Sequential hierarchical Bayesian model and particle filter estimation with two-step RJMCMC resampling

Huan, Yue; Wang, Guoqiang; Lin, Hai Xiang

DOI
10.1016/j.csda.2025.108304

Publication date
2026

Document Version
Final published version

Published in
Computational Statistics and Data Analysis

Computational Statistics and Data Analysis 216 (2026) 108304

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Sequential hierarchical Bayesian model and particle filter estimation with two-step RJMCMC resampling

Yue Huan [a,*], Guoqiang Wang [b], Hai Xiang Lin [a]

[a] Delft Institute of Applied Mathematics (DIAM), Delft University of Technology, Delft, 2628 CD, The Netherlands
[b] School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai, 201620, China

ARTICLE INFO

Keywords:
Data assimilation
State-space models
Bayesian inference
Parameter estimation
Model identification

ABSTRACT

Data assimilation (DA) combines numerical model simulations with observed data to obtain the best possible description of a dynamical system and its uncertainty. Incorrect modeling assumptions can lead to filter divergence, making model identification an important issue in the field of DA. Variations in dynamic model structures can result in differences in parameter dimensions, complicating the resampling step in PFs. To meet this challenge, the Sequential Hierarchical Bayesian Model (SHBM) is proposed in this paper, which integrates the evolution model along with observation model from the DA scheme, and the hierarchical parameter model. A two-step resampling method are also proposed to estimate the SHBM: the first step uses the resampling scheme in the bootstrap filter to resample new particles based on weights, which may produce some duplicate particles; the second step utilizes the Reversible Jump Markov Chain Monte Carlo (RJMCMC) methods to draw new particles from the target distribution. This approach ensures particle diversity, with the first step aiming at avoiding particle degeneracy, and the second step intends to prevent the sample impoverishment. The performance in the Advection Equation example and Lorenz 96 example demonstrates the effectiveness of the proposed method.

1. Introduction

Dynamical System is used to describe systems that evolve over time, which is widely applied in fields such as meteorological science and signal processing. Building numerical models and estimating with observational data are two important tools for studying dynamical systems. However, computational errors and stability issues often make it difficult for numerical models to unbiasedly and accurately characterize dynamical systems, while observations can only provide indirect and partial information.

Data assimilation (DA) combines the numerical models and observations to estimate the interesting states of the dynamical systems (Gordon et al., 1993; Evensen, 1994). The numerical models are also called the evolution equations, which describe the evolution of the systems, as well as the interesting states. Typically, the DA problem is solved sequentially over a sequence of assimilation time windows. Let $M : \mathbb{R}^n \to \mathbb{R}^n$ denote the evolution model, which is a mapping of the n-dimensional state space into itself. It is assumed that the state of the dynamic system x_t at time t evolves according to the nonlinear difference equation

$$x_t = M(x_{t-1}|\theta) + \epsilon_t, \tag{1}$$

where $\theta \in \Theta$ is the parameter in the evolution model, and ϵ_t is the $n-$dimensional vector denoting the external forcing. The subscript t is a time index and $t = 1, 2 \cdots$, and x_0 is the initial condition. It is usually assumed that ϵ_t is a white noise sequence (Gordon et al., 1993;

* Corresponding author.
E-mail address: Y.Huan@tudelft.nl (Y. Huan).

https://doi.org/10.1016/j.csda.2025.108304
Received 27 February 2025; Received in revised form 28 October 2025; Accepted 10 November 2025
Available online 17 November 2025

Lewis et al., 2006; Vrugt et al., 2013), and $E(\epsilon_t) = 0$. It is serially uncorrelated, that is, $E(\epsilon_t \epsilon_r) = 0$ for $r \neq t$ and $Cov(\epsilon_t) = E(\epsilon_t {\epsilon_t}^T) = Q_t \in \mathbb{R}^{n\times n}$, is a known symmetric and positive definite matrix. The observation operator $h : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping from the model space, $\mathbb{R}^n$ to the observation space $\mathbb{R}^m$,

$$z_t = h(x_t) + \upsilon_t, \tag{2}$$

defines in general, a nonlinear relationship between the observations z_t and the state x_t, $t = 1, 2, \ldots$. $\upsilon_t \in \mathbb{R}^m$ is a white noise sequence with $E(\upsilon_t) = 0$ and $Cov(\upsilon_t) = R_t \in \mathbb{R}^{m\times m}$ and R_t is a real symmetric and positive definite matrix (Lewis et al., 2006).

To improve the assimilation and forecast accuracy, the structures of the evolution equation $M(\cdot|\cdot)$ and the parameter uncertainty of the evolution model has been infused in the DA scheme, and DA techniques have been developed into a broader class of methods that can estimate parameters and even include model identification while assimilating states (Wiese et al., 2015; Chang and Zhang, 2019; Bach and Ghil, 2023; Feng and Wang, 2024). The flexibility of particle filter (PF) (Gordon et al., 1993) in model assumptions makes it particularly suitable for addressing this class of problems, as opposed to other DA methods like the Ensemble Kalman filter (Evensen, 1994).

The integration of Markov Chain Monte Carlo (MCMC) techniques, known for their efficiency in sampling from high dimensional probability distributions, substantially improves the performance of PF (Andrieu et al., 2010). Moradkhani et al. (2012) and Vrugt et al. (2013) presented joint inference of model parameters and states by incorporating MCMC into PF, effectively mitigating the challenge of sample impoverishment (Elfring et al., 2021). Wiese et al. (2015) combined PF with Reversible Jump Markov Chain Monte Carlo (RJMCMC) to estimate the directions of arrival of sources when the observation system is uncertain. By infusing RJMCMC into the DA scheme, Huan and Lin (2024) proposed an innovative Bayesian DA framework using data-driven methods for simultaneously and recursively identifying physical processes and estimating states and model parameters. Our previous work focuses on the parameter estimation when the parameters are independent. In this work, we propose a novel approach to enhance the performance of the resampling step in PF using RJMCMC, aiming at better parameter estimation without the assumption of parameter independence.

Assume that the structure of $M(\cdot|\cdot)$ in the evolution Eq. (1) is uncertain but belongs to a set of K alternative models $\{M^1(\cdot|\xi_1), M^2(\cdot|\xi_2), \ldots, M^K(\cdot|\xi_k)\}$ with the unknown parameter ξ_k, which defined on an r_k-dimensional parameter space $\Theta_k \subset \mathbb{R}^{r_k}$. The dimension r_k of ξ_k can vary across different models M^k. Let $\theta \triangleq (k, \xi_k)$, which has different dimensions in different parts of whole parameter space Θ.

Almost invariably in a frequentist setting, inference about these two kinds of unknowns, which are the model indicator k and the corresponding model parameter ξ_k, is based on different logical principles, but, at least formally, the Bayes paradigm offers the opportunity of a single logical framework – it is the joint posterior $f_{k,\xi_k|z_{1:t}} \triangleq f_{\theta|z_{1:t}}$ of model indicator and parameter given data, which is also named observation in DA literature, $z_{1:t}$. In this case, the DA problem needs to consider three objectives simultaneously:

- The posterior distribution of the state x_t at each time t, $f_{x_t|z_{1:t}}$,
- The posterior distribution of θ, $f_{\theta|z_{1:t}}$,
- The probability that the true model is $M^k(\cdot|\xi_k), k = 1, 2, \ldots, K$.

These three objectives correspond to the uncertainties in the state vector, parameters, and model, respectively. Feng and Wang (2024) proposed a Beta-Dirichlet switching state-space transmission model to track underlying dynamics of diseases, while the method is suitable for transformation under a limited number of evolution models, and they considered constant parameters in each evolution model. Hierarchical models can better characterize these three levels of uncertainty, which are introduced in Section 2. To achieve these objectives, we propose a Sequential Hierarchical Bayesian Model (SHBM) in Eq. (3) to model the dynamic spatial-temporal process with unknown parameters. The SHBM includes the identification of unknown models, incorporating both an evolution model and an observation equation for DA proceedings. The SHBM is an extension to the Hierarchical State-Space Model (HSSM), whose Bayesian inference is usually estimated by Monte Carlo methods (Katzfuss et al., 2020). Unlike the HSSM considered in the previous literature, the SHBM incorporates multiple distinct parameter spaces.

To estimate SHBM, we utilize the PF and propose a two-step resampling method tailored to the needs of model identification: the first step uses a resampling scheme in the bootstrap particle filter (Gordon et al., 1993) to resample new particles based on weights, which may produce some duplicate particles; the second step employs the RJMCMC methods to draw new particles from the target distribution. The first step aims at avoiding particle degeneracy, while the second step is intended to prevent sample impoverishment. This approach ensures the diversity of particles at each time step. With our proposed method, we can sequentially assimilate states, estimate parameters, and identify models. Our contribution also includes adjusting the RJMCMC method to make it more suitable for the DA framework and providing detailed theoretical derivations for this model.

The remainder of this paper is organized as follows. In Section 2, we present the theoretical model, and the proposed PF with a two-step resampling is introduced in Section 3. In Section 4, we validate the effectiveness of our new method using a 1-dimensional advection equation. Conclusions are provided in Section 6.

2. The sequential hierarchical Bayesian model

We propose the SHBM here for model selection with sequentially obtained observation data:

$$\begin{aligned} z_t &= h(x_t) + \upsilon_t, \\ x_t &= M^k(x_{t-1}|\xi_k) + \epsilon_t, \\ k &\sim \rho(k), \end{aligned} \tag{3}$$

$$\xi_k \sim \pi_{\xi_k}.$$

The first two equations are the State-Space Model (SSM), describing the evolution of the state of different points in space over time, and the relationship between the observed data and the state (Corenflos et al., 2021). For different k, the dimension of ξ_k can be different from each other. The "variable-dimension" issue can be seen as a joint inference about a model indicator k and a parameter vector ξ_k (Green and Hastie, 2009).

The hierarchical parameter θ for SHBM is composed of k and ξ_k. The probability density function of θ is defined as:

$$\begin{aligned} &\pi_\theta = \pi_{k,\xi_k} = \rho(k) \cdot \pi_{\xi_k}, \quad k = 1, 2, \dots, K, \\ &\theta = (k, \xi_k) \in \Theta = \cup_{k=1}^{K} (\{k\} \times \Theta_k) \subset \cup_{k=1}^{K} (\{k\} \times \mathbb{R}^{r_k}), \end{aligned} \tag{4}$$

where π_{ξ_k} is the conditional density of $\pi_{\xi_{\tilde{k}}}$ given the random index $\tilde{k} = k$, and $\rho(k) = P(\tilde{k} = k)$. Bayesian inference allows us to formulate the DA problem as the joint probability distribution over all observations and unobservable components of interest, such as the states and parameters. Given the prior distribution of θ

$$\pi_\theta^0 = \pi_{k,\xi_k}^0 = \rho^0(k) \cdot \pi_{\xi_k}^0, \tag{5}$$

the posteriors distribution of θ given observations till t $z_{1:t}$ can be acquired through Bayesian Theorem.

The SHBM integrates the DA methods from the geophysics literature (Gordon et al., 1993; Katzfuss et al., 2020) with the trans-dimensional parameter inference from the statistics literature (Green, 1995; Green and Hastie, 2009). It serves as a more general framework than the traditional Hierarchical Bayesian Models (HBM), which offering greater flexibility in representing the uncertainty from multiple sources (Rouder and Lu, 2005; Carlin and Louis, 2008; Gelfand, 2012). An HBM is a probabilistic model that incorporates multiple levels of parameters, each corresponding to a different source of uncertainty. It assumes that the data are generated by a set of parameters at different hierarchical levels, where each level is conditioned on the parameters from the previous level. For further details, we refer readers to Van Leeuwen (2009). Given this hierarchical structure, the SHBM is particularly well-suited for exploring parameter spaces of varying dimensions.

3. Methodological framework

The SHBM can be estimated sequentially using PFs. However, the resampling step of PFs faces two challenges: one is how to resample from the variable-dimensional parameter space $\cup_{k=1}^{K}(\{k\} \times \Theta_k)$, and the other is how to calculate resampling weights corresponding to parameters of different dimensions.

We propose a PF method that incorporates a two-step RJMCMC resampling scheme. In Section 3.1, we describe how to estimate the SSM using PFs when the dimension of θ is fixed. Section 3.2 outlines the estimation of the distribution of hierarchical parameters. In Section 3.3, we present the details of the two-step RJMCMC resampling procedure.

3.1. Sequential Monte Carlo methods

PFs aim to solve a recursive state and parameter estimation problem using an ensemble of weighted particles (Elfring et al., 2021). Each particle represents a sample of (x_t, θ) from the target distribution, denoted as $(x_t^{(i)}, \theta^{(i)})$, and $i \in 1, 2, \dots, N$. For simplicity, we adopt following notational conventions:

$$\begin{aligned} &M(x_t|\theta) = M^k\left(x_t \mid \xi_k\right), \\ &M(x_t^{(i)}|\theta^{(i)}) = M^{k_i}\left(x_t^{(i)} \mid \xi_{k_i}^{(i)}\right), \end{aligned}$$

where $M^{k_i}\left(x_t^{(i)} \mid \xi_{k_i}^{(i)}\right)$ denotes the evolution equation for the state $x_t^{(i)}$ and parameter $\xi_{k_i}^{(i)}$ under the k_i model. The index i refers to the ith ensemble member, and k_i is the model index associated with the ith ensemble member. To simplify the discussion, we first consider θ as a whole instead of the hierarchical parameters k and ξ_k.

At the initial time step $t = 0$, given the initial state x_0 at time $t = 0$ with prior distribution π_{x_0} and the prior distribution $\pi_\theta^0 = \rho^0(k) \cdot \pi_{\xi_k}^0$ for θ, where π_{x_0} and π_θ^0 are independent. The joint density function of x_1, x_0, and parameter θ is given by:

$$\begin{aligned} f_{x_1,x_0,\theta} &= f_{x_1|x_0,\theta} \cdot f_{x_0,\theta} \\ &= \phi(x_1; M(x_0|\theta), \sigma^2 R_n) \cdot \pi_{x_0} \cdot \pi_\theta^0, \end{aligned} \tag{6}$$

where ϕ is a probability density function of the multivariate Gaussian distribution. PFs generate samples - are also called ensemble members - from the distribution for the SHBM (3) at every time step. To obtain an ensemble of N samples from the joint distribution $f_{x_1,x_0,\theta}$, we follow the sampling procedure for random vectors as outlined in Algorithm 1.

Using Algorithm 1, we obtain an ensemble of samples $\{x_{0:1}^{(i)}, \theta^{(i)}\}_{i=1}^N$ from distribution $f_{x_{0:1},\theta}$. Assume an ensemble of samples $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ from distribution $f_{x_{0:t},\theta|z_{1:t-1}}$ is also available. Then at every time step t, $t = 2, 3, \dots$, for any function $G(x_{0:t}, \theta)$, given the observations $z_{1:t}$, its posterior estimate is:

$$\begin{aligned} \hat{G} &= E[G(x_{0:t}, \theta)|z_{1:t}] \\ &\triangleq \sum_{i=1}^{N} w_t^{(i)} \cdot G(x_{0:t}^{(i)}, \theta^{(i)}), \end{aligned} \tag{7}$$

Algorithm 1 Generate the ensemble from $f_{x_1,x_0,\theta}$.

1: First, generate N independent samples $\{x_0^{(i)}, \theta^{(i)}\}_{i=1}^N$ from $\pi_{x_0} \cdot \pi_\theta^0$, where $\theta^{(i)} = (k_i, \xi_{k_i}^{(i)})$. For $i = 1, 2, \ldots, N$, draw k_i from $\rho^0(k)$, then draw $\theta_{k_i}^{(i)}$ from $\pi_{\xi_{k_i}}^0$, thus obtaining the ith sample of θ as $(k_i, \xi_{k_i}^{(i)}) = \theta^{(i)}$.

2: Draw samples $x_1^{(i)}$ of x_1 from $f_{x_1|x_0,\theta}(x_1|x_0 = x_0^{(i)}, \theta = \theta^{(i)}) = \phi(x_1; M(x_0^{(i)}|\theta^{(i)}), \sigma^2 R_n)$. Specifically, draw $\epsilon_1^{(i)} \sim \mathcal{N}_n(0, \sigma^2 R_n)$ so that $x_1^{(i)} = M(x_0^{(i)}|\theta^{(i)}) + \epsilon_1^{(i)}$.

where

$$w_t^{(i)} = \frac{\phi(z_t; h(x_t^{(i)}), \delta^2 R_m)}{\sum_{j=1}^N \phi(z_t; h(x_t^{(j)}), \delta^2 R_m)}. \tag{8}$$

The effective sample size (ESS) is defined as:

$$ESS = \frac{1}{\sum_{i=1}^N w_t^{(i)}}. \tag{9}$$

It serves as a criterion for evaluating the effectiveness of an ensemble of weighted particles (Elvira et al., 2018; Albarakati et al., 2022). A higher ESS indicates that the particle weights are more evenly distributed, suggesting a more representative sample of the target distribution.

Applying the resampling strategy of the bootstrap PF (Gordon et al., 1993), we draw N samples with replacement from the ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ with weights $\{w_t^{(i)}\}_{i=1}^N$ as sampling probabilities. This results in $\{x_{0:t}^{(i)\prime}, \theta^{(i)\prime}\}_{i=1}^N$. For details on the bootstrap filter, refer to Gordon et al. (1993) and Godsill (2019). These N particles are sample realizations from the posterior distribution $f_{x_{1:t},\theta|z_{1:t}}$. For convenience, we relabel the resampled N particles as $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$.

Using conditional probability density sampling, for each $\{x_t^{(i)}, \theta^{(i)})\}$, where $i = 1, 2, \ldots, N$, draw $x_{t+1}^{(i)}$ from the conditional distribution $f_{x_{t+1}|x_{0:t},\theta,z_{1:t-1}} = \phi(x_{t+1}; M(x_t^{(i)}|\theta^{(i)}), \sigma^2 R_n)$. This yields an ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ from the joint distribution $f_{x_{0:t+1},\theta|z_{1:t-1}}$. Thus at time step t, we have an ensemble $\{x_{0:t+1}^{(i)}, \theta^{(i)}\}_{i=1}^N$ from the joint distribution $f_{x_{0:t+1},\theta|z_{1:t-1}}$. The details of the sequential Monte Carlo (SMC) method are provided in Appendix A. However, due to the presence of large weights $w_t^{(i)}$, many duplicate particles may appear after resampling. To mitigate this, we apply Markov chain Monte Carlo (MCMC) methods to move the duplicate particles. The resampling procedure is described in detail in Section 3.3.

