Advanced Bayesian approaches for state-space models with a case study on soil carbon sequestration

Mohammad Javad Davoudabadi*[1,2,3], Daniel Pagendam[4], Christopher Drovandi[1,2,3], Jeff Baldock[5], and Gentry White[1,2,3]

[1]*School of Mathematical Sciences, Queensland University of Technology, Australia;*
[2]*Australian Research Council Centre of Excellence for Mathematical & Statistical Frontiers (ACEMS);*
[3]*QUT Centre for Data Science, Queensland University of Technology, Australia;*
[4]*CSIRO Data61, GPO Box 2583, Brisbane, QLD 4001, Australia;*
[5]*CSIRO Agriculture & Food, Glen Osmond, South Australia, Australia;*

Abstract

Sequestering carbon into the soil can mitigate the atmospheric concentration of greenhouse gases, improving crop productivity and yield financial gains for farmers through the sale of carbon credits. In this work, we develop and evaluate advanced Bayesian methods for modelling soil carbon sequestration and quantifying uncertainty around predictions that are needed to fit more complex soil carbon models, such as multiple-pool soil carbon dynamic models. This paper demonstrates efficient computational methods using a one-pool model of the soil carbon dynamics previously used to predict soil carbon stock change under different agricultural practices applied at Tarlee, South Australia. We focus on methods that can improve the speed of computation when estimating parameters and model state variables in a statistically defensible way. This paper also serves as a tutorial on advanced Bayesian methods for fitting complex state-space models, which will be of interest to soil scientists and other environmental scientists more generally.

Keywords: Soil carbon sequestration; State-space model; Rao-Blackwellised particle filter; Correlated pseudo-marginal method.

*mohammadjavad.davoudabadi@hdr.qut.edu.au

1 Introduction

Global carbon cycling describes the exchange of carbon between and within five main pools: earth's oceans, its atmosphere, vegetation, fossil fuels, and its soil. Increases in atmospheric CO_2 and other greenhouse gases are considered responsible for global climate change. Soils store approximately fourfold more carbon than the atmosphere and sixfold more than earth's vegetation (Ahrens et al., 2014). Therefore, it is important that we have reliable statistical approaches for monitoring this fundamentally important carbon pool. Large-scale transfer of carbon from earth's soil to the atmosphere not only affects global climate change, but also has the potential to globally reduce soil quality and agricultural productivity (Sanderman and Chappell, 2013). Accurate models of soil carbon dynamics and sequestration for forecasting future changes are useful tools to help understand these systems and potentially maintain or increase soil carbon stocks.

Soil carbon sequestration, has many benefits: reducing the atmospheric concentration of greenhouse gases, improving agricultural productivity, and yielding financial gains for farmers. There have been numerous studies in which researchers have used models to predict soil organic carbon change (Cable et al., 2009; Clifford et al., 2014; Huang et al., 2017; Luo et al., 2016; Skjemstad et al., 2004; Stamati et al., 2013). Challenges in modelling the soil carbon cycle include our imperfect understanding of the carbon cycle and sequestration process, and that the volume and quality of data are often limited. Assessing uncertainty in models is important because it affects parameter estimates, model inputs, and ultimately, model predictions. A number of studies have been conducted into quantifying uncertainty in soil carbon model outputs, typically under simulated climate change or via a sensitivity analysis by running models for different sets of parameter values (Juston et al., 2010; Paul et al., 2003; Stamati et al., 2013). More recently, uncertainty quantification in model inputs, model parameters, and dynamics has been conducted by Clifford et al. (2014).

Quantifying trends in soil organic carbon (SOC) stocks sometimes needs measurements over more than a decade because of slow rates of temporal change as well as temporal variability in flows of carbon into and out of the soil introduced by environmental factors and management practices (Sanderman and Baldock, 2010). Because such data can be lacking, computer-simulation experiments performed with models are used to explore temporal dynamics of soil carbon and the impact of management practice *in silico*. Such models include RothC (Parton et al., 1988), and Century (Jenkinson et al., 1987) which were developed in response to this requirement of a

long-term sampling period. These models have been used to estimate changes in soil carbon pools derived by variations in climatic conditions (Senapati et al., 2013; Wan et al., 2011) or variations in applied agricultural management practices (Li Liu et al., 2009, 2011; Robertson and Nash, 2013). It is noteworthy that these models differentiate soil carbon into a series of conceptual pools. The carbon contained within a pool is considered homogeneous in its physical and chemical attributes and thus process dynamics; however, across pools, significant differences in these traits exist (Elliott et al., 1996).

