $\hat{r}_0$ based on $f(r_0|d_{0:T})$. This MMAP criterion is used as the marginal posterior model may be multi-modal. For uni-modal symmetric posterior distributions such as the Gaussian one, the MMAP predictor coincides with the expectation predictor.

3. The MMAP prediction and the associated 0.80 prediction interval along a horizontal profile A-A', see **Figure 1**. The prediction interval is computed as the highest density interval (HDI) [31], which entails that the prediction intervals may consist of several intervals for multimodal posterior pdfs.

4. Realizations From the Posterior Pdfs $f(r_{A,0}|d_{0:T})$ and $f(r_0|d_{0:T})$.

Figure 8 displays the marginal posterior pdfs based on the selection Kalman model at the four monitoring locations,
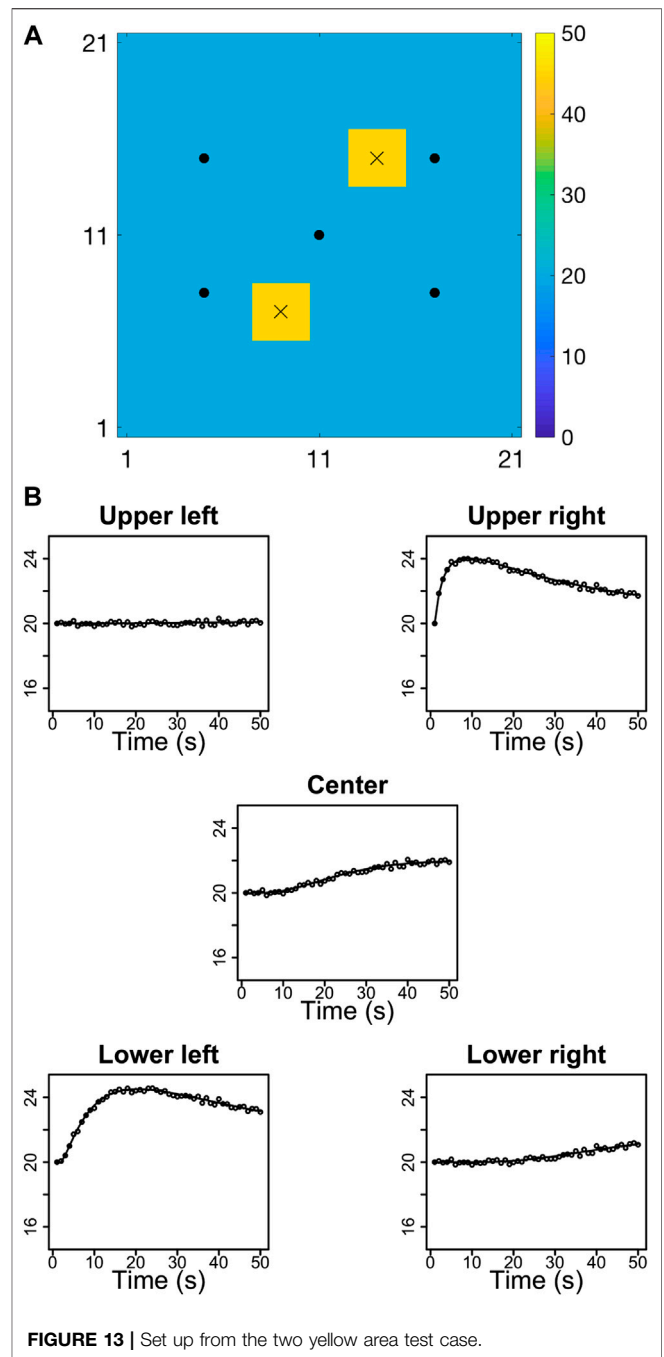

FIGURE 13 | Set up from the two yellow area test case.

vertically, for increasing current time T, horizontally. At current time $T = 0$, all pdfs are virtually identical to the marginal pdf of the stationary initial model. As current time T increases, and the observations are assimilated, one observes substantial differences between the marginal pdfs at the monitoring locations. The height of the high-value mode increases depending on the proximity of monitoring location to the yellow area, as expected. The posterior marginal pdf at observation location one clearly indicates that it lies in the yellow area already at current time $T = 20$ as the high-value mode is increasing. At location two the high-value mode also