3.2. The estimation of the conditional distribution of parameters

The conditional density of θ given $x_{0:t}$ and $z_{1:t}$ $f_{\theta|x_{0:t},z_{1:t}}$ can be expressed recursively

$$f_{\theta|x_{0:t},z_{1:t}} = \frac{\phi(z_t; h(x_t), \delta^2 R_m) \cdot \phi(x_t; M(x_{t-1} \mid \theta), \sigma^2 R_n) \cdot f_{\theta|x_{0:t-1},z_{1:t-1}}}{E_{\theta|x_{0:t-1},z_{1:t-1}}[\phi(z_t; h(x_t), \delta^2 R_m) \cdot \phi(x_t; M(x_{t-1} \mid \theta), \sigma^2 R_n)]}, \tag{10}$$

This recursive expression is difficult to compute in practices, but can be approximated by samples $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ drawn from the distribution $f_{x_{0:t},\theta|z_{1:t}}$ at each time step. The conditional density can also be written as:

$$f_{\theta|x_{0:t},z_{1:t}} = \frac{f_{\theta,x_{0:t}|z_{1:t}}}{f_{x_{0:t}|z_{1:t}}}. \tag{11}$$

However, this form is not easily directly derived from the samples $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$. Nonetheless, it is natural to consider $\{\theta^{(i)}\}_{i=1}^N$ as samples from the marginal posterior $f_{\theta|z_{1:t}}$, and $\{x_{0:t}^{(i)}\}_{i=1}^N$ as samples drawn with equal probability from the marginal distribution of the state trajectory.

Given the observations $z_{1:t}$ and the ensemble $\{x_{0:t}^{(i)}\}_{i=1}^N$, the distribution $f_{x_{0:t}|z_{1:t}}$ can be approximated using importance sampling, based on the expression:

$$f_{x_{0:t}|z_{1:t}} \propto \prod_{t_q=1}^t \phi\left(z_{t_q}; h\left(x_{t_q}\right), \delta^2 R_m\right) \cdot f_{x_{0:t}}. \tag{12}$$

Using formula (11), we treat $\{\theta^{(i)}\}_{i=1}^N$ as an ensemble of samples from $f_{\theta|x_{0:t},z_{1:t}}$, and assign $\theta^{(i)}$ weights $u^{(i)}$ (Doucet et al., 2001; Bishop and Nasrabadi, 2006). Since $\{x_{0:t}^{(i)}\}_{i=1}^N$ are an ensemble of equally probable samples from $f_{x_{0:t}}$, the importance weights are computed as:

$$u^{(i)} = \frac{1/\prod_{t_q=1}^t \phi\left(z_{t_q}; h\left(x_{t_q}^{(i)}\right), \delta^2 R_m\right)}{\sum_{j=1}^N [1/\prod_{t_q=1}^t \phi\left(z_{t_q}; h\left(x_{t_q}^{(j)}\right), \delta^2 R_m\right)]}. \tag{13}$$

Given the ensemble of particles, the posterior probability density function $f_{\theta|x_{0:t},z_{1:t}}$ can be estimated using either parametric or non-parametric methods. The estimation methods and other details are provided in Appendix B.

3.3. The two-step RJMCMC resampling method.

We propose a new PF with a two-step resampling scheme. After the arrival of a new observation and the weights calculation, the first resampling step applies the bootstrap PF method (Evensen et al., 2022). This generates a preliminary ensemble of particles with equal weights, though it may result in multiple instances of the same particle. This approach is a computationally efficient approach and helps mitigate particle degeneracy. To reduce the particle duplication and improve diversity, the second step utilizes the RJMCMC methods to draw new particles from the target distribution. The target distribution defined over $\theta = (k, \xi_k)$, spans multiple models and a variable-dimensional parameter space, which is a challenge not addressed in the previous works.

Given the samples $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ after the bootstrap resampling from $f_{x_{0:t},\theta|z_{1:t}}$, for $t \geq 2$, which has duplicate particles. We denote the ensemble of particles as:

$$\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N} \triangleq \{x_{0:t}^{(i)}, k^{(i)}, \xi_k^{(i)}\}_{i=1}^{N}. \tag{14}$$

To obtain an ensemble of samples that better represents $f_{x_{0:t},\theta|z_{1:t}}$ (Gilks and Berzuini, 2001), the RJMCMC resampling is applied. For the particle $(x_{0:t}^{(i)}, k_i, \xi_k^{(i)})$ required to be resampled, a new particle $(x_{0:t}^{*}, k^{*}, \xi_{k^*}^{*})$ is proposed using a Markov transition function $g_{((x_{0:t},\theta)\to(x_{0:t}^{*},\theta^{*}))}$. The proposed particle is accepted with an acceptance probability $\alpha_{((x_{0:t},\theta)\to(x_{0:t}^{*},\theta^{*}))}$ such that the detailed balance condition is satisfied (Doucet et al., 2000; Elvira et al., 2018).

If $(x_{0:t}^{(i)}, \theta^{(i)})$ is a duplicate particle that requires resampling, note that only $(x_t^{(i)}, \theta^{(i)})$ is used to generate the next state $x_{t+1}^{(i)}$. Therefore, we retain the trajectory $x_{0:t-1}^{(i)}$ and draw a new pair $(x_t^{(i)}, \theta^{(i)})$. At this point, the target distribution becomes the posterior distribution for x_t and θ:

$$\begin{aligned} & f_{x_t,\theta|z_t,x_{0:t-1}=x_{0:t-1}^{(i)}} \\ & \propto \phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1}^{(i)} \mid \theta\right), \sigma^2 R_n\right) \cdot f_{\theta|x_{0:t-1}=x_{0:t-1}^{(i)},z_{1:t-1}} \\ & \approx \phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1}^{(i)} \mid \theta\right), \sigma^2 R_n\right) \cdot \hat{\rho}(k) \cdot \hat{\pi}_{\xi_k}(\xi_k). \end{aligned} \tag{15}$$

The current ith particle is $(x_t^{(i)}, \theta^{(i)}) = (x_t^{(i)}, k_i, \theta_{k_i}^{(i)})$, and the proposed particle is $(x_t^{*}, \theta^{*}) = (x_t^{*}, k^{*}, \xi_{k^*}^{*})$. The sampling scheme can be specified as:

$$\begin{aligned} x_t^{(i)} &= M^{k_i}\left(x_{t-1}^{(i)} \mid \xi_k^{(i)}\right) + \epsilon_t^{(i)}, \\ x_t^{*} &= M^{k^*}\left(x_{t-1}^{*} \mid \xi_{k^*}^{*}\right) + \epsilon_t^{(i)}, \end{aligned} \tag{16}$$

i.e., the same noise term $\epsilon_t^{(i)}$ is used when proposing x_t^{*}, the acceptance ratio is given by:

$$\alpha = \frac{g(\theta^{(i)}|\theta^{*}, x_t^{*}) a_{k^*\to k_i} \phi\left(z_t; h\left(x_t^{*}\right), \delta^2 R_m\right) \cdot \hat{\rho}(k^*) \cdot \hat{\pi}_{\xi_{k^*}}(\xi_{k^*}^{*})}{g(\theta^{*}|\theta^{(i)}, x_t^{(i)}) a_{k_i\to k^*} \phi\left(z_t; h\left(x_t^{(i)}\right), \delta^2 R_m\right) \cdot \hat{\rho}(k_i) \cdot \hat{\pi}_{\xi_k}(\xi_{k_i}^{(i)})} |J|, \tag{17}$$

where the Jacobian determinant of the transformation is:

$$J = \left| \frac{dq(\xi_k, \mu)}{d(\xi_k, \mu)} \right|_{(\xi_k,\mu)=(\xi_{k_i}^{(i)},\mu^{(i)})}. \tag{18}$$

The resampling step for selecting new particles is summarized in Algorithm 2, with further details provided in Appendix C. At time t, for any function $G(x_{0:t}, \theta)$ the state trajectory and parameters, we now have an ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ from the posterior distribution $f_{x_{0:t},\theta|z_{1:t}}$. Therefore, the posterior estimate of $G(x_{0:t}, \theta)$ given the observations $z_{1:t}$ is:

$$\begin{aligned} \hat{G} &= E[G(x_{0:t}, \theta)] \\ &= \int G(x_{0:t}, \theta) \cdot f_{x_{0:t},\theta|z_{1:t}} \, dx_{0:t} \, d\theta \\ &= E_{x_{0:t},\theta|z_{1:t}}[G(x_{0:t}, \theta)] \\ &\approx \frac{1}{N} \sum_{i=1}^{N} G(x_{0:t}^{(i)}, \theta^{(i)}). \end{aligned} \tag{19}$$

The Algorithm 3 concludes our proposed PF with RJMCMC resampling method.

4. Case study: One-dimensional advection equation

4.1. Problem background

The advection equation describes the local change of a physical quantity due solely to the advection of that quantity. The simplest form is the one-dimensional advection equation:

$$\frac{\partial x}{\partial t} = -v \frac{\partial x}{\partial s}. \tag{21}$$

Algorithm 2 The second resampling step with RJMCMC method.

1: **Initialization:** An ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ after the first resampling step.
2: For each particle i need to be resampled, the current particle is $(x_t^{(i)}, \theta^{(i)}) = (x_t^{(i)}, k_i, \theta_{k_i}^{(i)})$.
3: Propose a consistent transition from the k_ith model to the k^*th model by drawing $\mu^{(i)}$ from $a_{k_i \to k^*}$, and construct a corresponding μ^* such that

$$(\xi_{k^*}^*, \mu^*) = q(\xi_{k_i}^{(i)}, \mu^{(i)}), \tag{20}$$

where q is a given bijective mapping from the space of $(\xi_{k_i}^{(i)}, \mu^{(i)})$ to the space of $(\xi_{k^*}^*, \mu^*)$.
4: Compute the acceptance ratio by Eq. (17).
5: Accept or reject the proposed particle based on the acceptance ratio α.

- If the proposal is rejected, retain $(x_t^{(i)}, k_i, \xi_{k_i}^{(i)})$.
- If the proposal is accepted, replace $(x_t^{(i)}, k_i, \xi_{k_i}^{(i)})$ with $(x_t^*, k^*, \xi_{k^*}^*)$ and reassign $(x_t^{(i)}, k_i, \xi_{k_i}^{(i)})$.

6: $(x_t^{(i)}, k_i, \xi_{k_i}^{(i)}) \triangleq (x_t^{(i)}, \theta^{(i)})$ along with the given $x_{0:t-1}$ serves as a sample from $f_{x_{0:t},\theta|z_{1:t}}$.
7: A new ensemble of particles $\{x_{0:t}^{(i)}, k_i, \xi_{k_i}^{(i)}\}_{i=1}^{N} \triangleq \{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ is obtained as samples from $f_{x_{0:t},\theta|z_{1:t}}$.

Algorithm 3 The PF with RJMCMC resampling method.

Initialization:
Give the prior distribution π_{x_0} and $\pi_\theta^0 = \rho^0(k) \cdot \pi_{\xi_k}^0$.
samples $\{x_{0:1}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ from $f_{x_1,x_0,\theta}$ by Algorithm 1.
For time step t, $t = 1, 2, \ldots$:
    An ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ from $f_{x_1,x_0,\theta}$ (when $t = 1$) or $f_{x_{0:t},\theta|z_{1:t-1}}$ (when $t \geq 2$) and new observation z_t are available.
    Calculate $w_t^{(i)}$ by Eqs. (A.9) or (A.17).
    Resample $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ with probabilities $\{w_t^{(i)}\}_{i=1}^{N}$ and relabel as $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$, then they are an ensemble of samples from $f_{x_{0:t},\theta|z_{1:t}}$.
    For the particles in $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ which need to be resampled,
        Propose a proposal particle $(x_{0:t-1}^{(i)}, x_t^*, \theta^*)$ and accept/reject it by Algorithm 2.
    End For
    The output is $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$, which is an ensemble of samples from $f_{x_{0:t},\theta|z_{1:t}}$. For any function $G(x_{0:t}, \theta)$ of $x_{0:t}, \theta$, given the observations $z_{1:t}$, its posterior estimate is given by Eq. (19), and the corresponding estimate of $f_{\theta|x_{0:t},z_{1:t}}$ is $\hat{\rho}(k) \cdot \hat{\pi}_{\xi_k}(\xi_k)$.
    $t = t + 1$.
End For

Here, the state vector x represents the value of a physical quantity at different spatial coordinates, such as the concentration of a substance; v is the system velocity; s is the spatial coordinate in the horizontal direction; and t is time. This equation indicates that the local change in the physical quantity is due to the uneven horizontal distribution of the quantity and the movement of the system. If v is constant, meaning the system moves with a constant velocity along the s direction, the shape of the wave remains unchanged. If s changes with time t or varies along the s direction, then the equation becomes nonlinear.

Due to terrain or other factors, the velocity v often changes with the position s. To model spatially varying velocity, we define v_s as a piecewise function, resulting in a series of interval points at the locations of interest. The velocity in the ith segment is v_i, where i is the segment index, $i = 1, 2, \ldots, k+1$. The number of discrete points k is also an unknown parameter, but its range is known: $k \in \{1, 2, \ldots, K_{\max}\}$.

Considering the Eq. (21) on a set of discrete points within the spatial interval $[0, L]$ denoted as $\{s_0, s_1, s_2, \ldots, s_n\}$, where $s_1 = 0$ and $s_n = L$. The number n corresponds to the dimension of the state vector x. Let $\{c_1, c_2, \ldots, c_k\}$ represent the critical points that separate different velocity segments.

The unknown parameters in Eq. (21) are:

$$\theta = (k, \xi_k) = (k, v_1, v_2, \ldots, v_{k+1}, c_1, c_2, \ldots, c_k). \tag{22}$$

The parameter space is $\Theta = \cup_{k \in \mathcal{K}}(\{k\} \times \Theta_k)$, for each k, $\Theta_k \subset \mathbb{R}^{2k+1}$. Denote the model M^k as the advection equation with k discontinuities. The parameters for model k are $\xi_k = (v_1, v_2, \ldots, v_{k+1}, c_1, c_2, \ldots, c_k)$.

The prior of the true model under the kth model, i.e., the probability of having k discontinuities, follows a truncated Poisson distribution:

$$\rho(k) = \frac{e^{-\lambda}\frac{\lambda^k}{k!}}{\sum_{i=1}^{K_{max}} e^{-\lambda}\frac{\lambda^i}{i!}}, \quad k = 1, 2, \ldots, K_{max}. \tag{23}$$

where λ and $K_{\max}$ are given. The velocity parameters $v_1, v_2, \ldots, v_{k+1}$ follow a $\Gamma(\alpha, \beta)$ distribution with the density function $\beta^\alpha v_j^{\alpha-1} e^{-\beta v_j}/\Gamma(\alpha)$, for $j = 1, 2, \ldots, k+1$, where α and β are given. The prior distribution of the parameters is

$$\pi_\theta(k, \xi_k) = \rho(k) \cdot \pi_{\xi_k}(\xi_k) \propto e^{-\lambda} \frac{\lambda^k}{k!} \cdot \prod_{j=1}^{k+1} [\beta^\alpha v_j^{\alpha-1} e^{-\beta v_j}/\Gamma(\alpha)] \cdot f_c, \tag{24}$$

where f_c denotes the prior distribution of location parameter. The proposal is implemented through four types of transitions: (a) adding a discontinuity point, (b) removing a discontinuity point, (c) changing one of the velocity, and (d) changing the position of one discontinuity point. A hyper-parameter c controls the proposal density of model type through Eqs. (D.6) and (D.7). Complete specifications of the prior distributions and proposal mechanism are provided in Appendix D.

4.2. The simulation settings

Consider the one-dimensional advection equation on 401 integer points within the interval $[0, 400]$. The locations $s \in \{0, 1, 2, \ldots, 400\}$. The dimension of the state vector is $n = 401$. We set the initial state vector as

$$x_0(s) = \frac{1}{5} \sin\left(\frac{3\pi}{20} s\right) s\left(\frac{2}{3} - s/400\right) \exp\left(-\frac{1}{200} s\right). \tag{25}$$

Assume that observations can only be acquired at 40 points, meaning the observation dimension is approximately one-tenth of the state vector dimension. In the simulation, these points are randomly selected. Each observation is modeled as the true plus Gaussian noise with mean 0 and variance $\sigma^2 = 0.2$. Observations are collected every 10 time steps and used in the assimilation process. The total number of simulation time steps is 600. The ensemble of initial state vectors is generated as:

$$x_0^{(i)} = x_0 \cdot (1 + \epsilon_0^{(i)}),$$

where $\epsilon_0^{(i)} \sim N(0, 1)$. There are 2 discrete spatial points at $s = 100$ and $s = 250$. Velocities in these segments are 0.7, 0.2, and 0.4, respectively.

4.3. Investigating the impact of priors

It is well-known that RJMCMC methods are particularly sensitive to prior choices (Barker and Link, 2013). We first examine the prior sensitivity of the proposed algorithm through simulation.

For the prior of the model parameters, we take a truncated Poisson distribution with strength $\lambda = 2$ and the maximum number of discrete spatial points $K_{\max} = 3$. The velocity parameter in each segment follows a $\Gamma(\alpha, \beta)$ prior distribution, where $\alpha = 0.4$. To sample the ensemble of parameters, we first draw the model index k from the truncated Poison distribution, and then draw the corresponding parameters. At each DA step, the particles in the ensemble are resampled using Algorithm 3.

To systematically examine the impact of the prior parameter β on the results, we first perform simulations with β ranging from 0.55 to 1.55 in increments of 0.2, conducting 30 replicate runs for each value.

We use the following Mean Squared Error (MSE) at the 600th time step as an indicator of the assimilation error:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} \left(z_{t,i} - x_{t,i}\right)^2, \tag{26}$$

where $z_{t,i}$ are the observations at the ith observation location, and m is the total number of observations. To evaluate the forecasting performance, we also compute the Mean Squared Prediction Error (MSPE) using predicted states 50 time steps ahead:

$$\text{MSPE} = \frac{1}{m} \sum_{i=1}^{m} \left(z_{t,i} - x_{t,i}^f\right)^2, \tag{27}$$

where $x_{t,i}^f$ denotes the foretasted state at the ith location.

The simulation results are presented in Table 1 and illustrated in the left panel of Fig. 1. The results indicate that the mean values of MSE and MSPE exhibit minimal sensitivity to variations in β. This stability is attributed to our method's implementation of RJMCMC in each resampling step: as the algorithm sequentially assimilates more observational data over time, the influence of prior information gradually diminishes. Next, we fix $\beta = 0.95$ along with other parameters and vary λ across values 1, 2, and 3. Based on the

Table 1
The average MSE and MSPE over 30 simulations across different β values. Results demonstrate that the proposed method shows relatively low sensitivity to prior parameter β.

β	0.55	0.75	0.95	1.15	1.35	1.55
MSE	0.481901	0.637324	0.528046	0.385506	0.548772	0.464786
MSPE	0.970376	1.414456	0.847632	0.963518	1.274058	1.116736

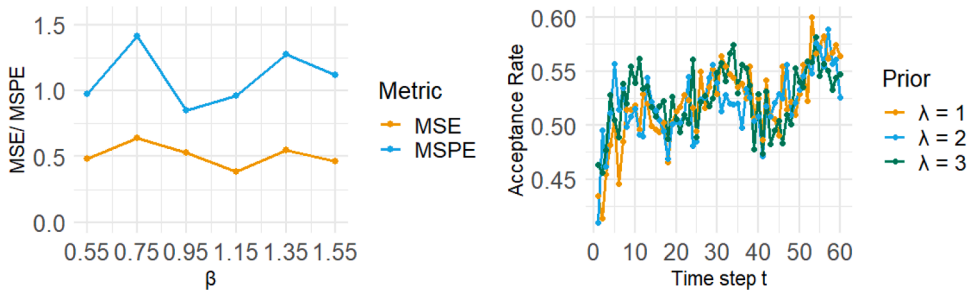

Fig. 1. The effect of the prior on the proposed method. The left subfigure shows the average MSE and MSPE over 30 simulations as β varies. The right subfigure displays the average acceptance rate at each time point over 30 simulations under different λ values. The results indicate that the prior has a limited influence on the outcomes of our method.