Some effort has been made to quantify uncertainties in modelled carbon stocks using statistical models, embodying a set of statistical assumptions about the data, as a formal approach to fitting the model and estimating uncertainty (Jones et al., 2007; Koo et al., 2007; Post et al., 2008). Clifford et al. (2014), addressed uncertainties using a physical-statistical model for carbon dynamics within a Bayesian hierarchical modelling (BHM) framework, and demonstrated this approach for a one-pool model. The one-pool model of Clifford et al. (2014) treats the soil carbon as a single pool of carbon undergoing exponential decay with carbon being lost to the atmosphere. Since the carbon stock cannot be observed directly (i.e. without error), it can be considered a latent variable that evolves with time as a Markov chain, meaning that it can be modelled as a state-space model. State-space modelling allows us to estimate unobserved state variables, unknown parameters, and future observations of the model from observed data. In Clifford et al. (2014), this model was fit using a particle Markov chain Monte Carlo (pMCMC, (Andrieu et al., 2010)) algorithm. They used the most standard pMCMC algorithm in the literature. More specifically, it used the bootstrap particle filter for estimating the likelihood for a given parameter value, and used the particle marginal Metropolis-Hastings (PMMH) algorithm to explore the parameter space.

In this paper, we explore how the modelling approach employed by Clifford et al. (2014) can be implemented with greater computational efficiency (improve the speed of computation) using more advanced Bayesian algorithms for state-space models. These advanced Bayesian methods can play a significant role in fitting more complex soil carbon models, such as multiple-pool soil carbon dynamic models. Here we will use the one-pool model of Clifford et al. (2014) to illustrate the methods and their benefit. We demonstrate that these advanced methods lead to substantial computational efficiency for model fitting and parameter estimation compared to the methods employed in Clifford et al. (2014). This is important as it demonstrates that such methods have the potential to be used as part of decision-support systems, to rapidly obtain insights using standard computer hardware. In addition to the aforementioned aim, this paper can also be considered as a

tutorial on more advanced Bayesian state-space model fitting methods for soil scientists and other environmental scientists working with similar classes of models.

The rest of the paper is organised as follows. In Section 2, the state-space model as well as the Kalman filter and particle filters are introduced. The PMMH algorithms and the implementation of the aforementioned methods on the model are presented in Section 3. Section 4 defines the soil carbon model of Clifford et al. (2014), which motivates the paper. The results are shown in Section 5 with a focus on quantities that may be of interest to a scientist or landowner interested in soil carbon sequestration. The paper concludes with a discussion in Section 6.

2 State-space model

The state-space model uses indirectly observable variables known as state (or latent) variables and observable measurement variables to describe a system. These state variables cannot be measured directly but can be estimated based on observational data that depend on the state variables (Andrieu et al., 2010; Fearnhead and Künsch, 2018). The values of state variables evolve through time in a way that depends on the past state variable values and external input variables. The main goal of state-space modelling is to gain knowledge of the latent states $\mathbf{X}$ given the observations $\mathbf{Y}$. Therefore, the aim is producing estimators for the state variable $X_{(t)}$, given the observed data $Y_{1:s} = \{Y_{(1)}, \ldots, Y_{(s)}\}$. In this case, if $s < t$, the problem is called prediction or forecasting, if $s = t$, the problem is known as filtering, and the problem is called smoothing when $s > t$ (Shumway and Stoffer, 2000, Chapter 6). In addition to producing these point estimates, measuring their precision is also important. This can be accomplished using approaches such as the Kalman filter and particle filters.

A generic representation of a state-space model with Gaussian noise is:

$$\begin{aligned} X_{(t)} &= f(X_{(t-1)}) + \boldsymbol{B}u_{(t)} + \epsilon_{(t)} \\ Y_{(t)} &= g(X_{(t)}) + \nu_{(t)}; \end{aligned} \tag{1}$$

where $\epsilon_{(t)} \sim N(\mu, \sigma_X^2)$ and $\nu_{(t)} \sim N(\lambda, \sigma_Y^2)$ are state and measurement noise components, respectively (Shumway and Stoffer, 2000, Chapter 6). The control-input matrix $\boldsymbol{B}$ is applied to a known vector of inputs $u_{(t)}$, also f and g are functions of $X_{(t-1)}$ and $X_{(t)}$, respectively. The first equation is called the state transition equation (state model) and the second one is the observation equation