FIGURE 14 | Results from the two-yellow-area test case.
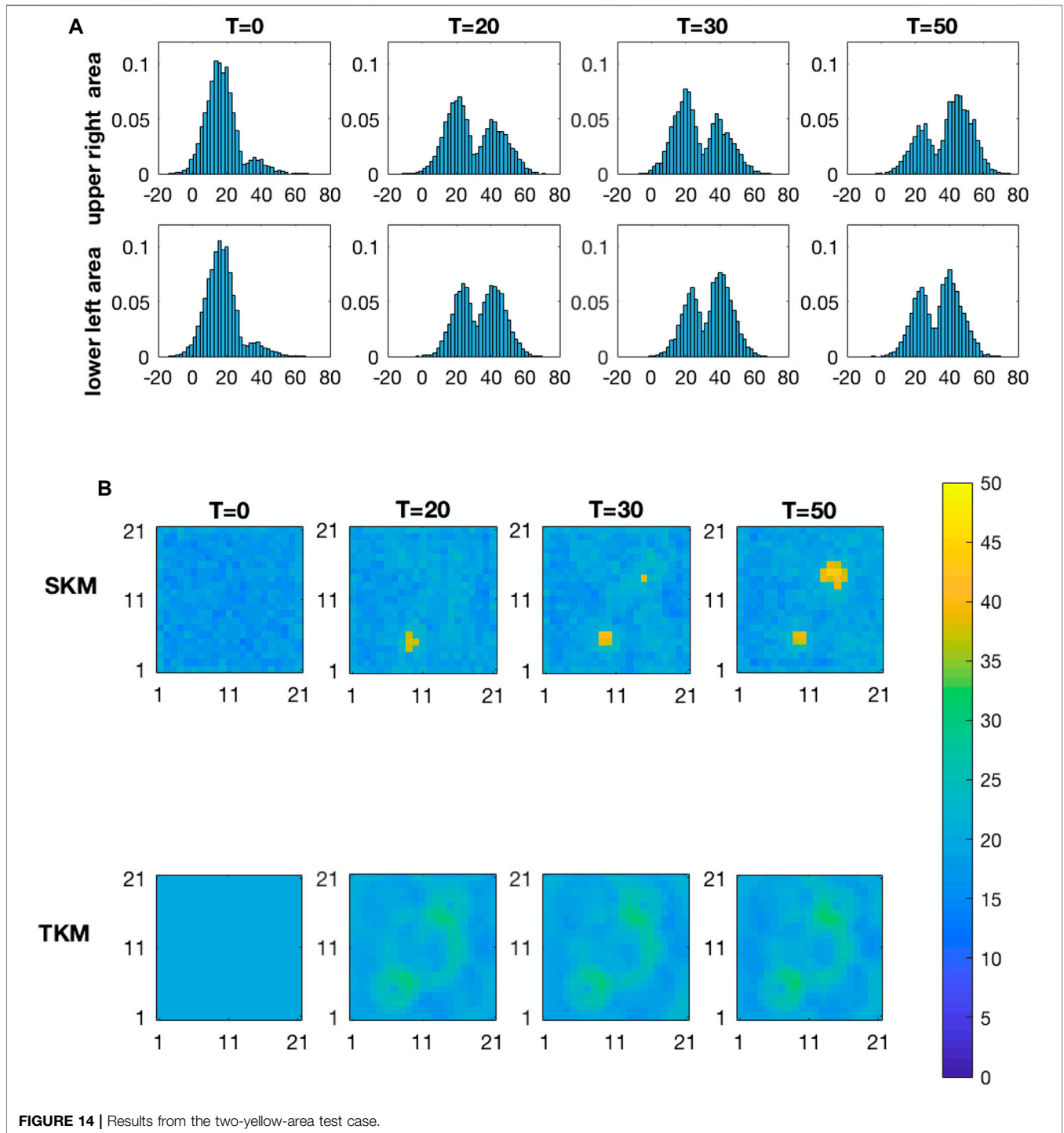
increases somewhat at $T = 20$, but does not increase more thereafter. This monitoring location is outside the yellow area, although fairly close to it. Location three is far from both the yellow area and observation locations and the posterior marginal pdf remains almost identical to the prior model. Lastly, location four is far from the yellow area but close to an observation location at which the observations remain

stationary, hence the low-value mode grows to be completely dominant.