Table 2
The average MSE and MSPE over 30 simulations across different λ values.

λ	1	2	3
MSE	0.672550	0.449901	0.632588
MSPE	1.253647	0.925636	1.574184

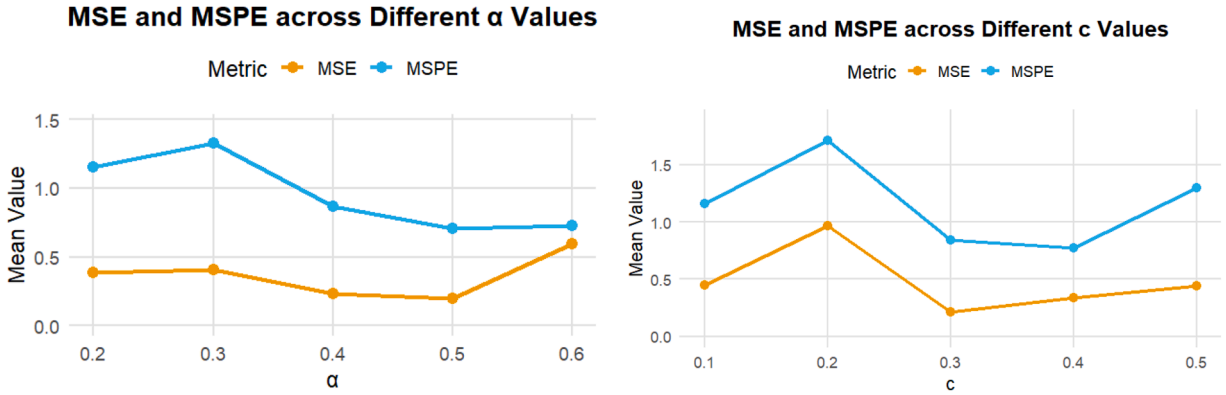

Fig. 2. The average MSE and MSPE over 30 simulations as α (left) and c (right) varies.

Table 3
The average MSE and MSPE over 30 simulations across different α and c values. Results demonstrate that the proposed method shows relatively low sensitivity to prior parameter β.

α	0.2	0.3	0.4	0.5	0.6
MSE	0.3858896	0.4064508	0.2306556	0.1938879	0.5921764
MSPE	1.1510757	1.3306954	0.8654554	0.7029578	0.7251362
c	0.1	0.2	0.3	0.4	0.5
MSE	0.4506588	0.9674379	0.2110012	0.3375425	0.4396401
MSPE	1.1570758	1.7142358	0.8448592	0.7692091	1.2960018

averaged results from 30 simulations (summarized in Table 2), $\lambda = 2$ yields relatively smaller errors. The right panel of Fig. 1 illustrates the variation in acceptance rates over time under different λ values, indicating that λ has only a minor effect on the acceptance rates.

In addition, we also conduct simulations with α ranging from 0.2 to 0.6, and c from 0.1 to 0.3, both in increments of 0.1, conducting 20 replicate runs for each value. The simulation results are presented in Fig. 2 and Table 3, which also indicate the low sensitivity to variations in α and c. The selection of c is relatively sensitive for both MSEs and MSPEs, because c is the hyper-parameter that controls the model proposal probability. A reasonable c can help with model identification. Therefore, caution should be exercised when selecting c. These findings suggest that prior information on model types (e.g., number of segments) requires more careful consideration, whereas the priors on model parameters have a relatively weaker influence on the outcomes. Therefore weakly informative prior distributions with $\lambda = 2$, $\alpha = 0.4$, $\beta = 0.95$, and $c = 0.3$ are used utilized in the following simulations.

Table 4
The average MSEs and MSPEs over 20 simulations for three different bandwidth choices when ensemble size $N = 60$ and $N = 100$.

Ensemble Size		Original bandwidths	0.1 times bandwidths	10 times bandwidths
$N = 60$	MSE ($t = 600$)	**0.5106497**	0.5127474	0.6047179
	MSPE ($t = 650$)	1.2501598	0.9938791	**0.9683806**
$N = 100$	MSE ($t = 600$)	**0.1813500**	0.2157441	0.2054888
	MSPE ($t = 650$)	**0.6733092**	0.9300546	0.7207709

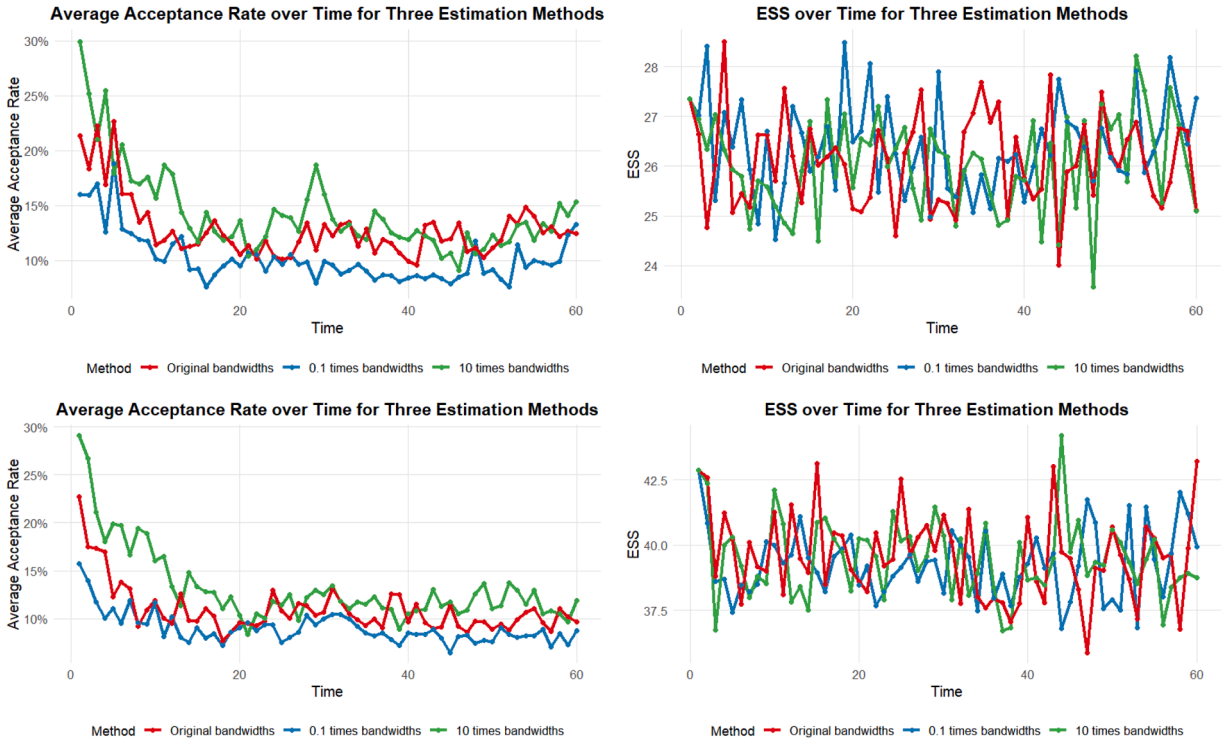

Fig. 3. The two upper subplots display the average acceptance rates and ESSs under different bandwidth selections when the ensemble size $N = 60$. Red, blue, and green lines represent the original bandwidth, 0.1 times the bandwidth, and 10 times the bandwidth, respectively. The two lower subplots show the corresponding results for $N = 100$. Larger bandwidths lead to higher acceptance rates. However, due to the resampling step, the choice of bandwidth has almost no effect on ESSs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4. Comparison between different parameter estimation methods

Both parametric and nonparametric methods introduced in Appendix B can be used to estimate parameter's prior distributions at each time step, which are necessary for computing the acceptance probability in Eq. (17).

When employing the nonparametric estimation method, the selection of a bandwidth requires attention. We initially conduct 20 repeated simulations under the default bandwidth, 0.1 times the bandwidth, and 10 times the bandwidth with a set size of $N = 60$ and $N = 100$. The average MSEs and MSPEs for the three scenarios are listed in Table 4. The variation of average acceptance probability over time under the three scenarios is presented in Fig. 3. The results indicate that although increasing the bandwidth improves the average acceptance rate, it does not necessarily lead to a reduction in MSE and MSPE. The choice of bandwidth depends on several factors, most notably the ensemble size, as well as observation errors, among others. More studies on determining the optimal bandwidth are needed in future research.

We also compare the two estimation methods with setting the informative prior distributions. The average acceptance rates, ESSs, MSEs, and MSPEs from 20 independent simulations with ensemble size $N = 60$ are presented in Table 5 and Fig. 4. The results reflect that due to the resampling process and the gradual integration of information from new observations over time, the three methods show no significant differences in ESSs, MSEs, and MSPEs.

The left subplot of Fig. 4 shows that the average acceptance rate curve for the nonparametric estimation method is decreasing, while the curves for the other two methods show an increasing trend in the initial time steps.

Table 5
The average MSEs and MSPEs over 20 simulations for three prior distribution estimation methods.

	Preset Prior	Nonparametric estimation	Parametric estimation
MSE ($t = 600$)	0.6479366	**0.3487666**	0.7887746
MSPE ($t = 650$)	0.7515748	1.4288716	**0.5269815**

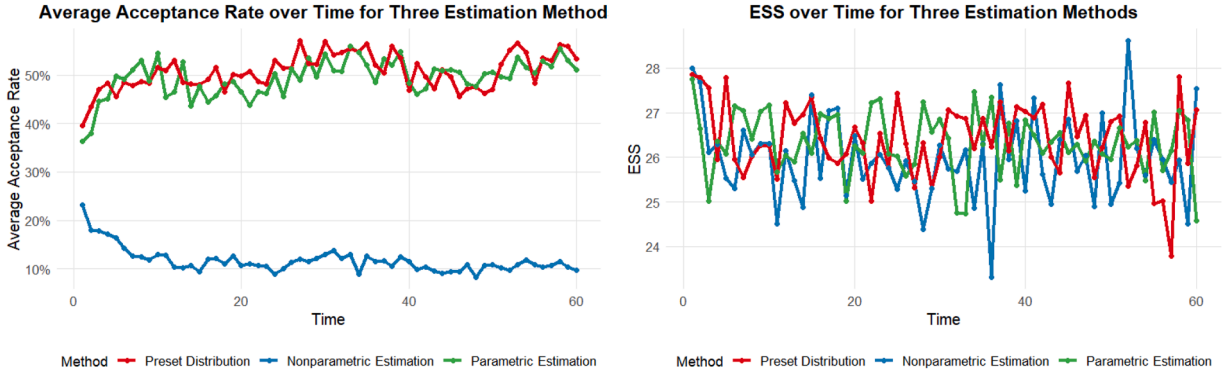

Fig. 4. The acceptance ratios and ESS of three estimation methods. The average acceptance rate curve for the nonparametric estimation method (blue) is decreasing, while the curves for the parametric method (green) and preset informative prior (red) show an increasing trend in the initial time steps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

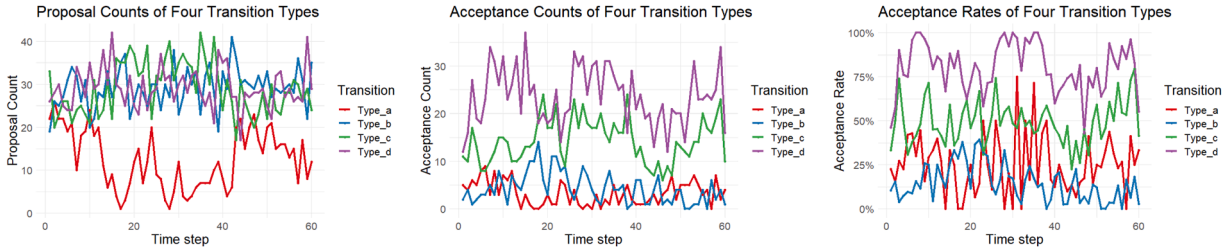

Fig. 5. Proposal counts, acceptance counts, and acceptance rates for the four types of proposals (from left to right).

The nonparametric estimation method uses a set of kernel functions centered on the samples to estimate the parameter distribution. As the resampled particles gradually approach the true values over time, the estimated distribution becomes very sharp, particularly with some model probabilities dropping to zero. This results in proposed parameters being accepted only when they are near the peak of the estimated distribution. However, once a parameter proposal is accepted, the corresponding particle tends to be very close to the true value. In contrast, the other two methods start with more dispersed particles in the initial ensemble. As these particles converge toward the true values, the proposal distribution becomes relatively closer to the target distribution, leading to the initial upward trend in acceptance rates. This trend eventually stabilizes within a certain range, indicating that the parameter estimation is gradually converging.

Weakly informative prior distributions are used in the following simulations. We ran one DA process with an ensemble size $N = 100$ to demonstrate the acceptance of the four RJMCMC resampling proposals. The three subplots in Fig. 5, from left to right, show the proposal counts, acceptance counts, and acceptance rates of the four proposals. The figure illustrates that the acceptance rates of proposal types (a) and (b) are relatively lower compared to the other two types, which also ensures that particles can more accurately search for the true parameter values under the current model Fig. 5.

4.5. Comparison with the SMC and pMCMC methods

We compare our method with the pMCMC methods (Moradkhani et al., 2012; Vrugt et al., 2013), which address similar application scenarios. In pMCMC methods, the resampling step is also divided into two stages: first, particles are resampled using SIR, and then a proposal is generated for the resampled particles, followed by an accept/reject decision. Compared to Vrugt et al. (2013), the method by Moradkhani et al. (2012) places greater emphasis on maintaining consistency between parameters and states during proposal generation. A primary distinction between the two methods lies in the different MCMC techniques used to generate proposals. A review of the follow-up studies to these two articles reveals that subsequent work has not introduced significant changes to the core theoretical framework. Therefore, we focus our comparison on the method proposed by Moradkhani et al. (2012). We also include a

Table 6
The average MSEs and MSPEs of 30 simulations for the three methods with a varying ensemble size N.

	Ensemble Size	Proposed method	pMCMC (k = 1)	pMCMC (k = 2)	pMCMC (k = 3)	SMC (k = 2)
MSE ($t = 600$)	20	**0.4952799**	0.8646048	0.9798698	1.2214387	2.747683
	40	**0.2582725**	2.2935382	0.3983601	0.5978624	3.174342
	60	**0.2299756**	0.3906565	0.3605533	0.7396862	3.000455
	80	**0.2051728**	0.8072615	0.2835920	0.6896150	3.061852
	100	0.2497857	0.2830056	**0.2056855**	0.4091987	2.884548
	120	**0.2001557**	0.3826623	0.2677825	0.5180791	2.806456
MSPE ($t = 650$)	20	**1.2250328**	1.6850393	2.1043719	1.8136866	1.848615
	40	**0.4443996**	2.9192545	1.8131713	1.1547955	2.007069
	60	**0.5458624**	2.0246323	1.5278236	1.4413560	1.969988
	80	**0.4311186**	2.0697486	1.6385844	1.9707802	1.993398
	100	**0.3342038**	1.8474003	1.4226972	0.9666893	1.899599
	120	**0.2762879**	1.9304805	1.2572423	1.3673224	1.984324

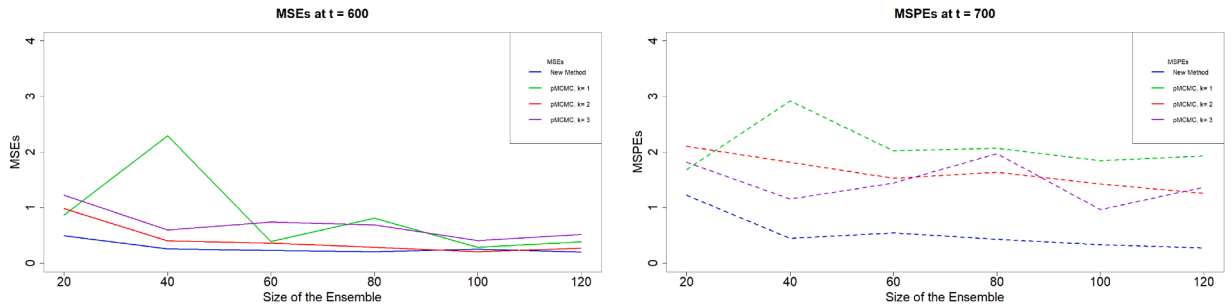

Fig. 6. Variation of MSEs at $t = 600$ (left) and MSPEs at $t = 650$ (right) with respect to the ensemble sizes. Blue, green, red, and purple represent our proposed method, pMCMC with $k = 1$, pMCMC with $k = 2$, and pMCMC with $k = 3$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Average runtime (in seconds) for different methods under varying ensemble sizes.

Ensemble Size	Proposed method	pMCMC (k = 1)	pMCMC (k = 2)	pMCMC (k = 3)	SMC (k = 2)
40	507.756	820.624	931.988	920.902	339.614
60	567.380	733.970	945.788	1189.910	423.112
80	1162.936	2107.070	2415.552	2551.018	971.342

comparison with the SMC method as a benchmark (Vetra-Carvalho et al., 2018). Since the pMCMC and the SMC are not designed to handle the issue of variable-dimensional parameter spaces, we fix the number of change points k to specific values for each method: $k = 1, 2, 3$ for pMCMC, and $k = 2$ for SMC. This avoids requiring these methods to search across multiple parameter spaces ensuring a fair comparison.

4.6. The results and discussion

We studied the variation in MSEs and MSPEs across three methods with respect to the ensemble size. For each configuration, we conducted 20 independent simulations, calculated the average MSEs and MSPEs values. The results are summarized in Table 6. Since the SMC method does not include MCMC or RJMCMC-based particle movement after resampling, it exhibits higher MSE and MSPE values under identical settings. For clarity, only the results of our proposed method and the pMCMC methods are illustrated in Fig. 6, which shows that our method performs relatively well.

Additionally, we further investigated the computational cost of these methods for ensemble sizes of 40, 60, and 80. For each ensemble size, we recorded the average computation time across 10 independent program runs. These results are summarized in Table 7, and visualized in Fig. 7. The findings demonstrate that our proposed method does not significantly increase computational burden compared to basic SMC, while achieving shorter computation time than the pMCMC method.

The SMC method demonstrates lower computational costs primarily because it avoids recomputing evolution equations after resampling to generate new particles. While our proposed method requires Jacobian determinant calculations during cross-dimensional proposals, it avoids proposing parameters across all dimensions, which contributes to its greater efficiency in parameter proposal.