(observation or measurement model) (Fearnhead and Künsch, 2018). In general, $\epsilon_{(t)}$ and $\nu_{(t)}$ need not be Gaussian nor additive, but in many models this is assumed to be the case (Andrieu et al., 2010). A soil carbon dynamics model can be considered as an example of a state-space model. The observation model in this setting includes the measured values of SOC at time t (i.e. $Y_{(t)}$) conditioning on unobserved masses of SOC in various pools and the process model includes unobserved masses of SOC (i.e. $X_{(t)}$) in various pools conditioning on its past value. These models can be formed from some sub-models including linear and non-linear sub-models. The general state-space model can also be represented graphically as depicted in Figure 1.

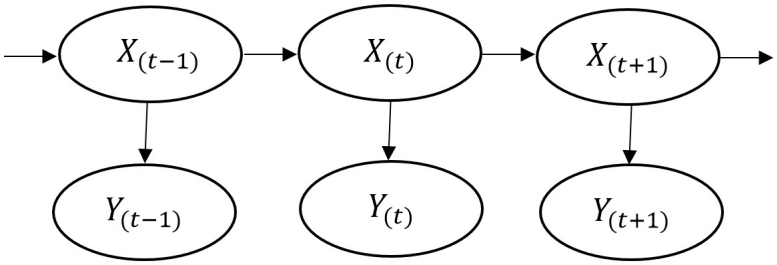

Figure 1: Graphical representation of a state-space model.

To calculate the filtering distribution of $X_{(t)}$ (i.e. $p(X_{(t)}|Y_{1:t})$) we can use Bayes rule as shown in (2)

$$p(X_{(t)}|Y_{1:t}) \propto p(Y_{(t)}|X_{(t)}) \int p(X_{(t)}|X_{(t-1)})p(X_{(t-1)}|Y_{1:t-1})dX_{(t-1)}; \tag{2}$$

where $p(X_{(t)}|X_{(t-1)})$ is the transition density and $p(Y_{(t)}|X_{(t)})$ is the observation density (Andrieu et al., 2010). When $p(X_{(t)}|Y_{1:t})$ and the right side of (2) do not have closed form expressions, some methods such as particle filters (PFs) can be used to approximate them. When the model is a linear-Gaussian, the Kalman filter provides closed-form solution results. In the next section we introduce the Kalman filter. It is worth noting that the state-space model shown in equation (1) often depends on an unknown static parameter $\boldsymbol{\theta}$, as is the case for our application. For the remainder of this section, we assume the static parameter $\boldsymbol{\theta}$ is fixed, and thus drop it for notational convenience. In Section 3, we re-introduce $\boldsymbol{\theta}$ into the notation.

2.1 Kalman filter

Filtering, also known as data assimilation in geophysics, is an engineering term which is used to extract information of a signal from noisy and partial observations and that is mostly undertaken using state-space models. It was conceived by Rudolf Emil Kalman in the 1960s (Kalman, 1960), who used the word *filter*, since it is a process of filtering out the noise from noisy data to find the best estimate. The method is a recursive filtering algorithm for linear-Gaussian state-space models that allows new measurements to be processed as they arrive, making it suitable for online applications (Künsch, 2013). There are many methods for evaluating state-space models and the most common is the Kalman filter. The KF is an optimal estimator in the sense of minimising the variance of the estimated states.

The main advantage of this filter is that it provides closed-form solution results for linear-Gaussian models, which occurs when the general model in (1) has the form:

$$\begin{aligned} X_{(t)} &= \boldsymbol{A}X_{(t-1)} + \boldsymbol{B}u_{(t)} + \epsilon^*_{(t)} \\ Y_{(t)} &= \boldsymbol{C}X_{(t)} + \nu^*_{(t)}; \end{aligned}$$

where $\epsilon^*_{(t)} \sim N(\mathbf{0}, \boldsymbol{Q}^*)$, $\nu^*_{(t)} \sim N(\mathbf{0}, \boldsymbol{R}^*)$, $\boldsymbol{A}$ is the state-transition matrix, the control-input matrix $\boldsymbol{B}$ is applied to a known vector of inputs $u_{(t)}$, and $\boldsymbol{C}$ is the observation matrix (Carter and Kohn, 1994; Orlande et al., 2011; Wikle and Berliner, 2007). The Kalman filter includes two steps: prediction and update (Orlande et al., 2011). The prediction step is:

$$\begin{aligned} X^{t-1}_{(t)} &= \boldsymbol{A}\hat{X}_{(t-1)} + \boldsymbol{B}u_{(t)} \\ \boldsymbol{P}^{t-1}_{(t)} &= \boldsymbol{A}\hat{\boldsymbol{P}}_{(t-1)}\boldsymbol{A}^{'} + \boldsymbol{Q}^*; \end{aligned}$$

where $\boldsymbol{A}^{'}$ is the transpose of matrix $\boldsymbol{A}$, and the update step is:

$$\begin{aligned} \boldsymbol{K}_{(t)} &= \boldsymbol{P}^{t-1}_{(t)}\boldsymbol{C}^{'}(\boldsymbol{C}\boldsymbol{P}^{t-1}_{(t)}\boldsymbol{C}^{'} + \boldsymbol{R}^*)^{-1} \\ \hat{X}_{(t)} &= X^{t-1}_{(t)} + \boldsymbol{K}_{(t)}(Y_{(t)} - \boldsymbol{C}X^{t-1}_{(t)}) \\ \hat{\boldsymbol{P}}_{(t)} &= (\boldsymbol{I} - \boldsymbol{K}_{(t)}\boldsymbol{C})\boldsymbol{P}^{t-1}_{(t)}. \end{aligned}$$

The difference $(Y_{(t)} - \boldsymbol{C}X^{t-1}_{(t)})$ is called the residual and $\boldsymbol{I}$ is the identity matrix of appropriate dimension. Matrix $\boldsymbol{K}_{(t)}$ is called the Kalman gain, $\boldsymbol{R}^*$ and $\hat{\boldsymbol{P}}_{(t)}$ are the covariance matrix of the

measurement noise and the updated estimate covariance matrix of the process noise, respectively (Cressie and Wikle, 2015; Orlande et al., 2011). $\hat{X}_{(t)}$ is our best estimate of the state vector at time t given all observations up to and including time t, and $X_{(t)}^{t-1}$ and $\boldsymbol{P}_{(t)}^{t-1}$ are the expectations of state variable and the process noise, respectively given all observations up to and including time $t-1$. The KF method is presented in Algorithm 1, here MVN$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Algorithm 1 Kalman filter (KF) algorithm

1: Initialize with initial state $\hat{X}_{(0)} = x_{(0)}$ and $\hat{\boldsymbol{P}}_{(0)} = \boldsymbol{Q}^*$ at $t = 0$;
2: **for** $t = 1, ..., T$ **do**
3: $X_{(t)}^{t-1} = \boldsymbol{A}\hat{X}_{(t-1)} + \boldsymbol{B}u_{(t)}$, State estimate extrapolation;
4: $\boldsymbol{P}_{(t)}^{t-1} = \boldsymbol{A}\hat{\boldsymbol{P}}_{(t-1)}\boldsymbol{A}^{'} + \boldsymbol{Q}^*$, State covariance extrapolation;
5: $\boldsymbol{K}_{(t)} = \boldsymbol{P}_{(t)}^{t-1}\boldsymbol{C}^{'}[\boldsymbol{R}^* + \boldsymbol{C}\boldsymbol{P}_{(t)}^{t-1}\boldsymbol{C}^{'}]^{-1}$, Kalman gain matrix;
6: $\hat{X}_{(t)} = X_{(t)}^{t-1} + \boldsymbol{K}_{(t)}[Y_{(t)} - \boldsymbol{C}X_{(t)}^{t-1}]$, State estimate update;
7: $\hat{\boldsymbol{P}}_{(t)} = [\boldsymbol{I} - \boldsymbol{K}_{(t)}\boldsymbol{C}]\boldsymbol{P}_{(t)}^{t-1}$, State covariance update;
8: Compute the log-likelihood contribution, $l_{(t)}^{\text{KF}}$, at time t through the density MVN$(Y_{(t)} - \boldsymbol{C}X_{(t)}^{t-1}, \boldsymbol{R}^* + \boldsymbol{C}\boldsymbol{P}_{(t)}^{t-1}\boldsymbol{C}^{'})$;
9: The complete log-likelihood can be calculated as $L = \sum_t l_{(t)}^{\text{KF}}$

In non-linear cases, the extended Kalman filter (EKF) can be applied to transform an estimate of the current mean and covariance into a linear form. It is not an optimal estimator as it tends to underestimate the true value of the covariance matrix and risks becoming inconsistent in the statistical sense without the addition of "stabilising noise", so it is only reliable for systems that are almost linear (Huang et al., 2008; Ljung, 1979). An improved form of the EKF is known as the unscented Kalman filter (UKF) that picks minimal sample points around the mean by using a deterministic sampling technique and propagating the sample points through the non-linear functions so that a new mean and covariance estimate are formed. The UKF also has to apply linearisation, which illustrates the difficulty in implementing any type of Kalman filter for non-linear state transitions (Grewal, 2011). More generally, particle filters, introduced in the next subsection, can be used for non-linear and/or non-Gaussian state-space models, without making any approximation to the model.