Figure 9 displays the marginal pdfs from the traditional Kalman model. These marginal posterior pdfs are also virtually identical at current time $T = 0$. As current time T increases the marginal pdfs at the monitoring locations are indeed different as they are shifting. However, this shift is

difficult to observe. By using the selection Kalman model, the indications of a yellow area at the correct location can be observed from current time $T = 20$, while one can hardly observe any indications of it if the traditional Kalman model is used.

The upper panels of **Figure 10** display the MMAP spatial prediction based on the selection Kalman model for increasing current time T. At current time $T = 0$, the predictions are virtually constant bar some boundary effect as the initial prior model is stationary. As current time T increases, indications of the yellow appear at $T = 30$, it is however at $T = 50$ that correct location and spatial extent are identified. The prediction value of the yellow area is very close to the correct value of 45. The blue area value is predicted with some variability around the expected 20. The lower panels of **Figure 10** present the corresponding spatial predictions based for the traditional Kalman model. As current time T increases, indications of something occurring in the yellow area appears, but the location is uncertain and the spatial extent only vaguely outlined. Moreover the predicted value in the yellow area is much lower than the correct value 45. The background value is however fairly precisely predicted around the expected 20. The circular features centered about the observation locations that appear on the predictions based on the selection Kalman model in **Figure 10** are not artifacts. These features are also present on the predictions based on the traditional Kalman model, although less prominent.

We evaluate the root mean square error (RMSE) values of the two models at time $T = 50$. The RMSE criterion is used to quantify the difference between the MMAP predictions in **Figure 10** and the truth in **Figure 1**. The RMSE for the selection Kalman model is 2.76 while the RMSE of the traditional Kalman model is 3.33 The selection Kalman model therefore offers a 18% reduction in RMSE compared to the RMSE of the traditional Kalman model.

Figure 11 displays the MMAP predictions with associated 0.80 prediction intervals along the horizontal profile A-A'. The prediction from the selection Kalman model captures the yellow area while the prediction from the traditional Kalman model barely indicates the yellow area. The prediction intervals follow the same pattern. Note, however, that the prediction intervals of the selection Kalman model may appear as two intervals close to the yellow area since the marginal posterior models are bimodal. By using the selection Kalman model, the location, spatial extent and value of the yellow area is very precisely predicted at current time $T = 50$. Predictions based on the traditional Kalman model are less precise and rather blurred.

Figure 12 displays realizations from the posterior pdfs at $T = 50$. For the selection Kalman model, see **Figure 12A**, the yellow area is precisely reproduced in the majority of realizations while for traditional Kalman model, see **Figure 12B**, the yellow area is only vaguely indicated. Note however that the realizations from the selection Kalman model reflect the bimodality of the prior model outside the central area where the five spot observation design provides the most information. These observations are consistent with the results observed in **Figures 8**, **9**.

Conditioning on the observed data takes the same time for both methods but the selection Kalman model requires sampling from a high dimensional truncated Gaussian pdf in order to evaluate the posterior distribution which means that the computational demand for the selection Kalman model is higher than that of the traditional Kalman model. For $n = 441$, as in our study, it takes an average of $7.4s$ to generate 100 realizations from a selection-Gaussian on our laptop computer, so a little over 12 min to generate the 10,000 realizations used to estimate the MMAP for the selection Kalman model. Note that the sampling becomes increasingly more resource consuming as the grid dimension increases and the computational time can be reduced by introducing parallelization in the algorithm.