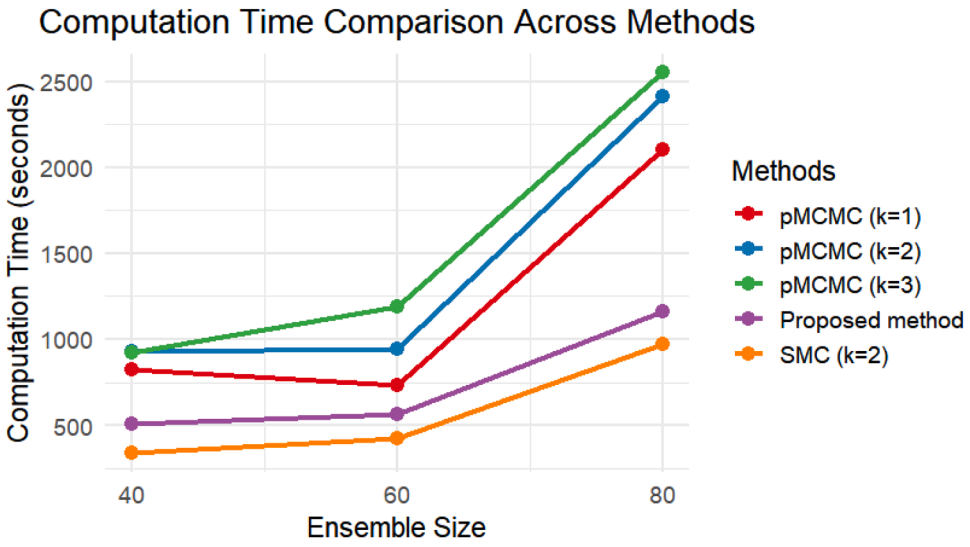

Fig. 7. Average single-run computation time (in seconds) versus ensemble size across different methods. The results show that our proposed method maintains a comparable computational burden to the basic SMC, while achieving shorter computation time than the pMCMC method.

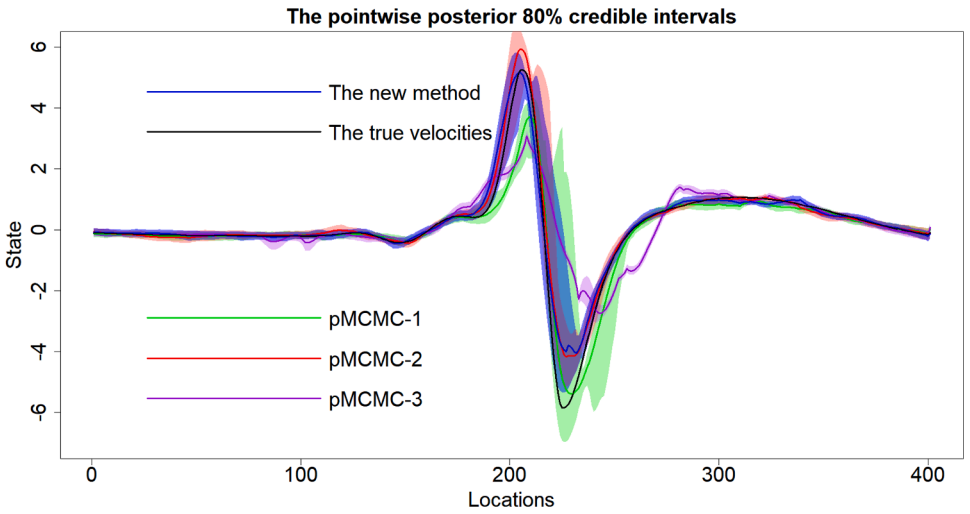

Fig. 8. The pointwise posterior 80 % credible intervals. Blue, green, red, and purple represent the credible intervals and means obtained by our proposed method, the pMCMC method with $k = 1$, the pMCMC method with $k = 2$, and the pMCMC method with $k = 3$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The dominant computational burden in the entire DA process stems from the computation of evolution equation. The RJMCMC component incurs significantly lower costs compared to recomputing evolution equations, meaning that the additional expense introduced by our method is mainly due to repeated evolution equation calculations.

As evidenced by the error metrics in Table 6 and computation time comparisons in Table 7, both our proposed method and pMCMC increase computational demands, but achieve substantial error reduction that justifies the additional cost. For the remaining experiments, we focus on comparing the proposed method with the pMCMC approach.

To more clearly demonstrate the effectiveness of the proposed method, we present the posterior mean of the state and the pointwise posterior 80 % credible intervals at time $t = 600$ from a single simulation. Results from the proposed method, and the pMCMC method are shown in Fig. 8. The corresponding MSEs for four runs are 0.1481, 0.2232, 0.1336, and 1.1317, respectively. The pointwise estimates of the velocities obtained by each method are displayed in Fig. 9. In the figure, the black line segments represent the true velocity values, which exhibit discontinuities at the 100th and 250th locations. The blue line shows the estimated velocities (the posterior mean) by the proposed method, and the blue shaded band indicates the corresponding pointwise posterior 80% credible interval. The figure also displays the posterior means (solid lines) and pointwise 80% posterior credible intervals (shaded bands) obtained using the pMCMC method with fixed $k = 1$ (green), $k = 2$ (red), and $k = 3$ (purple).

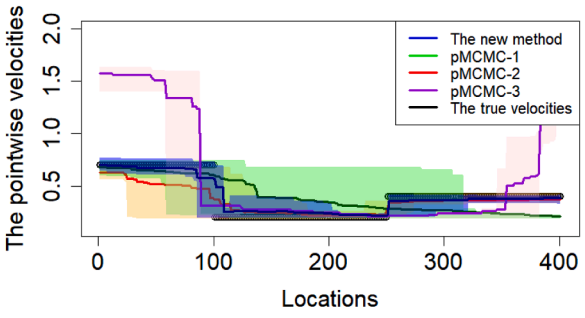

Fig. 9. Estimated velocities. The true pointwise velocity is shown by the black line, alongside the posterior means (solid lines) and pointwise posterior 80% credible intervals (shaded bands). These intervals represent the regions between the upper 0.1 quantile and lower 0.1 quantile of the N particles, as obtained using various methods. Notably, fixing $k = 1$ (green line) or $k = 3$ (purple line) causes significant deviations in the velocity estimates. In contrast, the proposed approach yields the most accurate results, closely matching the true velocity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

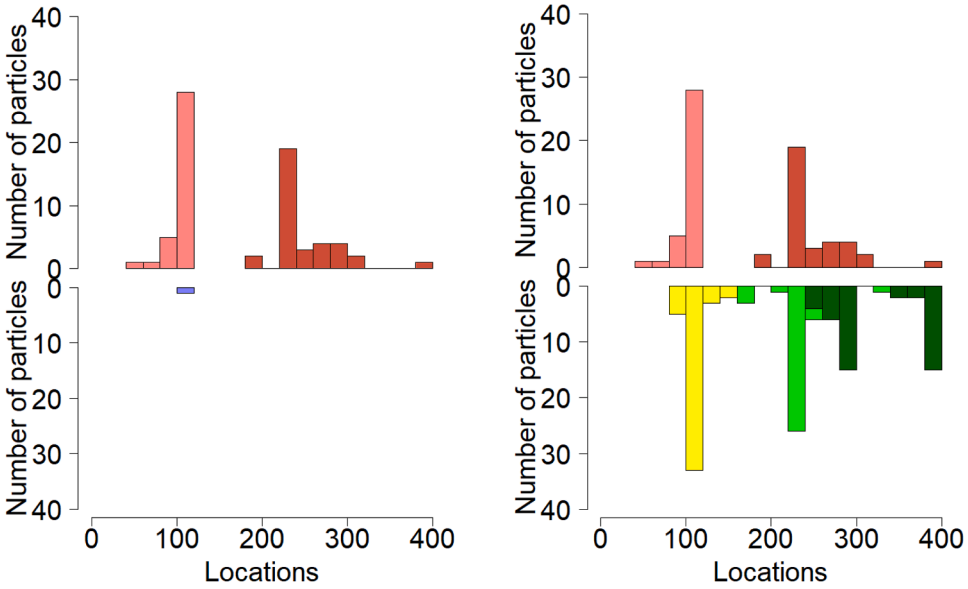

Fig. 10. The results of our method for model identification and location parameter estimation with $N = 80$. The upper half of both subplots shows the bar plots of the position parameters corresponding to the particles with $k = 2$. The light red and red bars represent the positions of the first and second discrete spatial points, respectively. In the left subplot, the blue bar plot represents the position parameters corresponding to the particles with $k = 1$. The bottom-right subplot shows the bar plot of the position parameters for particles with $k = 3$, where yellow, light green, and dark green bars correspond to the positions of the three discrete spatial points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The results demonstrate that both the proposed method and pMCMC ($k = 2$) achieve reasonable velocity estimates. In contrast, the pMCMC with $k = 1$ or $k = 3$ produces significantly poorer estimates. The case with $k = 1$ is particularly problematic - its constant-velocity assumption causes dramatic estimation errors on the left portion of the domain, as clearly shown by the wide divergence between the faded purple confidence band and the true values. Notably, the proposed method yields more accurate estimates than even the pMCMC approach with $k = 2$. This improved performance is attributed to the novel particle proposal mechanism employed in our method, which jointly updates both velocity and position parameters. This integrated approach enhances estimation precision across all dimensions.

The results of our method for model identification and location parameter estimation are shown in Fig. 10. The figure shows that nearly all particles identify models with either $k = 2$ or $k = 3$. Although some identification ambiguity remains - approximately half of the particles recognizing $k = 3$ - the identified models closely approximate the true model. As illustrated in the plot, the third discrete spatial point (shown in dark green) tends to be distributed either near the true second discrete spatial position or toward the right endpoint. This spatial distribution effectively diminishes the practical impact of the third point. The left subfigure of Fig. 11 displays the temporal evolution of particle counts identifying model indices, revealing that the majority of particles recognize models with either $k = 2$ or $k = 3$. The right subfigure presents the time-varying counts of particles detecting two discrete spatial points,

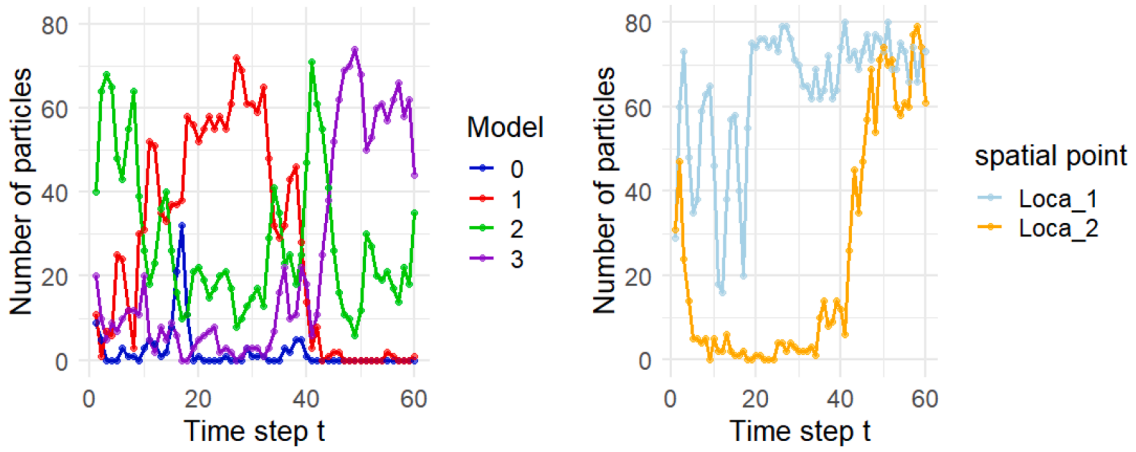

Fig. 11. Particle identification performance for model indices and discrete spatial point detection with ensemble size $N = 80$. The left subfigure displays the temporal evolution of particle counts identifying model indices, showing that the majority of particles recognize models with either $k = 2$ or $k = 3$. The right subfigure presents the time-varying counts of particles detecting two discrete spatial points, demonstrating that nearly all particles successfully identify both discrete spatial points as time progresses. To quantify this, the number of particles falling within the intervals [75,125] and [225,275] are respectively used to represent the number of particles identifying the two breakpoints.

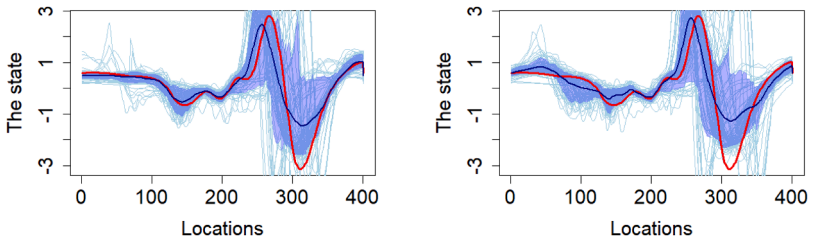

Fig. 12. State predictions at time $t = 650$ using two methods with ensemble size $N = 100$ based on observations from the time interval $[0, 600]$. The left subplot shows the predicted mean and uncertainty intervals produced by our method, while the right subplot shows the results from the pMCMC method with $k = 2$. The dark blue line represents the predicted mean, the blue shaded area represents the pointwise 80 % confidence interval, and the light blue lines correspond to individual particle predictions. The red line represents the true values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

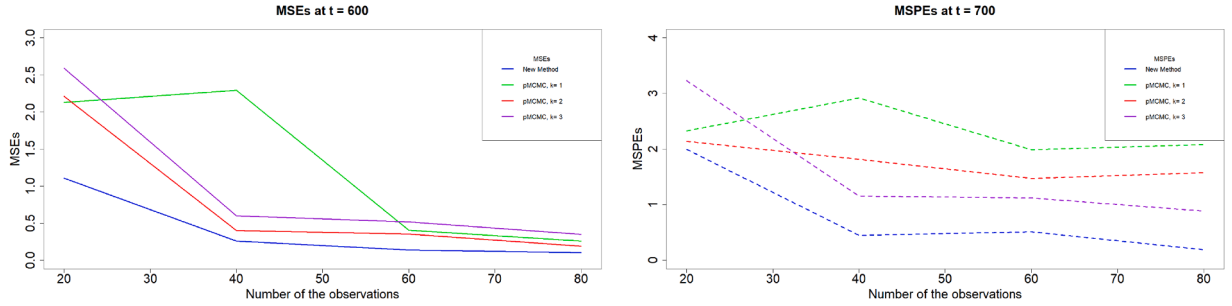

Fig. 13. The variation of MSEs at $t = 600$ (left) and MSPEs at $t = 650$ (right) as a function of the number of observations. Blue, green, red, and purple lines represent our proposed method, pMCMC with $k = 1$, pMCMC with $k = 2$, and pMCMC with $k = 3$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

indicating that nearly all particles successfully identify both discrete spatial points as time progresses Fig. 12. The prediction results of the proposed method and the pMCMC method with $k = 2$ at the 650-th time step are presented in Fig. 12. The results demonstrate that the proposed method achieves predictions closer to the true values.

Next, we investigated how the MSE ($t = 600$) and MSPE ($t = 650$) vary with the number of observations m, comparing the proposed method with the pMCMC approach. The averages of the results from 30 independent simulations are shown in Fig. 13 and summarized in Table 8. These findings show that the proposed method consistently achieves significantly lower MSEs and MSPEs across different observed datasets, outperforming pMCMC in terms of estimation and prediction accuracy.

Table 8
Average results from 30 simulations reporting the MSEs and MSPEs for four methods, evaluated across varying numbers of observations m.

	m	Proposed Method	pMCMC **k = 1**	pMCMC **k = 2**	pMCMC **k = 3**
MSE ($t = 600$)	20	**1.1082071**	2.1316883	2.2123202	2.5927941
	40	**0.2582725**	2.2935382	0.3983601	0.5978624
	60	**0.1383734**	0.4028563	0.3532119	0.5166435
	80	**0.1035852**	0.2596242	0.1869100	0.3470573
MSPE ($t = 650$)	20	**1.9961356**	2.3305962	2.1398791	3.2322982
	40	**0.4443996**	2.9192545	1.8131713	1.1547955
	60	**0.5079265**	1.9862264	1.4702752	1.1174952
	80	**0.1872246**	2.0801137	1.5717806	0.8841347

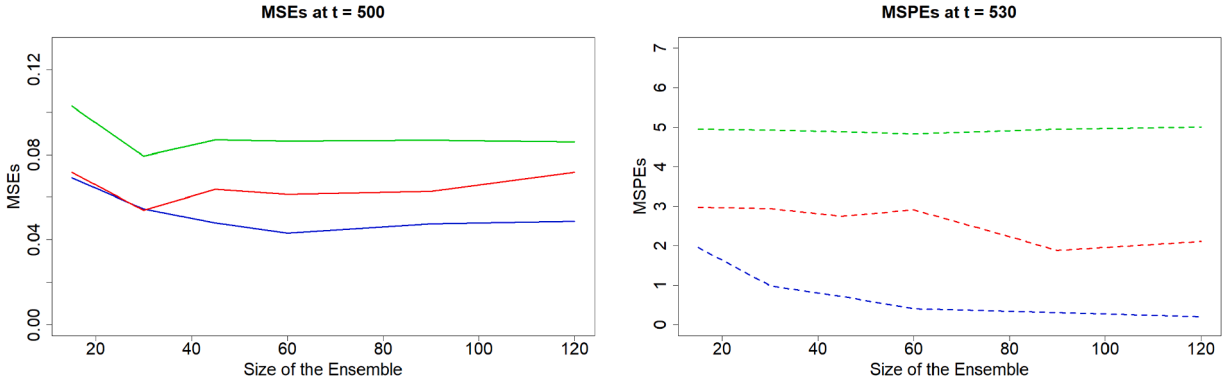

Fig. 14. The figure shows the variation of MSEs of the Lorenz 96 model at $t = 500$ (left) and MSPEs at $t = 530$ (right) with respect to the number of observations. In the figure, blue, green, and red represent our proposed method, pMCMC with $k = 0$, and pMCMC with $k = 1$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Nonlinear evolution case: Lorenz 96 model

In the next experiment, we apply our method to the extensively used medium-dimensional dynamical system known as the Lorenz'96 model (L96) (Lorenz, 1996). This model, which is developed by Edward Lorenz, represents a nonlinear chaotic dynamic system governed by the equation:

$$\frac{dx_j}{dt} = -x_{j-1}(x_{j-2} - x_{j+1}) - x_j + F. \tag{28}$$

where $x = \{x_j; j = 1, ..., n\}$ is the $n-$dimensional state vector. Previous studies, such as Herrera et al. (2010), Van Kekem and Sterk (2018) have considered more complex variants of the L96 model by modifying the forcing terms. Here adopt the following modified form:

$$\frac{dx_j}{dt} = -x_{j-1}(x_{j-2} - x_{j+1}) - x_j + F_j. \tag{29}$$

where $F = \{F_j; j = 1, ..., n\}$ is the corresponding $n-$dimensional force parameter assumed to be piecewise constant. We consider a 40-dimensional state vector $x = (x_1, \dots, x_{40})^T$, where x_1 to x_{20} correspond to forcing $F = 1$, and x_{21} to x_{40} correspond to forcing $F = 6$. Thus, the true parameter vector is $p_{\text{true}} = (1, 20, 1, 6)^T$. The true solution $(x_{1k}, x_{2k}, \dots, x_{nk})^T$, where $n = 40$, is computed over the time interval $t \in [0, 30]$, with a time step $\Delta t = 0.03$, resulting in a total of 1001 time steps. The initial condition is set as $x_0 = (x_{1,0}, x_{2,0}, \dots, x_{n,0})^T = (1.01, 1, \dots, 1)^T$.