2.2 Particle filters

Particle filter methods also known as sequential Monte Carlo (SMC) methods are a class of Monte Carlo methods used broadly in Bayesian inference and signal processing (Fearnhead and Künsch,

2018; Liu and Chen, 1998). A PF provides a particle approximation of the sequence of filtering distributions, $p(X_{(t)}|Y_{1:t})$ for $t = 1, \ldots, T$. As a by-product, it also produces an unbiased estimate of the marginal density of the observed data, $p(Y_{1:T})$ (Fearnhead and Künsch, 2018). This term is often referred to as the marginal likelihood, as the latent states $X_{1:T}$ have been integrated out. However, we refer to it simply as the likelihood here, since as mentioned earlier, it is conditional on a fixed value of the static parameter.

Assume that we have a particle representation of size N of the filtering distribution at time $t-1$, $p(X_{(t-1)}|Y_{1:t-1})$ at time $t-1$:

$$\hat{p}(X_{(t-1)}|Y_{1:t-1}) = \sum_{k=1}^{N} W_{(t-1)}^{k} \delta_{X_{(t-1)}^{k}}(X_{(t-1)}),$$

where $W_{(t-1)}^{k}$ and $X_{(t-1)}^{k}$ are the normalised weight and location of particle k, respectively, and $\delta_{X_{(0)}}(\cdot)$ denotes the Dirac delta mass located at $X_{(0)}$ (Andrieu et al., 2010; Fearnhead and Künsch, 2018). The particle filter provides a mechanism for updating this empirical approximation so that it represents the next filtering distribution $p(X_{(t)}|Y_{1:t})$ in light of the new observation $Y_{(t)}$.

Different particle filters are distinguished by how the empirical approximation is updated. There are now many sophisticated particle filters in the literature, such as the auxiliary particle filter (Pitt and Shephard, 1999), Rao-Blackwellised particle filter (Doucet et al., 2000) and Gaussian particle filter (Kotecha and Djuric, 2003). However, the bootstrap particle filter of (Gordon et al., 1993), discussed next, has the simplest form.

2.2.1 Bootstrap particle filter

In the sampling importance resampling (SIR) algorithm, the particle values for time t are proposed according to a user-specified proposal distribution, $X_{(t)} \sim q(\cdot|X_{(t-1)})$. The updated particle values are then re-weighted using arguments from importance sampling

$$w_{(t)}^{k} = W_{(t-1)}^{k} \frac{p(Y_{(t)}|X_{(t)})p(X_{(t)}|X_{(t-1)})}{q(X_{(t)}|X_{(t-1)})}, \quad t = 1, \ldots, T, \text{ and } \quad k = 1, \ldots, N; \tag{3}$$

where $w_{(t)}^{k}$ are unnormalised weights at time t and $W_{(t)}^{k} = \frac{w_{(t)}^{k}}{\sum_j w_{(t)}^{j}}$ (Doucet and Johansen, 2009; Fearnhead and Künsch, 2018). After normalisation, a weighted sample from $p(X_{(t)}|Y_{1:t})$ is produced. These steps can be repeated until all data have been processed. However, the particle weights have a tendency to become degenerate; eventually placing all weight on a single particle,

resulting in a poor empirical approximation of the filtering distribution (Fearnhead and Künsch, 2018). This problem can be mitigated by including a resampling step at each iteration. This step selects N particle values randomly with replacement from the current particle set, where the probability of selecting a particular particle is given by its corresponding normalised weight (Andrieu et al., 2010; Doucet and Johansen, 2009; Fearnhead and Künsch, 2018). The resampling step has the effect of replicating particle values with high weight, and terminating those with low weight. After resampling, the normalised weights are re-initialised at $1/N$ for each particle (Fearnhead and Künsch, 2018). If the variance of unnormalised weights of particles at a specific time is small, the resampling step at that time might be unnecessary, as it introduces some additional variance into the estimator (Doucet and Johansen, 2009). More precisely, this is often assessed by looking at the effective sample size (ESS) criterion. The ESS provides a means of comparison between our correlated samples and a set of independently drawn samples. This is given by