To demonstrate the generality of the selection Kalman model, we define an alternative true initial state with two yellow areas, see **Figure 13A**. We used the same model parameters as in the primary case. The prior distribution for both the selection Kalman model and the traditional Kalman model are identical to the first case. Note in particular that the number of yellow areas is not specified. The observed time series will of course be different, see **Figure 13B**. These time series have many similarities with the ones from the primary case. We inspect the marginal pdfs at two monitoring locations, one inside each yellow area, as they evolve with current time T, see **Figure 14A**. Both marginal pdfs are identical at current time T = 0, and as current time T increases the height of the high-value mode increases, indicating that both monitoring locations are within the yellow areas. In **Figure 14B** the corresponding MMAP predictions are displayed for increasing current time T. We observe that location, areal extent and value of both yellow areas are well reproduced, but not as well as for the first case since identifying two sources is more complicated. The identification challenge is of course increasing with an increasing number of yellow areas. **Figure 14B** also displays the MMAP prediction for the traditional Kalman model for the two-yellow-area case, where location, areal extent and value are hard to evaluate, similarly to the first test case.

5 CONCLUSION

We define a selection Kalman model based on a selection-Gaussian initial distribution and Gauss-linear dynamic and observation models. This model may represent spatial phenomena with initial states with spatial histograms that are multimodal, skewed and/or peaked. The selection Kalman model is demonstrated to be contained in the class of selection-Gaussian distributions and hence analytically tractable. The analytical tractability makes the assessment of the selection-Gaussian posterior model fast and reliable. Moreover, an efficient recursive algorithm for assessing the selection Kalman model is specified. Note that the traditional Kalman model is a special case of the selection Kalman model, hence the latter can be seen as a generalization of the former.

A synthetic spatio-temporal inversion case study with Gauss-linear forward and observation models is used to demonstrate the characteristics of the methodology. We specify both a selection Kalman model and a traditional Kalman model and evaluate their ability to restore the initial state based on the observed time series. The time series are noisy observations of the variable of interest collected at

a set of sites. The selection Kalman model outperforms the traditional Kalman model. The former model identifies location, areal extent and value of the yellow area very reliably. The traditional Kalman model only provides blurry indications with severe under-prediction of the yellow area. We conclude that for spatio-temporal inversion where the initial spatial state has bimodal or multimodal spatial histograms, the selection Kalman model is far more suitable than the traditional Kalman model.

The selection Kalman model has potential applications far beyond the simple case evaluated in this case study. For all spatio-temporal problems where multimodal spatial histograms appear, the selection Kalman model should be considered. The model can easily be extended to a selection extended Kalman model, along the lines of the extended Kalman model. A more challenging and interesting extension is the definition of a selection ensemble Kalman model including non-linear dynamic and observation models. Research along these lines is currently taking place.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MC and HO contributed equally to this manuscript in every matter except implementation for which was carried out exclusively by MC.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2021.636524/full#supplementary-material.

REFERENCES

1. Handcock MS, Wallis JR. An approach to statistical spatial-temporal modeling of meteorological fields. *J Am Stat Assoc* (1994). 89:368–78. doi:10.1080/01621459.1994.10476754

2. Cappé O, Moulines E, Ryden T. *Inference in hidden Markov models (Springer series in statistics)*. Berlin, Heidelberg: Springer-Verlag (2005).

3. Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng* (1960). 82:35–45. doi:10.1115/1.3662552

4. Emerick AA, Reynolds AC. Ensemble smoother with multiple data assimilation. *Comput Geosci* (2013). 55:3–15. doi:10.1016/j.cageo.2012.03.011

5. Evensen G, Raanes PN, Stordal AS, Hove J. Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Front Appl Math Stat* (2019). 5:47. doi:10.3389/fams.2019.00047

6. Liu M, Grana D. Time-lapse seismic history matching with iterative ensemble smoother and deep convolutional autoencoder. *Geophysics* (2019). 85:1–63. doi:10.1190/geo2019-0019.1

7. Hopke PK. Review of receptor modeling methods for source apportionment. *J Air Waste Manag Assoc* (2016). 66:237–59. doi:10.1080/10962247.2016.1140693

8. Cheng MD, Lin CJ. Receptor modeling for smoke of 1998 biomass burning in Central America. *J Geophys Res Atmos* (2001). 106:22871–86. doi:10.1029/2001JD900024