For the prior of the model parameters, the model index k follows a truncated Poisson distribution with strength $\lambda = 1$, and the upper bound $K_{\max} = 2$. The forcing term F on each segment follows a $\Gamma(\alpha, \beta)$ prior distribution, where $\alpha = 1.5$ and $\beta = 0.5$.

We compared our method with the pMCMC methods (with $k = 0$ and $k = 1$). The average MSEs and MSPEs from 30 independent simulations and across ensemble sizes $N = 15$, 30,45,60, 90, and 120, are shown in Fig. 14. The results clearly demonstrate that our proposed method consistently outperforms the pMCMC approaches.

6. Discussion and conclusions

In this paper, we proposed the Sequential Hierarchical Bayesian Model (SHBM), which integrates a State Space Model (SSM) to address the data assimilation (DA) problems with uncertainties in both the form of the equations and their parameters. At the initial stage of DA, employing a general model like SHBM helps prevent particle filter (PF) divergence caused by incorrect modeling

assumptions. We introduce a two-stage resampling PF method incorporating the RJMCMC method. This approach enables dynamic identification of the model's form, including the dimensionality of the unknown parameters, during the DA process.

We provided both theoretical derivations and practical applications to demonstrate the effectiveness of our method. By resampling across multiple models and different parameter space of varying dimensions, our approach effectively overcomes the common issue of particle diversity deficiency in PF and meets the resampling requirements in parameter spaces of different dimensions. The proposed method allows for sequential model identification, parameter estimation, and state assimilation throughout the DA process.

We applied our method to the advection model and the Lorenz 96 model, which are commonly used in spatial-temporal process studies. In both cases, our method achieved significantly lower MSPEs lower MSPEs compared to pMCMC methods that do not account for the evolution model uncertainty. This improvement stems from our method's ability to consider multiple candidate models and generate more diverse ensembles of particles, resulting in more robust predictions. This aligns with the concept of Bayesian Model Averaging (Zhu et al., 2016). However, existing methods often rely on fixed sample sizes per model or require higher-quality observations (Bach and Ghil, 2023). Inspired by a reviewer's suggestion, integrating our method with BMA presents a promising direction for future research. The large MSPEs of the pMCMC method with $k = 1$ in the advection model example and $k = 0$ in the Lorenz 96 model example highlight the risks of incorrect model assumptions, such as an inaccurate number of discontinuous points, which can lead to the PF divergence. Our method effectively mitigates this issue with only a modest increase in computational cost.

In the case of the advection model, as shown in Fig. 10, our resampling method effectively identifies the locations of velocity change points. After the resampling step, most particles have chosen the parameter $k = 2$ or 3, which means there are either 2 or 3 discontinuity points. For particles with 3 discontinuity points, the third discontinuity point is near the left boundary of the interval, which has little influence on the prediction. This suggest that our method accurately identifies the form of evolution equation. The presence of particles identifying a third discontinuity point may be influenced by the choice of the prior distribution, the number of observations, and the observational error. Overall, the results demonstrate that our method effectively identifies the model indices, and estimates parameters, and enables accurate prediction of future state vectors.

While promising, the proposed method may face challenges when applied to more complex systems. Extending this framework to higher-dimensional parameter spaces is an important direction for future research. The primary challenges stem from two key aspects: first, the increasing number of candidate models and higher parameter dimensionality complicate RJMCMC implementation, particularly in maintaining reasonable acceptance rates and designing effective proposal distributions; and second, higher dimensions exacerbate particle degeneracy in PF (Snyder et al., 2008). These challenges currently entail significant computational costs, a limitation shared by many advanced PF methods.

The proposal distribution in the proposed resampling method is influenced by the prior distributions of the model and parameters, which are application-specific. Appendix D provides the specific formula for the proposal distribution used in Section 4. The theoretical derivation of our resampling method is presented in Section 3 and Appendix A. From the derivation process, it can be seen that the resampled particles are an ensemble of more efficient samples drawn from the posterior distribution, enabling more accurate statistical inference.

In summary, the proposed SHBM provides a refined framework for parameter estimation in DA, effectively capturing the uncertainty in both model parameters and their dimensionality. By treating the dimensionality of parameters as an inferable parameter, our model and method provide enhanced flexibility for various applications. The proposed PF estimation with two-step RJMCMC resampling demonstrates strong performance in both case studies. The R code used to implement our methods on the examples is available at: github.com/lwhuanyue/Two-stage-Resampling-PF-with-RJMCMC.

Appendix A. The SMC procedure

A.1. The initial conditions of SMC

At the initial time $t = 0$, we assume the initial state x_0 follows a prior distribution π_{x_0}, and the parameter θ follows a prior distribution $\pi_\theta^0 = \rho^0(k) \cdot \pi_{\xi_k}^0$, where π_{x_0} and π_θ^0 are independent. The joint density function of x_1, x_0, and parameter θ is given by:

$$\begin{aligned} f_{x_1,x_0,\theta} &= f_{x_1|x_0,\theta} \cdot f_{x_0,\theta} \\ &= \phi(x_1; M(x_0|\theta), \sigma^2 R_n) \cdot \pi_{x_0} \cdot \pi_\theta^0, \end{aligned} \tag{A.1}$$

where ϕ denotes the probability density function of a multivariate Gaussian distribution. At every time step, a PF generates samples - also referred to as ensemble members - from the distribution defined by the SHBM in Eq. (3). An ensemble of N samples from the joint distribution $f_{x_1,x_0,\theta}$ is obtained using the sampling method for random vectors in Algorithm 1.

A.2. Estimation in the initial steps of SMC

At time $t = 1$, the joint density of z_1, x_1, x_0, θ is given by

$$\begin{aligned} f_{z_1,x_1,x_0,\theta} &= f_{z_1|x_1,x_0,\theta} \cdot f_{x_1,x_0,\theta} \\ &= \phi(z_1; h(x_1), \delta^2 R_m) \cdot \phi(x_1; M(x_0|\theta), \sigma^2 R_n) \cdot \pi_{x_0} \cdot \pi_\theta^0. \end{aligned} \tag{A.2}$$

Given the observation z_1

$$\begin{aligned} f_{x_1,x_0,\theta|z_1} &= \frac{1}{f_{z_1}} \cdot f_{z_1,x_1,x_0,\theta} \\ &= \frac{1}{f_{z_1}} \cdot \phi(z_1; h(x_1), \delta^2 R_m) \cdot \phi(x_1; M(x_0|\theta), \sigma^2 R_n) \cdot \pi_{x_0} \cdot \pi_\theta^0. \end{aligned} \tag{A.3}$$

Since $f_{x_0,\theta} = \pi_{x_0} \cdot \pi_\theta$, x_0 and θ are independent, we draw N independent samples $\{x_0^{(i)}, \theta^{(i)}\}_{i=1}^N$ according to π_{x_0} and π_θ respectively. These form an ensemble from the distribution $f_{x_0,\theta}$. Next, we draw samples x_1 from $f_{x_1|x_0,\theta}(x_1|x_0 = x_0^{(i)}, \theta = \theta^{(i)}) = \phi(x_1; M(x_0^{(i)}|\theta^{(i)}), \sigma^2 R_n)$. Specifically, we sample $\epsilon_1^{(i)} \sim \mathcal{N}_n(0, \sigma^2 R_n)$, and compute $x_1^{(i)} = M(x_0^{(i)}|\theta^{(i)}) + \epsilon_1^{(i)}$. This yields an ensemble $\{x_1^{(i)}, x_0^{(i)}, \theta^{(i)}\}_{i=1}^N$ from the joint distribution $f_{x_1,x_0,\theta}$. As the ensemble members are sampled independently from the same distribution, the initial weights are uniformly assigned as $w_0^{(i)} = \frac{1}{N}$.

For any function $G(x_1, x_0, \theta)$, the posterior estimate given z_1 is

$$\begin{aligned} \hat{G} &= E[G(x_1, x_0, \theta)|z_1] \\ &= \int G(x_1, x_0, \theta) \cdot f_{x_1,x_0,\theta|z_1} dx_1\ dx_0\ d\theta \\ &= \int G(x_1, x_0, \theta) \cdot \frac{f_{x_1,x_0,\theta|z_1}}{f_{x_1,x_0,\theta}} f_{x_1,x_0,\theta} dx_1\ dx_0\ d\theta, \end{aligned} \tag{A.4}$$

Using the identity:

$$\frac{f_{x_1,x_0,\theta|z_1}}{f_{x_1,x_0,\theta}} = \frac{f_{z_1|x_1,x_0,\theta} \cdot f_{x_1,x_0,\theta} / f_{z_1}}{f_{x_1,x_0,\theta}} = \frac{f_{z_1|x_1,x_0,\theta}}{f_{z_1}}, \tag{A.5}$$

we obtain the following expectation which can be approximated using the Monte Carlo estimator:

$$\begin{aligned} \hat{G} &= \int G(x_1, x_0, \theta) \cdot \frac{f_{z_1|x_1,x_0,\theta}}{f_{z_1}} f_{x_1,x_0,\theta} dx_1\ dx_0\ d\theta \\ &= E_{x_1,x_0,\theta}[G(x_1, x_0, \theta) \cdot \frac{f_{z_1|x_1,x_0,\theta}}{f_{z_1}}] \\ &= E_{x_1,x_0,\theta}[\frac{\phi(z_1; h(x_1), \delta^2 R_m)}{f_{z_1}} \cdot G(x_1, x_0, \theta)] \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{\phi(z_1; h(x_1^{(i)}), \delta^2 R_m)}{f_{z_1}} \cdot G(x_1^{(i)}, x_0^{(i)}, \theta^{(i)}). \end{aligned} \tag{A.6}$$

The marginal likelihood f_{z_1} can be approximated:

$$\begin{aligned} f_{z_1} &= \int f_{z_1|x_1} \cdot f_{x_1} dx_1 \\ &= E_{x_1}[f_{z_1|x_1}] \\ &\approx \frac{1}{N} \Sigma_{j=1}^N \phi(z_1; h(x_1^{(j)}), \delta^2 R_m) \end{aligned} \tag{A.7}$$

so

$$\begin{aligned} \hat{G} &\approx \frac{1}{N} \sum_{i=1}^N \frac{\phi(z_1; h(x_1^{(i)}), \delta^2 R_m)}{f_{z_1}} \cdot G(x_1^{(i)}, x_0^{(i)}, \theta^{(i)}) \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{\phi(z_1; h(x_1^{(i)}), \delta^2 R_m)}{\frac{1}{N} \Sigma_{j=1}^N \phi(z_1; h(x_1^{(j)}), \delta^2 R_m)} \cdot G(x_1^{(i)}, x_0^{(i)}, \theta^{(i)}) \\ &= \sum_{i=1}^N \frac{\phi(z_1; h(x_1^{(i)}), \delta^2 R_m)}{\Sigma_{j=1}^N \phi(z_1; h(x_1^{(j)}), \delta^2 R_m)} \cdot G(x_1^{(i)}, x_0^{(i)}, \theta^{(i)}) \\ &\triangleq \sum_{i=1}^N w_1^{(i)} \cdot G(x_1^{(i)}, x_0^{(i)}, \theta^{(i)}), \end{aligned} \tag{A.8}$$

where

$$w_1^{(i)} = \frac{\phi(z_1; h(x_1^{(i)}), \delta^2 R_m)}{\Sigma_{j=1}^N \phi(z_1; h(x_1^{(j)}), \delta^2 R_m)}. \tag{A.9}$$

Using the resampling strategy of the original PF (Gordon et al., 1993), draw N samples with replacement from $\{x_1^{(i)}, x_0^{(i)}, \theta^{(i)}\}_{i=1}^N$ with probabilities $\{w_1^{(i)}\}_{i=1}^N$. This results in a new set of particles $\{x_1^{(i)\prime}, x_0^{(i)\prime}, \theta^{(i)\prime}\}_{i=1}^N$, which are samples from the posterior distribution

$f_{x_1,x_0,\theta|z_1}$. For convenience, we relabel these N particles as $\{x_1^{(i)}, x_0^{(i)}, \theta^{(i)}\}_{i=1}^N$. Consequently, the ensemble $\{x_1^{(i)}, \theta^{(i)}\}_{i=1}^N$ samples from the marginal posterior $f_{x_1,\theta|z_1}$ (Gordon et al., 1993).

At time $t = 2$, we begin with the sample $\{x_0^{(i)}, x_1^{(i)}, \theta^{(i)}\}_{i=1}^N$ from the joint distribution $f_{x_0,x_1,\theta|z_1}$. The next step is to obtain a sample realization of the joint distribution $f_{x_{0:2},\theta|z_1}$. Using conditional probability density sampling, for each triplet $(x_1^{(i)}, x_0^{(i)}, \theta^{(i)})$, $i = 1, 2, \cdots, N$, draw $x_2^{(i)}$ from the conditional distribution $f_{x_2|x_1,x_0,\theta,z_1} = \phi(x_2; M(x_1^{(i)}|\theta^{(i)}), \sigma^2 R_n)$. This yields the samples $\{x_{0:2}^{(i)}, \theta^{(i)}\}_{i=1}^N$ from the joint distribution $f_{x_{0:2},\theta|z_1}$. The posterior distribution given observations z_1 and z_2 is:

$$\begin{aligned} f_{x_{0:2},\theta|z_{1:2}} &= \frac{f_{z_2,x_{0:2},\theta|z_1}}{f_{z_2|z_1}} \\ &= \frac{1}{f_{z_2|z_1}} f_{z_2|x_{0:2},\theta,z_1} \cdot f_{x_{0:2},\theta|z_1} \\ &= \frac{1}{f_{z_2|z_1}} \cdot \phi(z_2; h(x_2), \delta^2 R_m) \cdot f_{x_{0:2},\theta|z_1} \end{aligned} \tag{A.10}$$

For any t, let $dx_{0:t}$ denote $dx_0\, dx_1 \cdots dx_t$. Then, for any function $G(x_{0:2}, \theta)$ of x_2, x_1, θ, the posterior estimate given observations $z_{1:2}$ is

$$\begin{aligned} \hat{G} &= E[G(x_{0:2},\theta)|z_{1:2}] \\ &= \int G(x_{0:2},\theta) \cdot f_{x_{0:2},\theta|z_1,z_2} dx_{0:2} d\theta \\ &= \int G(x_{0:2},\theta) \cdot \frac{f_{x_{0:2},\theta|z_1,z_2}}{f_{x_{0:2},\theta|z_1}} f_{x_{0:2},\theta|z_1} dx_{0:2} d\theta \\ &= \int G(x_{0:2},\theta) \cdot \frac{\phi(z_2; h(x_2), \delta^2 R_m)}{f_{z_2|z_1}} \cdot f_{x_{0:2},\theta|z_1} dx_{0:2} d\theta \\ &= E_{x_{0:2},\theta|z_1}[G(x_{0:2},\theta) \cdot \frac{\phi(z_2; h(x_2), \delta^2 R_m)}{f_{z_2|z_1}}] \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{\phi(z_2; h(x_2^{(i)}), \delta^2 R_m)}{f_{z_2|z_1}} \cdot G(x_{0:2}^{(i)}, \theta^{(i)}), \end{aligned} \tag{A.11}$$

According to equation (63) in Doucet et al. (2000),

$$\begin{aligned} f_{z_2|z_1} &= \int f_{z_2|x_2,z_1} \cdot f_{x_2|z_1} dx_2 \\ &= E_{x_2|z_1}[f_{z_2|x_2,z_1}] \\ &\approx \frac{1}{N} \sum_{j=1}^N f_{z_2|x_2}(z_2|x_2 = x_2^{(j)}) \\ &= \frac{1}{N} \sum_{j=1}^N \phi(z_2; h(x_2^{(j)}), \delta^2 R_m), \end{aligned} \tag{A.12}$$

Therefore,

$$\begin{aligned} \hat{G} &= E[G(x_{0:2},\theta)|z_{1:2}] \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{\phi(z_2; h(x_2^{i)}), \delta^2 R_m)}{f_{z_2|z_1}} \cdot G(x_{0:2}^{(i)}, \theta^{(i)}) \\ &\approx \frac{1}{N} \sum_{i=1}^N \frac{\phi(z_2; h(x_2^{(i)}), \delta^2 R_m)}{\frac{1}{N}\sum_{j=1}^N \phi(z_2; h(x_2^{(j)}), \delta^2 R_m)} \cdot G(x_{0:2}^{(i)}, \theta^{(i)}) \\ &= \sum_{i=1}^N \frac{\phi(z_2; h(x_2^{(i)}), \delta^2 R_m)}{\sum_{j=1}^N \phi(z_2; h(x_2^{(j)}), \delta^2 R_m)} \cdot G(x_{0:2}^{(i)}, \theta^{(i)}) \\ &\triangleq \sum_{i=1}^N w_2^{(i)} \cdot G(x_{0:2}^{(i)}, \theta^{(i)}), \end{aligned} \tag{A.13}$$

where

$$w_2^{(i)} = \frac{\phi(z_2; h(x_2^{(i)}), \delta^2 R_m)}{\sum_{j=1}^N \phi(z_2; h(x_2^{(j)}), \delta^2 R_m)}. \tag{A.14}$$

Resample $\{x_{0:2}^{(i)}, \theta^{(i)}\}_{i=1}^N$ with probabilities $\{w_2^{(i)}\}_{i=1}^N$ and relabel the resampled particles as $\{x_{0:2}^{(i)}, \theta^{(i)}\}_{i=1}^N$, then these are N samples from the posterior distribution $f_{x_{0:2},\theta|z_{1:2}}$.

Assume that at time $t-1$, $\{x_{0:t-1}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ is a sample realization from the joint distribution $f_{x_{0:t-1},\theta|z_{1:t-1}}$. Using conditional probability density sampling, for each $\{x_{t-1}^{(i)}, \theta^{(i)})\}$, $i = 1, 2, \cdots, N$, draw $x_t^{(i)}$ from the conditional distribution $f_{x_t|x_{0:t-1},\theta,z_{1:t-1}} = \phi(x_t; M(x_{t-1}^{(i)}|\theta^{(i)}), \sigma^2 R_n)$. This yields an ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ from the joint distribution $f_{x_{0:t},\theta|z_{1:t-1}}$.