$$ESS = \left(\sum_{j=1}^{N} (W_{(t)}^{j})^2 \right)^{-1}.$$

We resample only when it is below a pre-specified threshold (typically $\frac{N}{2}$) (Doucet and Johansen, 2009). When the resampling step is omitted in each time step, we are required to store the normalised weights as these are an important quantity on the right-hand side of equation (3). When the resampling step is used at every time step, the $W_{(t-1)}^{k}$ on the right-hand side of (3) plays no role in the importance sampling re-weighting since they remain uniform throughout the algorithm.

The bootstrap particle filter (BPF) is a special case of the SIR, where the transition density is selected as the proposal density, $q(X_{(t)}|X_{(t-1)}) = p(X_{(t)}|X_{(t-1)})$. With this choice, the importance weights simplify to:

$$w_{(t)}^{k} = W_{(t-1)}^{k} p(Y_{(t)}|X_{(t)}), \quad t = 1, \ldots, T, \text{ and } \quad k = 1, \ldots, N.$$

The BPF algorithm is provided in Algorithm 2. In this algorithm, since the resampling step is applied at each time step, the normalised weights at $t-1$ are omitted from the weight update at t. The appealing features of this PF are that it is simple to implement and widely applicable, since it does not require evaluation of the transition density $p(X_{(t)}|X_{(t-1)})$, but simply the ability to simulate from it.

However, a drawback of the bootstrap particle filter is that, since the proposal is the transition density, the propagated particles $X_{(t)}^k$, $k = 1, \ldots, N$ may not be compatible with the next observation $Y_{(t)}$. This can lead to the particle system degenerating quickly (Künsch, 2013). One way to overcome this is to use a larger number of particles N, but this can be computationally burdensome. In the next subsection, we describe the auxiliary particle filter, which explicitly accounts for the next observation $Y_{(t)}$ when making proposals.

Algorithm 2 Bootstrap particle filter (BPF) algorithm

1: Initialise $X_{(0)}^{*k} \sim p(X_{(0)})$ for $k \in \{1, 2, \ldots, N\}$;
2: **for** $t = 1, ..., T$ **do**
3: **for** $k = 1, ..., N$ **do**
4: Draw sample $X_{(t)}^k \sim p(X_{(t)}|X_{(t-1)}^{*k})$;
5: Calculate weights $w_{(t)}^k = p(Y_{(t)}|X_{(t)}^k)$;
6: Estimate the log-likelihood component for the t^{th} observation, $\hat{l}_{(t)} = \log\left(\frac{\sum_j w_{(t)}^j}{N}\right)$;
7: Normalise weights $W_{(t)}^k = \frac{w_{(t)}^k}{\sum_j w_{(t)}^j}$ for $k \in \{1, 2, \ldots, N\}$;
8: Resample with replacement N particles, $X_{(t)}^{*k}$ ($k \in \{1, 2, \ldots, N\}$ and $X_{(t)}^{*k} \in \{X_{(t)}^k\}_{k=1}^N$), based on the normalised importance weights;
9: Estimate the overall log-likelihood $L = \sum_t \hat{l}_{(t)}$.

2.2.2 Auxiliary particle filter

In the presence of outliers or small variances of observation errors, the BPF can be very inefficient and the weights will be extremely unequally distributed. Therefore, a huge number of particles can be required to gain sensible estimation results (Pitt and Shephard, 1999). The compatibility between propagated particles at time t and $Y_{(t)}$ can be improved as well as having more balanced-weights through an advanced particle filter, called auxiliary particle filter (APF).

The APF is an advanced particle filter technique for state-space models which was introduced by Pitt and Shephard (1999). This particle filter was proposed to improve the performance of the BPF when the measurement density is informative relative to the process density (Pitt et al., 2010). To explain more clearly, before the resampling step, this particle filter employs knowledge about the current observation to improve the compatibility between particles and that observation through an auxiliary variable.