9. Kjeldsen P, Bjerg PL, Rügge K, Christensen TH, Pedersen JK. Characterization of an old municipal landfill (Grindsted, Denmark) as a groundwater pollution source: landfill hydrology and leachate migration. *Waste Manage Res* (1998). 16:14–22. doi:10.1177/0734242X9801600103

10. Wikle CK, Cressie N. A dimension-reduced approach to space-time Kalman filtering. *Biometrika* (1999). 86:815–29. doi:10.1093/biomet/86.4.815

11. Todescato M, Carron A, Carli R, Pillonetto G, Schenato L. Efficient spatio-temporal Gaussian regression via Kalman filtering. *Automatica* (2020). 118:109032. doi:10.1016/j.automatica.2020.109032

12. Jazwinski AH. *Stochastic processes and filtering theory*. New York, NY: Academic Press (1970).

13. Julier SJ, Uhlmann JK. New extension of the Kalman filter to nonlinear systems. In: I Kadar, editor. *Signal processing, sensor fusion, and target recognition VI (SPIE)*, 3068 (1997). p. 182–93. doi:10.1117/12.280797

14. Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J Geophys Res* (1994). 99:101–43. doi:10.1029/94JC00572

15. Evensen G, van Leeuwen PJ. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Rev* (2000). 128:1852–67. doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2

16. Li R, Prasad V, Huang B. Gaussian mixture model-based ensemble Kalman filtering for state and parameter estimation for a PMMA process. *Processes* (2016). 4:9. doi:10.3390/pr4020009

17. Katzfuss M, Stroud JR, Wikle CK. Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *J Am Stat Assoc* (2020). 115:866–85. doi:10.1080/01621459.2019.1592753

18. Robert C, Casella G. *Monte Carlo statistical methods. Springer Texts in Statistics*. New York, NY: Springer (2005).

19. Naveau P, Genton MG, Shen X. A skewed Kalman filter. *J Multivariate Anal* (2005). 94:382–400. doi:10.1016/j.jmva.2004.06.002

20. Ackerson G, Fu K. On state estimation in switching environments. *IEEE Trans Automatic Control* (1970). 15:10–7. doi:10.1109/TAC.1970.1099359

21. Chen R, Liu JS. Mixture Kalman filters. *J R Stat Soc Ser B (Statistical Methodology)* (2000). 62:493–508. doi:10.1111/1467-9868.00246

22. Smith KW. Cluster ensemble kalman filter. *Tellus A* (2007). 59:749–57. doi:10.1111/j.1600-0870.2007.00246.x

23. Dovera L, Della Rossa E. Multimodal ensemble Kalman filtering using Gaussian mixture models. *Comput Geosciences* (2011). 15:307–23. doi:10.1007/s10596-010-9205-3

24. Bengtsson T, Snyder C, Nychka D. Toward a nonlinear ensemble filter for high-dimensional systems. *J Geophys Res Atmos* (2003). 108. doi:10.1029/2002JD002900

25. Ulvmoen M, Omre H. Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 1—methodology. *Geophysics* (2010). 75(2):R21–R35. doi:10.1190/1.3294570

26. Omre H, Rimstad K. Bayesian spatial inversion and conjugate selection Gaussian prior models. *Methodology*. Preprint: arXiv:1812.01882 (2018).

27. Arellano-Valle RB, Branco MD, Genton MG. A unified view on skewed distributions arising from selections. *Can J Stat* (2006). 34:581–601. doi:10.1002/cjs.5550340403

28. Arellano-Valle R, del Pino G. From symmetric to asymmetric distributions: a unified approach. In: M Genton, editor. *Skew-elliptical distributions and their applications: a journey beyond normality*. New York, NY: Chapman and Hall/CRC (2004). p. 113–33.

29. Moja S, Asfaw Z, Omre H. Bayesian inversion in hidden Markov models with varying marginal proportions. *Math Geosci* (2018). 51:463–84. doi:10.1007/s11004-018-9752-z

30. Robert CP. Simulation of truncated normal variables. *Stat Comput* (1995). 121–5. doi:10.1007/BF00143942

31. Hyndman RJ. Computing and graphing highest density regions. *The Am Statistician* (1996). 50:120–6. doi:10.2307/2684423