A.3. Estimation in the general steps of SMC

At time t, for any function $G(x_{0:t}, \theta)$, given the observations $z_{1:t}$, its posterior estimate is given by:

$$
\begin{aligned}
\hat{G} &= E[G(x_{0:t},\theta)|z_{1:t}] \\
&= \int G(x_{0:t},\theta)\cdot f_{x_{0:t},\theta|z_{1:t}} dx_{0:t} d\theta \\
&= \int G(x_{0:t},\theta)\cdot \frac{f_{x_{0:t},\theta|z_{1:t-1},z_t}}{f_{x_{0:t},\theta|z_{1:t-1}}} f_{x_{0:t},\theta|z_{1:t-1}} dx_{0:t} d\theta \\
&= \int G(x_{0:t},\theta)\cdot \frac{\phi(z_t; h(x_t), \delta^2 R_m)}{f_{z_t|z_{1:t-1}}} \cdot f_{x_{0:t},\theta|z_{1:t-1}} dx_{0:t} d\theta \\
&= E_{x_{0:t},\theta|z_{1:t-1}}[G(x_{0:t},\theta)\cdot \frac{\phi(z_t; h(x_t), \delta^2 R_m)}{f_{z_t|z_{1:t-1}}}] \\
&\approx \frac{1}{N}\sum_{i=1}^{N} \frac{\phi(z_t; h(x_t^{(i)}), \delta^2 R_m)}{f_{z_t|z_{1:t-1}}} \cdot G(x_{0:t}^{(i)}, \theta^{(i)}) \\
&\approx \frac{1}{N}\sum_{i=1}^{N} \frac{\phi(z_t; h(x_t^{(i)}), \delta^2 R_m)}{\frac{1}{N}\sum_{j=1}^{N} \phi(z_t; h(x_t^{(j)}), \delta^2 R_m)} \cdot G(x_{0:t}^{(i)}, \theta^{(i)}) \\
&= \sum_{i=1}^{N} \frac{\phi(z_t; h(x_t^{(i)}), \delta^2 R_m)}{\sum_{j=1}^{N} \phi(z_t; h(x_t^{(j)}), \delta^2 R_m)} \cdot G(x_{0:t}^{(i)}, \theta^{(i)}) \\
&\triangleq \sum_{i=1}^{N} w_t^{(i)} \cdot G(x_{0:t}^{(i)}, \theta^{(i)}),
\end{aligned}
\tag{A.15}
$$

where

$$
\begin{aligned}
f_{x_{0:t},\theta|z_{1:t-1},z_t} &= \frac{f_{z_t,x_{0:t},\theta|z_{1:t-1}}}{f_{z_t|z_{1:t-1}}} \\
&= \frac{f_{z_t|x_{0:t},\theta,z_{1:t-1}} \cdot f_{x_{0:t},\theta|z_{1:t-1}}}{f_{z_t|z_{1:t-1}}} \\
&= \frac{\phi(z_t; h(x_t), \delta^2 R_m) \cdot f_{x_{0:t},\theta|z_{1:t-1}}}{f_{z_t|z_{1:t-1}}}
\end{aligned}
\tag{A.16}
$$

and

$$
w_t^{(i)} = \frac{\phi(z_t; h(x_t^{(i)}), \delta^2 R_m)}{\sum_{j=1}^{N} \phi(z_t; h(x_t^{(j)}), \delta^2 R_m)}. \tag{A.17}
$$

Resample $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$ with probabilities $\{w_t^{(i)}\}_{i=1}^{N}$, and relabel the resampled particles as $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$, then they are an ensemble of samples from $f_{x_{0:t},\theta|z_{1:t}}$. Therefore, at each time step, the PF maintains an ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^{N}$, representing samples from the posterior distribution $f_{x_{0:t},\theta|z_{1:t}}$.

Appendix B. Estimation of the conditional distribution of parameters

In this appendix, we derive the estimation of the conditional distribution of parameters $f_{\theta|x_{0:t},z_{1:t}}$. First, consider the conditional density of θ, $f_{\theta|x_{0:t},z_{1:t}}$, which can be derived from the recursive expression of $f_{\theta|x_{0:t-1},z_{1:t-1}}$.

At time $t = 1$, the joint distribution of x_1 and θ is:

$$
\begin{aligned}
f_{x_1,\theta|z_1,x_0} &= \frac{1}{f_{z_1|x_0}} \cdot f_{z_1,x_1,\theta|x_0} \\
&= \frac{1}{f_{z_1|x_0}} \cdot \phi(z_1; h(x_1), \delta^2 R_m) \cdot \phi(x_1; M(x_0|\theta), \sigma^2 R_n) \cdot \pi_\theta^0.
\end{aligned}
\tag{B.1}
$$

Thus, the conditional density function of θ given z_1 and $x_{0:1}$ is:

$$\begin{aligned} f_{\theta|z_1,x_{0:1}} &= \frac{f_{\theta,x_1,z_1|x_0}}{f_{x_1,z_1|x_0}} \\ &= \frac{f_{\theta,x_1,z_1|x_0}}{\int f_{\theta,x_1,z_1|x_0} d\theta} \\ &= \frac{\phi\left(z_1; h\left(x_1\right), \delta^2 R_m\right) \cdot \phi\left(x_1; M\left(x_0 \mid \theta\right), \sigma^2 R_n\right) \cdot \pi_\theta^0}{E_\theta[\phi\left(z_1; h\left(x_1\right), \delta^2 R_m\right) \cdot \phi\left(x_1; M\left(x_0 \mid \theta\right), \sigma^2 R_n\right)]}. \end{aligned} \tag{B.2}$$

At time $t = 2$, we have

$$\begin{aligned} f_{z_2,x_2,\theta|z_1,x_{0:1}} &= f_{z_2|x_2,\theta,z_1,x_{0:1}} \cdot f_{x_2|\theta,z_1,x_{0:1}} \cdot f_{\theta|z_1,x_{0:1}} \\ &= \phi\left(z_2; h\left(x_2\right), \delta^2 R_m\right) \cdot \phi\left(x_2; M\left(x_1 \mid \theta\right), \sigma^2 R_n\right) \cdot f_{\theta|z_1,x_{0:1}}. \end{aligned} \tag{B.3}$$

Thus, the conditional density function of θ given $z_{1:2}$ and $x_{1:2}$ is:

$$\begin{aligned} f_{\theta|z_{1:2},x_{0:2}} &= \frac{f_{\theta,x_2,z_2|z_1,x_{0:1}}}{f_{x_2,z_2|z_1,x_{0:1}}} \\ &= \frac{f_{\theta,x_2,z_2|z_1,x_{0:1}}}{\int f_{\theta,x_2,z_2|z_1,x_{0:1}} d\theta} \\ &= \frac{\phi\left(z_2; h\left(x_2\right), \delta^2 R_m\right) \cdot \phi\left(x_2; M\left(x_1 \mid \theta\right), \sigma^2 R_n\right) \cdot f_{\theta|z_1,x_{0:1}}}{E_{\theta|z_1,x_{0:1}}[\phi\left(z_2; h\left(x_2\right), \delta^2 R_m\right) \cdot \phi\left(x_2; M\left(x_1 \mid \theta\right), \sigma^2 R_n\right)]}. \end{aligned} \tag{B.4}$$

At time $t-1$, we assume the conditional density function of θ given $z_{1:t-1}$ and $x_{0:t-1}$ is:

$$f_{\theta|x_{0:t-1},z_{1:t-1}}. \tag{B.5}$$

Then, at time t, the conditional density of θ becomes:

$$\begin{aligned} f_{\theta|x_{0:t},z_{1:t}} &= \frac{f_{z_t,x_t,\theta|z_{1:t-1},x_{0:t-1}}}{f_{z_t,x_t|z_{1:t-1},x_{0:t-1}}} \\ &= \frac{f_{z_t,x_t,\theta|z_{1:t-1},x_{0:t-1}}}{\int f_{z_t,x_t,\theta|z_{1:t-1},x_{0:t-1}} d\theta} \\ &= \frac{\phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1} \mid \theta\right), \sigma^2 R_n\right) \cdot f_{\theta|x_{0:t-1},z_{1:t-1}}}{E_{\theta|x_{0:t-1},z_{1:t-1}}[\phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1} \mid \theta\right), \sigma^2 R_n\right)]}, \end{aligned} \tag{B.6}$$

From Eqs. (B.2), (B.4), and (B.6), we can derive a recursive expression for the conditional density $f_{\theta|x_{0:t},z_{1:t}}$.

Although this expression is analytically intractable in practice, it can be approximated by samples $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ from $f_{x_{0:t},\theta|z_{1:t}}$ at each time step.

The conditional density can also be expressed as:

$$f_{\theta|x_{0:t},z_{1:t}} = \frac{f_{\theta,x_{0:t},z_{1:t}}}{f_{x_{0:t},z_{1:t}}} = \frac{f_{\theta,x_{0:t}|z_{1:t}}}{f_{x_{0:t}|z_{1:t}}}. \tag{B.7}$$

However, this form is not directly computable from the particle ensemble $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ sampled from $f_{x_{0:t},\theta|z_{1:t}}$. However, naturally, $\{\theta^{(i)}\}_{i=1}^N$ from $\{x_{0:t}^{(i)}, \theta^{(i)}\}_{i=1}^N$ are samples of $f_{\theta|z_{1:t}}$, and $\{x_{0:t}^{(i)}\}_{i=1}^N$ are N samples drawn with equal probability.

Given $z_{1:t}$ and $\{x_{0:t}^{(i)}\}_{i=1}^N$, the posterior distribution of$x_{0:t}$ is:

$$\begin{aligned} f_{x_{0:t}|z_{1:t}} &= \frac{f_{z_{1:t}|x_{0:t}} \cdot f_{x_{0:t}}}{f_{z_{1:t}}} \\ &\propto f_{z_{1:t}|x_{0:t}} \cdot f_{x_{0:t}} \\ &= \prod_{t_q=1}^{t} \phi\left(z_{t_q}; h\left(x_{t_q}\right), \delta^2 R_m\right) \cdot f_{x_{0:t}}. \end{aligned} \tag{B.8}$$

This distribution can be estimated using the importance sampling. Given $\{x_{0:t}^{(i)}\}_{i=1}^N$ and $z_{1:t}$, it is difficult to draw corresponding θ samples from $f_{\theta|x_{0:t},z_{1:t}}$. However, through formula (B.7), let $\{\theta^{(i)}\}_{i=1}^N$ be an ensemble of samples directly from $f_{\theta|x_{0:t},z_{1:t}}$, and assign $\theta^{(i)}$ importance weights $u^{(i)}$ (Doucet et al., 2001; Bishop and Nasrabadi, 2006). Since $\{x_{0:t}^{(i)}\}_{i=1}^N$ are equally probable samples from $f_{x_{0:t}}$, the weights are given by:

$$u^{(i)} = \frac{1/\prod_{t_q=1}^{t} \phi\left(z_{t_q}; h\left(x_{t_q}^{(i)}\right), \delta^2 R_m\right)}{\sum_{j=1}^{N}[1/\prod_{t_q=1}^{t} \phi\left(z_{t_q}; h\left(x_{t_q}^{(j)}\right), \delta^2 R_m\right)]}. \tag{B.9}$$

There are many methods to estimate the distribution of parameters $\hat{f}_{\theta|x_{0:t},z_{1:t}} = \hat{\rho}(k) \cdot \hat{\pi}_k(\xi_k)$. Under the assumption that random variables $\rho(k)$ and $\pi_k(\xi_k)$ follow a known distribution family with unspecified parameters, parametric estimation methods can be employed. For instance, if $\rho(k)$ follows a Poisson distribution, its intensity parameter λ can be estimated from samples via maximum likelihood or moment estimation. This approach achieves high efficiency with minimal samples when model specifications are correct, as it only requires estimating a limited number of parameters. When estimating the prior distribution, one may also specify the distribution form and corresponding parameters based on empirical knowledge. In such cases, appropriate parameters are often selected to make the distribution relatively flat. The hyperparameter of $\rho(k)$ and $\pi_k(\xi_k)$ can be estimated after the bootstrap resampling, which can generate an ensemble of equally weighted samples.

If no assumption is made about the distribution, nonparametric estimation methods can be utilized. For $k = 1, 2, \cdots, K$, let

$$\hat{\rho}(k) = \sum_{j=1}^{N} u^{(i)} \mathbb{I}(k_i = k). \tag{B.10}$$

$\mathbb{I}(\cdot)$ is the indicator function, which equals 1 when the ith ensemble member follows the kth model, i.e., when its model index $k_i = k$, $\mathbb{I}(k_i = k) = 1$; otherwise, $\mathbb{I}(k_i = k) = 0$. Let N_k be the number of particles for $k_i = k$. The corresponding probability density function denoted as $\pi_{\xi_k}(\xi_k)$, which may represent either a prior distribution or a posterior distribution. As in DA scheme, the posterior distribution of parameter at time step t will be the prior distribution for the next time step. The estimated density function of ξ_k given k from the N_k samples of $k_i = k$ is denoted as $\hat{\pi}_{\xi_k}(\xi_k)$.

For $k = 1, 2, \cdots, K$, the kernel density estimate of the conditional distribution $\pi_k(\xi_k)$ is given by:

$$\hat{\pi}_k(\xi_k) = \sum_{i_k=1}^{N_k} \frac{u^{(i_k)}}{\sum_{i_k=1}^{N_k} u^{(j_k)}} K_h(\xi_k - \xi_k^{(i_k)}), \tag{B.11}$$

where $\int \hat{\pi}_k(\xi_k) d\xi_k = 1$. Here, K_h is a multidimensional kernel function. A common choice for K_h is the multidimensional Gaussian kernel:

$$K_h(\xi_k - \xi_k^{(i_k)}) = \frac{1}{(2\pi)^{r_k/2}|H|^{1/2}} \exp\left(-\frac{1}{2}(\xi_k - \xi_k^{(i_k)})^T H^{-1}(\xi_k - \xi_k^{(i_k)})\right), \tag{B.12}$$

where H is the bandwidth matrix and r_k is the dimension of ξ_k. Thus, the kernel density estimate of $\pi_{\xi_k}(\xi_k)$ is obtained as $\hat{\pi}_{\xi_k}(\xi_k)$, and the corresponding estimate of $f_{\theta|x_{0:t},z_{1:t}}$ is $\hat{\rho}(k) \cdot \hat{\pi}_{\xi_k}(\xi_k)$. Similarly, this method can also be used to estimate the prior distribution of θ at time t, $f_{\theta|x_{0:t-1},z_{1:t-1}}$, which is required in the acceptance ratio in Eq. (17).

Appendix C. Details of the RJMCMC resampling step for APF

For a particle $(x_{0:t}^{(i)}, k_i, \xi_k^{(i)})$ that needs to be resampled, a new particle $(x_{0:t}^*, k^*, \xi_{k^*}^*)$ is proposed using a Markov transition function $g_{((x_{0:t},\theta)\to(x_{0:t}^*,\theta^*))}$, with an acceptance probability $\alpha_{((x_{0:t},\theta)\to(x_{0:t}^*,\theta^*))}$ (Doucet et al., 2000; Elvira et al., 2018), such that:

$$\begin{aligned} &g_{((x_{0:t},\theta)\to(x_{0:t}^*,\theta^*))}\alpha_{((x_{0:t},\theta)\to(x_{0:t}^*,\theta^*))} f_{x_{0:t},\theta|z_{0:t}} \\ =&g_{((x_{0:t}^*,\theta^*)\to(x_{0:t},\theta))}\alpha_{((x_{0:t}^*,\theta^*)\to(x_{0:t},\theta))} f_{x_{0:t}^*,\theta^*|z_{0:t}}, \end{aligned} \tag{C.1}$$

where $\theta = (k, \xi_k)$.

If $(x_{0:t}^{(i)}, \theta^{(i)})$ is a duplicate particle, only $(x_t^{(i)}, \theta^{(i)})$ is used to draw $x_{t+1}^{(i)}$ at the next time step. Therefore, we retain $x_{0:t-1}^{(i)}$ and draw a new $(x_t^{(i)}, \theta^{(i)})$ from $f_{x_t,\theta|z_{1:t},x_{0:t-1}}$. The joint posterior distribution of x_t and θ given $x_{0:t-1}$ and observations $z_{1:t}$ is:

$$f_{x_t,\theta|z_{1:t},x_{0:t-1}} = \frac{f_{z_t|x_t,x_{0:t-1},\theta,z_{1:t-1}} \cdot f_{x_t|x_{0:t-1},\theta,z_{1:t-1}} \cdot f_{\theta|x_{0:t-1},z_{1:t-1}}}{f_{z_t|x_{0:t-1},z_{1:t-1}}}. \tag{C.2}$$

Then the conditional density of θ given $x_{0:t}$ and $z_{1:t}$ is:

$$\begin{aligned} f_{\theta|x_t,x_{0:t-1},z_{1:t}} &= \frac{f_{z_t,x_t,\theta|z_{1:t-1},x_{0:t-1}}}{f_{z_t,x_t|z_{1:t-1},x_{0:t-1}}} \\ &= \frac{f_{z_t,x_t,\theta|z_{1:t-1},x_{0:t-1}}}{\int f_{z_t,x_t,\theta|z_{1:t-1},x_{0:t-1}} d\theta} \\ &= \frac{\phi(z_t; h(x_t), \delta^2 R_m) \cdot \phi(x_t; M(x_{0:t-1} \mid \theta), \sigma^2 R_n) \cdot f_{\theta|x_{0:t-1},z_{1:t-1}}}{E_{\theta|x_{0:t-1},z_{1:t-1}}[\phi(z_t; h(x_t), \delta^2 R_m) \cdot \phi(x_t; M(x_{0:t-1} \mid \theta), \sigma^2 R_n)]}, \end{aligned} \tag{C.3}$$

Similarly, the conditional density of x_t given θ and $z_{1:t}$ is

$$f_{x_t|\theta,x_{0:t-1},z_{1:t}} = \frac{1}{f_{z_t|\theta,x_{0:t-1},z_{1:t-1}}} \cdot \phi(z_t; h(x_t), \delta^2 R_m) \cdot \phi(x_t; M(x_{t-1} \mid \theta), \sigma^2 R_n). \tag{C.4}$$

From Eq. (C.2), the joint conditional probability density function of x_t and θ given $x_{0:t-1}$ and $z_{1:t}$ can be further written as

$$\begin{aligned} f_{x_t,\theta|z_{1:t},x_{0:t-1}} &= \frac{f_{z_t|x_t,x_{0:t-1},\theta,z_{1:t-1}} \cdot f_{x_t|x_{0:t-1},\theta,z_{1:t-1}} \cdot f_{\theta|x_{0:t-1},z_{1:t-1}}}{f_{z_t|x_{0:t-1},z_{1:t-1}}} \\ &= \frac{1}{f_{z_t|x_{0:t-1},z_{1:t-1}}} \cdot \phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1} \mid \theta\right), \sigma^2 R_n\right) \cdot \\ & f_{\theta|x_{0:t-1},z_{1:t-1}}. \end{aligned} \tag{C.5}$$

For a fixed $i \in \{1, 2, \cdots, N\}$, given $x_{0:t-1}^{(i)}$ and observations $z_{1:t}$, the target distribution for the MCMC step is:

$$\begin{aligned} & f_{x_t,\theta|z_{1:t}=z_{1:t},x_{0:t-1}=x_{0:t-1}^{(i)}} \\ &= \frac{\phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1}^{(i)} \mid \theta\right), \sigma^2 R_n\right) \cdot f_{\theta|x_{0:t-1}=x_{0:t-1}^{(i)},z_{1:t-1}}}{f_{z_t|x_{0:t-1}=x_{0:t-1}^{(i)},z_{1:t-1}=z_{1:t-1}}}. \end{aligned} \tag{C.6}$$

Therefore, an MCMC method can be used to construct a Markov chain whose stationary distribution is the target posterior distribution $f_{x_t,\theta|z_{1:t}=z_{1:t},x_{0:t-1}=x_{0:t-1}^{(i)}}$. Existing particle filters using MCMC methods as resampling methods only apply to cases where the dimension of the particles does not change. However, since $\theta = (k, \xi_k)$, the parameter space dimension can change from pre-resampling to post-resampling, so we refer to Green's RJMCMC (Green, 1995), which allows the dimension of the Markov chain to change.