We first describe the ideal, but usually not implementable, version of the APF, called the fully

adapted APF. It relies on being able to compute the following terms

$$p(X_{(t)}|Y_{(t)}, X_{(t-1)}) = \frac{p(Y_{(t)}|X_{(t)})p(X_{(t)}|X_{(t-1)})}{p(Y_{(t)}|X_{(t-1)})};$$

where

$$p(Y_{(t)}|X_{(t-1)}) = \int p(Y_{(t)}|X_{(t)})p(X_{(t)}|X_{(t-1)})dX_{(t)};$$

$$p(Y_{(t)}|X_{(t)})p(X_{(t)}|X_{(t-1)}) = p(Y_{(t)}|X_{(t-1)})p(X_{(t)}|Y_{(t)}, X_{(t-1)}). \tag{4}$$

Assume that we have particles $X^k_{(t-1)}$, with corresponding weights $W^k_{(t-1)}$, $k = 1, \ldots, N$ at time $t-1$. To proceed to time t, a set of auxiliary indices $\gamma^k \in \{1, \ldots, N\}$ for $k = 1, \ldots, N$ are firstly drawn from a categorical distribution with probabilities proportional to $\lambda^k_{(t)} = W^k_{(t-1)}p(Y_{(t)}|X^k_{(t-1)})$ for $k = 1, \ldots, N$. These are often referred to as the first stage weights (Pitt and Shephard, 1999). The idea is to pre-weight the particles so that particles supported by the next observation $Y_{(t)}$ are given higher weight. Then, the states are simulated forward to time t using the exact conditional of $X_{(t)}$ given $Y_{(t)}$ and $X_{(t-1)}$, i.e. $X^k_{(t)} \sim p(X_{(t)}|Y_{(t)}, X^{\gamma^k}_{(t-1)})$ for $k = 1, \ldots, N$. It can be shown (see below) that this results in particle weights that remain constant throughout the algorithm, and thus the ESS remains at N after each iteration. The fully adapted APF is generally the optimal filter in terms of a single time step ahead (Pitt et al., 2010).

However, in complex models it is often not feasible to obtain an analytical expression for $p(Y_{(t)}|X_{(t-1)})$ and/or to simulate directly from $p(X_{(t)}|Y_{(t)}, X_{(t-1)})$. Therefore, the fully adapted APF can often not be implemented in practice, producing the typical APF algorithm. We assume that $p(Y_{(t)}|X_{(t-1)})$ can be approximated with some $q(Y_{(t)}|X_{(t-1)})$ (Johansen and Doucet, 2008). This is often selected to be $q(Y_{(t)}|X_{(t-1)}) = p(Y_{(t)}|\mu_{(t)})$, i.e. simply replacing $X_{(t-1)}$ with $\mu_{(t)}$ in the observation density where $\mu_{(t)}$ is given by some location measure of the transition distribution $X_{(t)}|X_{(t-1)}$ (e.g. mean) when such a location measure is analytically available (Pitt and Shephard, 1999). Further, we can simulate from some proposal $q(X_{(t)}|Y_{(t)}, X_{(t-1)})$ when $p(X_{(t)}|Y_{(t)}, X_{(t-1)})$ cannot be directly simulated. One simple example is by simulating according to the transition density, i.e. by setting $q(X_{(t)}|Y_{(t)}, X_{(t-1)}) = p(X_{(t)}|X_{(t-1)})$, but others are possible in the general APF framework.

By using importance sampling arguments, it can be shown that the second stage weights of

the APF are given by

$$w_{(t)}^{k} = \frac{p(Y_{(t)}|X_{(t)}^{k})p(X_{(t)}|X_{(t-1)}^{\gamma^k})}{q(Y_{(t)}|X_{(t-1)}^{\gamma^k})q(X_{(t)}|Y_{(t)}, X_{(t-1)}^{\gamma^k})}, \quad t = 1, \ldots, T, \text{ and } \quad k = 1, \ldots, N. \tag{5}$$

The method reduces to the fully adapted algorithm if the denominator in (5) equals equation (4) in that $w_{(t)}^{k} = 1$ and $W_{(t-1)}^{k} = \frac{1}{N}$. The APF algorithm is presented in Algorithm 3.