The posterior distribution of the x_t and θ given observations $z_{1:t}$ and a specific particle of the previous state $x_{0:t-1}^{(i)}$ is

$$\begin{aligned} & f_{x_t,\theta|z_t,x_{0:t-1}=x_{0:t-1}^{(i)}} \\ & \propto \phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1}^{(i)} \mid \theta\right), \sigma^2 R_n\right) \cdot f_{\theta|x_{0:t-1}=x_{0:t-1}^{(i)},z_{1:t-1}} \\ & \approx \phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \phi\left(x_t; M\left(x_{t-1}^{(i)} \mid \theta\right), \sigma^2 R_n\right) \cdot \hat{\rho}(k) \cdot \hat{\pi}_{\xi_k}(\xi_k). \end{aligned} \tag{C.7}$$

The current ith particle is given by $(x_t^{(i)}, \theta^{(i)}) = (x_t^{(i)}, k_i, \theta_{k_i}^{(i)})$, and the proposed particle is $(x_t^*, \theta^*) = (x_t^*, k^*, \xi_{k^*}^*)$. If the dimensions of $\theta^{(i)}$ and θ^* are the same, it is easy to satisfy the balance condition. We define a Markov transition function $g(\cdot|\cdot)$ to denote the density from the current particle to the proposed particle (or denoted as $g_{\cdot\to\cdot}$) with an associated acceptance probability function $\alpha(\cdot \to \cdot)$ such that

$$\begin{aligned} & g(\theta^*, x_t^*|\theta^{(i)}, x_t^{(i)})\alpha(\theta^{(i)}, x_t^{(i)} \to \theta^*, x_t^*) f_{x_t^{(i)},\theta^{(i)}|z_t,x_{t-1}=x_{t-1}^{(i)}} = \\ & g(\theta^{(i)}, x_t^{(i)}|\theta^*, x_t^*)\alpha(\theta^*, x_t^* \to \theta^{(i)}, x_t^{(i)}) f_{x_t^*,\theta^*|z_t,x_{t-1}=x_{t-1}^{(i)}}. \end{aligned} \tag{C.8}$$

Substituting from Eq. (C.7),this becomes:

$$\begin{aligned} & g(\theta^*, x_t^*|\theta^{(i)}, x_t^{(i)})\alpha(\theta^{(i)}, x_t^{(i)} \to \theta^*, x_t^*) \cdot \phi\left(z_t; h\left(x_t^{(i)}\right), \delta^2 R_m\right) \cdot \\ & \phi\left(x_t^{(i)}; M\left(x_{t-1}^{(i)} \mid \theta^{(i)}\right), \sigma^2 R_n\right) \cdot \hat{\rho}(k_i) \cdot \hat{\pi}_{\xi_k}(\xi_{k_i}^{(i)}) \\ = & g(\theta^{(i)}, x_t^{(i)}|\theta^*, x_t^*)\alpha(\theta^*, x_t^* \to \theta^{(i)}, x_t^{(i)}) \cdot \phi\left(z_t; h\left(x_t^*\right), \delta^2 R_m\right) \cdot \\ & \phi\left(x_t^*; M\left(x_{t-1}^{(i)} \mid \theta^*\right), \sigma^2 R_n\right) \cdot \hat{\rho}(k^*) \cdot \hat{\pi}_{\xi_{k^*}}(\xi_{k^*}^*). \end{aligned} \tag{C.9}$$

The proposal density $g(\cdot|\cdot)$ differs from the standard RJMCMC formulation due to two key considerations. First, model-specific proposal design is still required. In the DA framework where observations arrive sequentially, both states and parameters are updated iteratively. Second, to maintain the dynamical system balance during proposals Moradkhani et al. (2012), our algorithm first proposes parameter, followed by the corresponding states. The likelihood function is not directly related to the parameters, but is instead derived through the observation equation that links the states and observations. We simplify the detailed balance condition via strategic state vector sampling, thereby completing the algorithmic design. Given the current particle $(x_t^{(i)}, k_i, \xi_{k_i}^{(i)})$, the proposed density can be decomposed as:

$$\begin{aligned} & g(\theta^*, x_t^*|\theta^{(i)}, x_t^{(i)}) \\ = & g(k^*, \xi_{k^*}^*, x_t^*|k = k_i, \xi_k = \xi_{k_i}^{(i)}, x_t = x_t^{(i)}) \\ = & g_1(k^*|k = k_i, \xi_k = \xi_{k_i}^{(i)}, x_t = x_t^{(i)}) \cdot g_2(\xi_{k^*}^*|k^*, k = k_i, \xi_k = \xi_{k_i}^{(i)}, x_t = x_t^{(i)}) \cdot \\ & g_3(x_t^*|k^*, \xi_{k^*}^*, k = k_i, \xi_k = \xi_{k_i}^{(i)}, x_t = x_t^{(i)}) \\ = & g_1(k^*|k = k_i, \xi_k = \xi_{k_i}^{(i)}, x_t = x_t^{(i)}) \cdot g_2(\xi_{k^*}^*|k^*, k = k_i, \xi_k = \xi_{k_i}^{(i)}, x_t = x_t^{(i)}) \cdot \\ & \phi\left(x_t^*; M\left(x_{t-1}^{(i)} \mid \theta^*\right), \sigma^2 R_n\right). \end{aligned} \tag{C.10}$$

Here, $g_1(\cdot|k=k_i,\xi_k=\xi_{k_i}^{(i)},x_t=x_t^{(i)})$ denotes the probability of selecting a new model k^* from the state $(x_t,k,\xi_k)=(x_t^{(i)},k_i,\xi_{k_i}^{(i)})$. The second part $g_2(\xi_{k^*}^*|k^*,k=k_i,\xi_k=\xi_{k_i}^{(i)},x_t=x_t^{(i)})$ is the probability of selecting new model parameters $\xi_{k^*}^*|k^*$ after choosing the new model k^*.

Substituting Eqs. (C.10) into (C.9), the balance condition becomes:

$$\begin{aligned}
&g_1(k^*|k=k_i,\xi_k=\xi_{k_i}^{(i)},x_t=x_t^{(i)})\cdot g_2(\xi_{k^*}^*|k^*,k=k_i,\xi_k=\xi_{k_i}^{(i)},x_t=x_t^{(i)})\\
&\alpha(\theta^{(i)},x_t^{(i)}\to\theta^*,x_t^*)\phi\left(z_t;h\left(x_t^{(i)}\right),\delta^2R_m\right)\cdot\hat{\rho}(k_i)\cdot\hat{\pi}_{\xi_k}(\xi_{k_i}^{(i)})\\
&=\\
&g_1(k_i|k=k^*,\xi_k=\xi_{k^*}^*,x_t=x_t^{(i)})\cdot g_2(\theta_{k_i}^{(i)}|k_i,k=k^*,\xi_k=\xi_{k^*}^*,x_t=x_t^{(i)})\\
&\alpha(\theta^*,x_t^*\to\theta^{(i)},x_t^{(i)})\phi\left(z_t;h\left(x_t^*\right),\delta^2R_m\right)\cdot\hat{\rho}(k^*)\cdot\hat{\pi}_{\xi_{k^*}}(\xi_{k^*}^*).
\end{aligned}\tag{C.11}$$

Therefore, the sampling scheme can be specified as:

$$\begin{aligned}
x_t^{(i)}&=M^{k_i}\left(x_{t-1}^{(i)}\mid\xi_k^{(i)}\right)+\epsilon_t^{(i)},\\
x_t^*&=M^{k^*}\left(x_{t-1}^*\mid\xi_{k^*}^*\right)+\epsilon_t^{(i)},
\end{aligned}\tag{C.12}$$

i.e., the same noise $\epsilon_t^{(i)}$ is used for both the current and proposed states. Finally, the proposal densities can be redefined as:

$$\begin{aligned}
g(\theta^{(i)}|\theta^*,x_t^*)&=g_1(k_i|k=k^*,\xi_k=\xi_{k^*}^*,x_t=x_t^{(i)})\cdot\\
&\qquad g_2(\theta_{k_i}^{(i)}|k_i,k=k^*,\xi_k=\xi_{k^*}^*,x_t=x_t^{(i)}),\\
g(\theta^*|\theta^{(i)},x_t^{(i)})&=g_1(k^*|k=k_i,\xi_k=\xi_{k_i}^{(i)},x_t=x_t^{(i)})\cdot\\
&\qquad g_2(\xi_{k^*}^*|k^*,k=k_i,\xi_k=\xi_{k_i}^{(i)},x_t=x_t^{(i)}).
\end{aligned}\tag{C.13}$$

The design of the proposal function $g(\cdot|\cdot)$ must account for transitions between parameter spaces while ensuring reversibility. Let $\xi_{k^*}^{(i)*}\in r_{k^*}$ and $\xi_k^{(i)}\in r_k$. The key to proposing parameters from the space Θ_k to the space Θ_{k^*} lies in introducing random auxiliary variables. Since $\xi_{k^*}^{(i)*}$ is r_{k^*}-dimensional, and $\xi_{k_i}^{(i)}$ is r_{k_i}-dimensional, the proposal distributions for the auxiliary random variables need to be designed based on the specific parameter spaces. In particular, proposing parameters from the Θ_{k_i} space to the Θ_{k^*} space requires a random variable $\mu^{(i)}\sim a_{k_i\to k^*}$, and similarly, proposing parameters from the Θ_{k^*} space to the Θ_k space requires a random variable $\mu^*\sim a_{k^*\to k_i}$. Here, $a_{\cdot\to\cdot}$ is a given probability density function. To ensure that $(\xi_k,\mu^{(i)})$ and (ξ_{k^*},μ^*) have the same dimensionality. Let $\hat{\pi}_{k,\xi_k\to k^*,\xi_{k^*}}$ be the given probability of transitioning from parameters ξ_k under model k to parameters ξ_{k^*} under model k^*. The detailed balance condition is then

$$\begin{aligned}
&g_1(k^*|k=k_i,\xi_k=\theta_{k_i}^{(i)},x_t=x_t^{(i)})\cdot g_2(\xi_{k^*}^*|k^*,k=k_i,\xi_k=\theta_{k_i}^{(i)},x_t=x_t^{(i)})\cdot\\
&a_{k_i\to k^*}\alpha(\theta^{(i)},x_t^{(i)}\to\theta^*,x_t^*)\phi\left(z_t;h\left(x_t^{(i)}\right),\delta^2R_m\right)\cdot\hat{\rho}(k_i)\cdot\hat{\pi}_{\xi_k}(\xi_{k_i}^{(i)})\\
&=\\
&g_1(k_i|k=k^*,\xi_k=\xi_{k^*}^*,x_t=x_t^{(i)})\cdot g_2(\theta_{k_i}^{(i)}|k_i,k=k^*,\xi_k=\xi_{k^*}^*,x_t=x_t^{(i)})\cdot\\
&a_{k^*\to k_i}\alpha(\theta^*,x_t^*\to\theta^{(i)},x_t^{(i)})\phi\left(z_t;h\left(x_t^*\right),\delta^2R_m\right)\cdot\hat{\rho}(k^*)\cdot\hat{\pi}_{\xi_{k^*}}(\xi_{k^*}^*).
\end{aligned}\tag{C.14}$$

This formulation leads directly to the acceptance ratio used in Eq 17.

Appendix D. Definition and derivation of the transition functions

Let the prior distribution of the initial state be denoted by π_{x_0}, and the prior distribution of the model and parameters by $\pi_\theta=\rho(k)\cdot\pi_{\xi_k}(\xi_k)$.

The probability of the true model following the kth model, i.e., the probability of having k discontinuities, follows a truncated Poisson distribution:

$$\rho(k)=\frac{e^{-\lambda}\frac{\lambda^k}{k!}}{\sum_{i=1}^{K_{max}}e^{-\lambda}\frac{\lambda^i}{i!}},\quad k=1,2,\cdots,K_{max}.\tag{D.1}$$

where λ and $K_{\max}$ are given.

Assuming the geographical space is finely discretized, i.e., the state vector x has sufficiently high dimensionality, we consider the positions of the k discontinuities $c_1,c_2,\cdots,c_k$ to be drawn from the uniform distribution over $[0,L]$. These positions correspond to the even-order statistics among $2k+1$ uniformly distributed points with $L=n$. Let F be the cumulative distribution function of the uniform distribution and f be the probability density function of the uniform distribution.

The joint density of the even-order statistics for these points is given by:

$$\begin{aligned}
&f_{2,4,\cdots,2k}\left(\xi_2,\xi_4,\cdots,\xi_{2j},\cdots,\xi_{2k+2}\right)\\
&=\int\cdots\int_{-\infty<\xi_1<\xi_2<\cdots<\xi_{2k+1}}(2k+1)!f\left(\xi_1\right)f\left(\xi_2\right)\cdots f\left(\xi_{2k}\right)f\left(\xi_{2k+1}\right)d\xi_1\ d\xi_3\cdots d\xi_{2k+1}\\
&=(2k+1)!\int_{-\infty<\xi_1<\xi_2}f\left(\xi_1\right)f\left(\xi_2\right)d\xi_1\cdot\int f\left(\xi_3\right)f\left(\xi_4\right)d\xi_3\cdot\cdots\cdot\\
&\quad\int f\left(\xi_{2k-1}\right)f\left(\xi_{2k}\right)d\xi_{2k-1}\int_{\xi_{2k}<\xi_{2k+1}<\infty}f\left(\xi_{2k+1}\right)d\xi_{2k+1}\\
&=(2k+1)!\cdot\frac{f\left(\xi_2\right)}{1!}\cdot F\left(\xi_2\right)\cdot\frac{f\left(\xi_4\right)}{1!}\left[F\left(\xi_4\right)-F\left(\xi_2\right)\right]\cdot\cdots\cdot\frac{1}{1!}\left[F\left(\xi_{2k}\right)-F\left(\xi_{2k-2}\right)\right]\cdot\\
&\quad\frac{1}{1!}\left[F(\infty)-F\left(\xi_{2k}\right)\right]\\
&=(2k+1)!\cdot F\left(\xi_2\right)\left[F\left(\xi_4\right)-F\left(\xi_2\right)\right]\cdots\left[F\left(\xi_k\right)-F\left(\xi_{k-2}\right)\right]\left[1-F\left(\xi_{2k}\right)\right]\cdot\\
&\quad f\left(\xi_2\right)\cdot f\left(\xi_4\right)\cdots\cdot f\left(\xi_{2k}\right).
\end{aligned}\tag{D.2}$$

Thus, $c_1,c_2,\cdots,c_k$ correspond to the 2nd, 4th, $\cdots$, $2k$th points from the $2k+1$ points in $[0,L]$:

$$\begin{aligned}
&f_{2,4,\cdots,2k}\left(\xi_2=c_1,\xi_4=c_2,\cdots,\xi_{2j}=c_j,\cdots,\xi_{2k}=c_k\right)\\
&=(2k+1)!\cdot\frac{c_1}{L}\cdot\frac{\left(c_2-c_1\right)}{L}\cdot\ldots\frac{\left(c_{j+1}-c_j\right)}{L}\cdot\ldots\cdot\frac{\left(L-c_k\right)}{L}\cdot\left(\frac{1}{L}\right)^k\\
&=\frac{(2k+1)!}{L^{2k+1}}c_1\left(c_2-c_1\right)\cdots\left(c_{j+1}-c_j\right)\cdots\left(L-c_k\right)\\
&\triangleq f_c,
\end{aligned}\tag{D.3}$$

The velocity parameters $v_1,v_2,\cdots,v_{k+1}$ are assumed to follow a $\Gamma(\alpha,\beta)$ distribution with the density function $\beta^\alpha v_j^{\alpha-1}e^{-\beta v_j}/\Gamma(\alpha)$, for $j=1,2,\cdots,k+1$, where α and β are given. The prior distribution over parameters is then:

$$\pi_\theta(k,\xi_k)=\rho(k)\cdot\pi_{\xi_k}(\xi_k)\propto e^{-\lambda}\frac{\lambda^k}{k!}\cdot\prod_{j=1}^{k+1}[\beta^\alpha v_j^{\alpha-1}e^{-\beta v_j}/\Gamma(\alpha)]\cdot f_c.\tag{D.4}$$

Here, π_{ξ_k} is the conditional density of π_{ξ_k} given that the model index $\tilde{k}=k$.

For simplicity in notation, the particle $(k_i,\xi_{k_i}^{(i)},x_t^{(i)})$ is abbreviated as (k,ξ_k,x_t). According to Eq. (C.10), new parameters are proposed in three steps through the RJMCMC method:

- First, propose a new model index k^* using $g_1(k^*|k=k,\xi_k=\xi_k,x_t=x_t)$,
- Second, propose new parameters $\xi_{k^*}^*$ corresponding to the k^*th model using $g_2(\xi_{k^*}^*|k^*,k=k,\xi_k=\xi_k,x_t=x_t)$,
- Third, given $(k^*,\xi_{k^*}^*)$, obtain the corresponding x_t^* using $g_3(x_t^*|k^*,\xi_{k^*}^*,k=k,\xi_k=\xi_k,x_t=x_t)$.

The model transition probability $g_1(k^*|k=k,\xi_k=\xi_k,x_t=x_t)$ defines transitions between models. Transitions are restricted to adjacent model indices or staying in the current model, resulting in four types of transitions:

- (a) Transition from model k to model $k+1$: Add a discontinuity point by drawing a point uniformly from $[0,L]$,
- (b) Transition from model k to model $k-1$: Remove a randomly selected discontinuity point randomly from the k discontinuity points,
- (c) Stay in the current model k and randomly change one of the velocity parameters v_j in one of the intervals,
- (d) Stay in the current model k and randomly change the position of one of the discontinuity points c_j.

When proposing a particle, the first step is to select the type of model change $\{(a),(b),(c),(d)\}$ with probabilities $\{b_k,d_k,\eta_k,\psi_k\}$:

$$g_1\left(k^*\mid k=k,\xi_k=\xi_k,x_t=x_t\right)=\begin{cases}b_k & k^*=k+1,\\ d_k & k^*=k-1,\\ 1-b_k-d_k & k^*=k.\end{cases}\tag{D.5}$$

The probabilities for transitions are defined as:

- For (a) - adding a discontinuity-point to increase from k to $k+1$:

$$b_k=c\cdot\min\{1,\rho(k+1)/\rho(k)\},\tag{D.6}$$

- For (b) - removing one discontinuity point:

$$d_k=c\cdot\min\{1,\rho(k-1)/\rho(k)\},\tag{D.7}$$

where $b_{K_{\max}}=d_1=0$.

- For transitions (c) proposing a new velocity η_k and (d) - proposing a new position ψ_k:

$$\eta_k = \psi_k = \frac{1}{2}(1 - b_k - d_k). \tag{D.8}$$

Here, $\lambda, \alpha, \beta, c$ are predefined hyper-parameters. Once these hyper-parameters are specified, the probabilities of the four transitions are determined, enabling the proposal of new particles across the entire parameter space Θ. For (a), the proposed new parameter $\theta^* = (k^*, \xi^*_{k^*})$ has two more dimensions than the original parameter $\theta = (k, \xi_k)$, while for (b), the proposed new parameter $\theta^* = (k^*, \xi^*_{k^*})$ has two fewer dimensions than the original parameter $\theta = (k, \xi_k)$. To reconcile this mismatch, the following steps are taken: If the proposed parameter dimension r_{k^*} differs from the original parameter dimension r_k, then: draw a random variable μ of dimension r_μ from $a_{k\to k^*}$; draw a random variable μ^* of dimension r_{μ^*} from $a_{k^*\to k}$; ensuring $(r_{k^*} + r_{\mu^*} = r_k + r_\mu)$, thus $(\xi^*_{k^*}, \mu^*) = q(\xi_k, \mu)$ is a bijection, where q is a predefined transition function. The Jacobian determinant of the transition is:

$$|J| = \left| \frac{\partial(\xi^*_{k^*}, \mu^*)}{\partial(\xi_k, \mu)} \right|. \tag{D.9}$$

For each parameter particle θ and the corresponding x_t, propose a proposal parameter particle θ^* and the corresponding x^*_t. The acceptance probability is given by the formula (17):

$$\begin{aligned}
\alpha &= \min\{1, \text{ (distribution ratio) } \times \text{ (proposal ratio) } \times |J|\} \\
&= \min\{1, \text{ (likelihood ratio) } \times (\text{ prior ratio}) \times \text{ (proposal ratio) } \times |J|\} \\
&= \min\left\{1, \frac{g(\theta|\theta^*, x^*_t) a_{k^*\to k} \phi\left(z_t; h\left(x^*_t\right), \delta^2 R_m\right) \cdot \rho(k^*) \cdot \pi_{\xi_{k^*}}(\xi^*_{k^*})}{g(\theta^*|\theta, x_t) a_{k\to k^*} \phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right) \cdot \rho(k) \cdot \pi_{\xi_k}(\xi_k)} |J|\right\},
\end{aligned} \tag{D.10}$$

where

$$\begin{aligned}
\text{(likelihood ratio)} &= \frac{\phi\left(z_t; h\left(x^*_t\right), \delta^2 R_m\right)}{\phi\left(z_t; h\left(x_t\right), \delta^2 R_m\right)} \triangleq \frac{l^*}{l^0} \\
\text{(prior ratio)} &= \frac{\rho(k^*) \cdot \pi_{\xi_{k^*}}(\xi^*_{k^*})}{\rho(k) \cdot \pi_{\xi_k}(\xi_k)} \\
\text{(proposal ratio)} &= \frac{g(\theta|\theta^*, x^*_t) a_{k^*\to k}}{g(\theta^*|\theta, x_t) a_{k\to k^*}}
\end{aligned} \tag{D.11}$$

When proposing a new parameter, first select the type of transition in the form $\{(a), (b), (c), (d)\}$ with probability distribution $\{b_k, d_k, \eta_k, \psi_k\}$, and propose new parameters $\xi^*_{k^*}$ and the corresponding x^*_t under that transition type.

Assume that the current particle follows model k, which has $k+1$ segments, the original parameter is $\theta = (k, v_1, v_2, \cdots, v_{k+1}, c_1, c_2, \cdots, c_k)$.

Type (c): Changing a velocity parameter

For (c), randomly change one of the velocity parameters v_j on an interval. The steps for proposing the new parameter are as follows:

(1) Select one of the $k+1$ velocities $\{v_1, v_2, \cdots, v_{k+1}\}$ with probability $\frac{1}{k+1}$ to change the jth velocity v_j, $j = 1, 2, \cdots, k$.
(2) Draw μ from a uniform distribution $U\left[-\frac{1}{2}, \frac{1}{2}\right]$ (where μ is the reparameterized variable).
(3) Set $\mu = \log\left(\frac{v^*_j}{v_j}\right)$. So,
(4) The proposed parameter becomes $\theta^* = (k, v_1, v_2, \cdots, v^*_j, \cdots, v_{k+1}, c_1, c_2, \cdots, c_k)$.
(5) Accept or reject the proposed parameter θ^*.

Since the proposal of new parameters is done by drawing from a uniform distribution,

$$\mu = \log\left(\frac{v^*_j}{v_j}\right) \sim U\left[-\frac{1}{2}, \frac{1}{2}\right], \tag{D.12}$$

the ratio of the acceptance probability is:

$$\begin{aligned}
p_v &= \text{ (likelihood ratio) } \times (\text{ prior ratio}) \times \text{ (proposal ratio) } \times |J| \\
&= \frac{l^*}{l^0} \cdot \frac{\beta^\alpha v^{*\alpha-1}_j e^{-\beta v^*_j} / \Gamma(\alpha)}{\beta^\alpha v^{\alpha-1}_j e^{-\beta v_j} / \Gamma(\alpha)} \left| \frac{\partial v^*_j}{\partial v_j} \right| \\
&= \frac{l^*}{l^0} \cdot \left(\frac{v^*_j}{v_j}\right)^\alpha \exp\left\{-\beta\left(v^*_j - v_j\right)\right\},
\end{aligned} \tag{D.13}$$

with the likelihood functions:

$$l_0 = \phi(z_t; h(M^k\left(x^{(i)}_{t-1} \mid \xi_k\right) + \epsilon^{(i)}_t), \delta^2 R_m) \tag{D.14}$$

and

$$l^* = \phi(z_t; h(M^{k^*}\left(x_{t-1}^{(i)} \mid \xi_{k^*}^*\right) + \epsilon_t^{(i)}), \delta^2 R_m), \tag{D.15}$$

where $x_{t-1}^{(i)}$ is the state of the ith particle at the previous time $t-1$.

The Jacobian determinant is:

$$\left|\frac{\partial v_j^*}{\partial v_j}\right| = \left|\frac{\partial v_j \cdot e^{\mu}}{\partial v_j}\right| = e^{\mu} = \frac{v_j^*}{v_j}. \tag{D.16}$$

Thus, the acceptance probability is:

$$\alpha_{velocity} = \min\{1, p_v\}. \tag{D.17}$$

Type (d): Changing a discontinuous point

For (d), changing the position of one of the discontinuity points c_j, the steps for proposing a new parameter are as follows:

(1) Select one of the k discontinuity points $j \in \{1, 2, \cdots, k\}$ with probability $\frac{1}{k}$ to change the jth discontinuity point c_j.
(2) Draw c_j^* from the uniform distribution $U\left[c_{j-1}, c_j\right]$ (i.e., uniformly select c_j^* from the coordinates contained in $U\left[c_{j-1}, c_j\right]$).
(3) Propose the parameter $\theta^* = (k, v_1, v_2, \cdots, v_{k+1}, c_1, c_2, \cdots, c_j^*, \cdots, c_k)$.
(4) Accept or reject the proposed parameter θ^*.

The ratio of the acceptance is:

$$\begin{aligned} p_c =& \frac{l^*}{l^0} \times \frac{\frac{\left(c_{j+1}-c_j^*\right)}{L} \cdot \frac{\left(c_j^*-c_{j-1}\right)}{L}}{\frac{\left(c_{j+1}-c_j\right)}{L} \cdot \frac{\left(c_j-c_{j-1}\right)}{L}} \\ =& \frac{l^*}{l^0} \cdot \frac{\left(c_{j+1}-c_j^*\right)\left(c_j^*-c_{j-1}\right)}{\left(c_{j+1}-c_j\right)\left(c_j-c_{j-1}\right)}. \end{aligned} \tag{D.18}$$

Thus, the acceptance probability is:

$$\alpha_c = \min\{1, p_c\}. \tag{D.19}$$

Type (a): Adding a discontinuity point

For (a), moving from model k to model $k+1$, which involves adding a discontinuity point. This transition increases the dimensionality by 2 (a new discontinuity point and a new velocity). The steps are:

(1) Draw c^* from the uniform distribution $U[0, L]$, which falls in a interval $[c_j, c_{j+1}]$ with velocity v_j.
(2) Draw μ from $U[0, 1]$ and determine v_j^* and v_{j+1}^* using the equations:

$$\begin{aligned} (c^* - c_j)\log(v_j^*) + (c_{j+1} - c^*)\log(v_{j+1}) &= (c_{j+1} - c_j)\log(v_j) \\ \frac{v_{j+1}^*}{v_j^*} &= \frac{1-\mu}{\mu}. \end{aligned} \tag{D.20}$$

(3) Propose the parameter

$$\theta^* = (k+1, v_1, v_2, \cdots, v_{j-1}, v_j^*, v_{j+1}^*, v_{j+1}, \cdots, v_{k+1}, c_1, c_2, \cdots, c_j, c_j^*, c_{j+1}, \cdots, c_k).$$

(4) Accept or reject the proposed parameter θ^* based on the acceptance probability.

The prior distribution of the k discontinuity points is given by the even-order statistics among $2k+1$ uniformly distributed points:

$$\begin{aligned} & f_{2,4,\cdots,2k}\left(\xi_2 = c_1, \xi_4 = c_2, \cdots, \xi_{2j} = c_j, \cdots, \xi_{2k+2} = c_k\right) \\ =& (2k+1)! \cdot \frac{c_1}{L} \cdot \frac{\left(c_2 - c_1\right)}{L} \cdot \ldots \cdot \frac{\left(c_{j+1} - c_j\right)}{L} \cdot \ldots \cdot \frac{\left(L - c_k\right)}{L} \cdot \left(\frac{1}{L}\right)^k \\ \triangleq & f_{c,k}, \end{aligned} \tag{D.21}$$

where F is the cumulative distribution function of the uniform distribution.

For the $k+1$ discontinuity points, the joint distribution is similarly defined using the even-order statistics of the $2k+3$ points:

$$\begin{aligned} & f_{2,4,\cdots,2k,2k+2}\left(\xi_2 = c_1, \xi_4 = c_2, \cdots, \xi_{2j} = c^*, \xi_{2j+2} = c_j, \cdots, \xi_{2k+2} = c_k\right) \\ =& (2k+3)! \cdot \frac{c_1}{L} \cdot \frac{c_2 - c_1}{L} \cdot \ldots \cdot \frac{\left(c^* - c_j\right)}{L} \cdot \frac{\left(c_{j+1} - c^*\right)}{L} \cdot \ldots \cdot \frac{L - c_k}{L} \cdot \left(\frac{1}{L}\right)^{k+1} \\ \triangleq & f_{c,k+1}. \end{aligned} \tag{D.22}$$

Thus, the ratio of the prior distributions for the discontinuity points is:

$$\frac{f_{c,k+1}}{f_{c,k}} = \frac{(2k+2)(2k+3)}{L^2} \cdot \frac{\left(c^* - c_j\right)\left(c_{j+1} - c^*\right)}{\left(c_{j+1} - c_j\right)}. \tag{D.23}$$

For the new velocities v_j^* and v_{j+1}^*, draw μ from $U[0,1]$.
The prior distribution ratio for the velocities is:

$$\begin{aligned}\frac{f_{v,k+1}}{f_{v,k}} &= \frac{\beta^\alpha v_j^{*\alpha-1} e^{-\beta v_j^*}/\Gamma(\alpha) \cdot \beta^\alpha v_{j+1}^{*\alpha-1} e^{-\beta v_{j+1}^*}/\Gamma(\alpha)}{\beta^\alpha v_j^{\alpha-1} e^{-\beta v_j}/\Gamma(\alpha)} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{v_j^* v_{j+1}^*}{v_j}\right)^{\alpha-1} \exp\left\{-\beta\left(v_j^* + v_{j+1}^* - v_j\right)\right\}.\end{aligned} \tag{D.24}$$

The Jacobian Determinant is:

$$|J| = \frac{\left(v_j^* + v_{j+1}^*\right)^2}{v_j} \tag{D.25}$$

The proposal distribution ratio is:

$$\frac{f_{p,k+1}}{f_{p,k}} = \frac{d_{k+1} \cdot \frac{1}{k+1}}{b_k \cdot \frac{1}{L}} = \frac{d_{k+1} \cdot L}{b_k \cdot (k+1)} \tag{D.26}$$

Let

$$p_b = \frac{l^*}{l^0} \cdot \frac{\rho(k+1)}{\rho(k)} \cdot \frac{f_{c,k+1}}{f_{c,k}} \cdot \frac{f_{v,k+1}}{f_{v,k}} \cdot \frac{f_{p,k+1}}{f_{p,k}} \cdot |J| \tag{D.27}$$

Then, the acceptance probability is:

$$\alpha_{birth} = \min\left\{1, p_b\right\}. \tag{D.28}$$

Type (b): Removing a discontinuity point

To maintain symmetry with Type (a), we derive the formulas for transitioning from model $k+1$ to model k, which involves removing a discontinuity point and reducing the dimensionality by 2.

The steps for proposing a new parameter when moving from model $k+1$ to model k are:

(1) Select one of the $k+1$ discontinuity points with probability $\frac{1}{k+1}$ to remove. Suppose the $(j+1)$th discontinuity point c_{j+1} is removed.
(2) Calculate the velocity parameter v_j^* for the merged interval $[c_j^*, c_{j+1}^*] = [c_j, c_{j+2}]$ using:

$$(c_{j+1} - c_j)\log(v_j) + (c_{j+2} - c_{j+1})\log(v_{j+1}) = (c_{j+1}^* - c_j^*)\log(v_j^*) \tag{D.29}$$

(3) Propose the parameter $\theta^* = (k, v_1, v_2, \cdots, v_{j-1}, v_j^*, v_{j+2}, \cdots, v_{k+1}, c_1, c_2, \cdots, c_j, c_{j+2}, \cdots, c_k)$.
(4) Accept or reject the proposed parameter θ^* with probability:

$$\alpha_{death} = \min\left\{1, \frac{1}{p_b}\right\}. \tag{D.30}$$

Note: the $(k+1)$th discontinuity point removed corresponds exactly to the added discontinuity point in Type (a).

Acknowledgements

The first author is supported by China Scholarship Council (Grant No. 202006030371) and the National Natural Science Foundation of China (Grant No. 12131001); The second author is financially supported by the National Natural Science Foundation of China (Grant Nos. 11971302 and 12171307).

References

Albarakati, A., Budišić, M., Crocker, R., Glass-Klaiber, J., Iams, S., Maclean, J., Marshall, N., Roberts, C., Van Vleck, E.S., 2022. Model and data reduction for data assimilation: particle filters employing projected forecasts and data with application to a shallow water model. Comput. Math. Appl. 116, 194–211.
Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle markov chain Monte Carlo methods. J. R. Stat. Soc. Series B (Stat. Methodol.) 72 (3), 269–342.
Bach, E., Ghil, M., 2023. A multi-model ensemble Kalman filter for data assimilation and forecasting. J. Adv. Model. Earth Syst. 15 (1), e2022MS003123.
Barker, R.J., Link, W.A., 2013. Bayesian multimodel inference by RJMCMC: a gibbs sampling approach. Am. Stat. 67 (3), 150–156.
Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. Vol. 4. Springer, New York.
Carlin, B.P., Louis, T.A., 2008. Bayesian Methods for Data Analysis. CRC press, New York.
Chang, H., Zhang, D., 2019. Identification of physical processes via combined data-driven and data-assimilation methods. J. Comput. Phys. 393, 337–350.
Corenflos, A., Thornton, J., Deligiannidis, G., Doucet, A., 2021. Differentiable particle filtering via entropy-regularized optimal transport. In: International Conference on Machine Learning. PMLR, pp. 2100–2111.

Doucet, A., De Freitas, N., Gordon, N.J., et al., 2001. Sequential Monte Carlo Methods in Practice. Vol. 1. Springer, New York.
Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput. 10, 197–208.
Elfring, J., Torta, E., van de Molengraft, R., 2021. Particle filters: a hands-on tutorial. Sensors 21 (2), 438.
Elvira, V., Martino, L., Bugallo, M.F., Djurić, P.M., 2018. In search for improved auxiliary particle filters. In: 2018 26th European Signal Processing Conference (EUSIPCO). IEEE, pp. 1637–1641.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans 99 (C5), 10143–10162.
Evensen, G., Vossepoel, F.C., Van Leeuwen, P.J., 2022. Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem. Springer Nature, Cham.
Feng, J., Wang, L., 2024. A switching state-space transmission model for tracking epidemics and assessing interventions. Comput. Stat. Data Anal. 197, 107977.
Gelfand, A.E., 2012. Hierarchical modeling for spatial data problems. Spat. Stat. 1, 30–39.
Gilks, W.R., Berzuini, C., 2001. Following a moving target–Monte Carlo inference for dynamic Bayesian models. J. R. Stat. Soc. Series B (Stat. Methodol.) 63 (1), 127–146.
Godsill, S., 2019. Particle filtering: the first 25 years and beyond. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 7760–7764.
Gordon, N.J., Salmond, D.J., Smith, A. F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEE Proceedings F (Radar and Signal Processing). Vol. 140. IET, pp. 107–113.
Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.
Green, P.J., Hastie, D.I., 2009. Reversible jump MCMC. Genetics 155 (3), 1391–1403.
Herrera, S., Fernández, J., Rodríguez, M.A., Gutiérrez, J.M., 2010. Spatio-temporal error growth in the multi-scale Lorenz'96 model. Nonlinear Process. Geophys. 17 (4), 329–337.
Huan, Y., Lin, H.X., 2024. Sequential model identification with reversible jump ensemble data assimilation method. Stat. Comput. 34 (6), 184.
Katzfuss, M., Stroud, J.R., Wikle, C.K., 2020. Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. J. Am. Stat. Assoc. 115 (530), 866–885.
Lewis, J.M., Lakshmivarahan, S., Dhall, S., 2006. Dynamic Data Assimilation: A Least Squares Approach. Vol. 13. Cambridge University Press.
Lorenz, E.N., 1996. Predictability: a problem partly solved. In: Proc. Seminar on Predictability. Vol. 1. Reading, pp. 1–18.
Moradkhani, H., DeChant, C.M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. Water Resour. Res. 48 (12). https://doi.org/10.1029/2012WR012144
Rouder, J.N., Lu, J., 2005. An introduction to Bayesian hierarchical models with an application in the theory of signal detection. Psychonomic Bull. Rev. 12 (4), 573–604.
Snyder, C., Bengtsson, T., Bickel, P., Anderson, J., 2008. Obstacles to high-dimensional particle filtering. Mon. Weather Rev. 136 (12), 4629–4640.
Van Kekem, D.L., Sterk, A.E., 2018. Wave propagation in the Lrenz-96 model. Nonlinear Process. Geophys. 25 (2), 301–314.
Van Leeuwen, P.J., 2009. Particle filtering in geophysical systems. Mon. Weather Rev. 137 (12), 4089–4114.
Vetra-Carvalho, S., Van Leeuwen, P.J., Nerger, L., Barth, A., Altaf, M.U., Brasseur, P., Kirchgessner, P., Beckers, J.-M., 2018. State-of-the-art stochastic data assimilation methods for high-dimensional non-gaussian problems. Tellus A Dyn. Meteorol. Oceanog. 70 (1), 1–43.
Vrugt, J.A., ter Braak, C. J.F., Diks, C. G.H., Schoups, G., 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: theory, concepts and applications. Adv. Water Resour. 51, 457–478. https://doi.org/10.1016/j.advwatres.2012.04.002
Wiese, T., Rosca, J., Claussen, H., 2015. Reversible jump particle filter (RJPF) for wideband DOA tracking. Excursions in Harmonic Anal. Vol. 3 The February Fourier Talks Norbert Wien. Center, 231–261.
Zhu, G., Li, X., Zhang, K., Ding, Z., Han, T., Ma, J., Huang, C., He, J., Ma, T., 2016. Multi-model ensemble prediction of terrestrial evapotranspiration across North China using Bayesian model averaging. Hydrol. Process. 30 (16), 2861–2879.