Algorithm 3 Auxiliary particle filter (APF) algorithm

for $k = 1, ..., N$ **do**
  Draw sample $X_{(1)}^{*k} \sim p(X_{(1)})$;
  $w_{(1)}^{k} = p(Y_{(1)}|X_{(1)}^{*k})$;
  Normalise weights $W_{(1)}^{k} = \dfrac{w_{(1)}^{k}}{\sum_j w_{(1)}^{j}}$;
for $t = 2, ..., T$ **do**
  for $k = 1, ..., N$ **do**
    Determine a suitable statistic, $\mu_{(t)}^{k}$, from $p(X_{(t)}|X_{(t-1)}^{*k})$ and the first stage weights $\lambda_{(t)}^{k} = p(Y_{(t)}|\mu_{(t)}^{k})W_{(t-1)}^{k}$;
    Draw indices γ^k from the set of indices $k = \{1, \ldots, N\}$ with normalised sampling probabilities $\lambda_{(t)}^{k}$;
    Draw sample $X_{(t)}^{*k} \sim q(X_{(t)}|Y_{(t)}, X_{(t-1)}^{\gamma^k})$;
    Calculate weights $w_{(t)}^{k} = \dfrac{p(Y_{(t)}|X_{(t)}^{*k})p(X_{(t)}|X_{(t-1)}^{\gamma^k})}{p(Y_{(t)}|\mu_{(t)}^{\gamma^k})q(X_{(t)}|Y_{(t)}, X_{(t-1)}^{\gamma^k})}$;
  Normalise weights $W_{(t)}^{k} = \dfrac{w_{(t)}^{k}}{\sum_j w_{(t)}^{j}}$ for $k = \{1, 2, \ldots, N\}$;
  Estimate the log-likelihood component for the t^{th} observation, $l_{(t)} = \log\left(\dfrac{\sum_j w_{(t)}^{j}}{N}\right) + \log(\sum_j \lambda_{(t)}^{j})$;
Estimate the overall log-likelihood $L = \sum_t l_{(t)}$.

There is a common misconception that the asymptotic variance of estimators based on the APF is always smaller than the variance of SIR-based estimators. However in contrary to this popular belief, not only can it be larger than those corresponding SIR-based estimators, but also it can perform worse than the BPF in the full adaptation case (Johansen and Doucet, 2008). The next section is allocated to describe a particle filter that is implemented on models which have a combination of linear and non-linear structures.

2.2.3 Rao-Blackwellised particle filters

Rao-Blackwellised particle filters (RBPFs) were introduced by Doucet et al. (2000) to improve the performance of particle filtering when part of the model is analytically tractable. Fundamentally, RBPFs use a particle filter for a few variables and marginalise out the rest using a filter such as the KF. This leads to more accurate estimates than the standard particle filter (Doucet et al., 2000). Suppose the state variable $X_{(t)}$ can be divided into two groups $X^{(1)}_{(t)}$ and $X^{(2)}_{(t)}$ so that the transition density can be rewritten as $p(X_{(t)}|X_{(t-1)}) = p(X^{(1)}_{(t)}|X^{(1)}_{(t-1)}, X^{(2)}_{(t-1)})p(X^{(2)}_{(t)}|X^{(2)}_{(t-1)})$. According to the chain rule, the filtering distribution decompose to the following:

$$p(X^{(1)}_{(t)}, X^{(2)}_{(t)}|Y_{1:t}) = p(X^{(1)}_{(t)}|Y_{1:t}, X^{(2)}_{(t)})p(X^{(2)}_{(t)}|Y_{1:t});$$

where $p(X^{(1)}_{(t)}|Y_{1:t}, X^{(2)}_{(t)})$ is analytically tractable and occurs where the model, conditional on $X^{(2)}_{(t)}$, is linear-Gaussian, hence, $p(X^{(1)}_{(t)}|Y_{1:t}, X^{(2)}_{(t)})$ is Gaussian and able to be calculated using the KF (Fearnhead and Künsch, 2018). The RBPF can be more efficient than the BPF as the particle filter is only required for a lower dimensional state-space and the remaining calculations can be done analytically with the KF. The RBPF algorithm is presented in Algorithm 4, where $l^{\text{KF}}_{(t)}$ denotes the log-likelihood computed by the KF at time t. As mentioned earlier in Section 2, a soil carbon dynamics model can be considered as a tangible example of a state-space model. If a soil carbon dynamics model is formed from non-linear and linear-Gaussian sub-models, the RBPF algorithm can be used to estimate unobserved masses of SOC.

Algorithm 4 Rao-Blackwellised particle filter (RBPF) algorithm

1: **for** $t = 1, ..., T$ **do**
2: Use the KF algorithm (1) on $X^{(1)}_{(t)}$ and compute the log-likelihood $l^{\text{KF}}_{(t)}$;
3: Apply a particle filter on $X^{(2)}_{(t)}$ to estimate the log-likelihoods $\hat{l}_{(t)}$;
4: Update the log-likelihood $l^{*}_{(t)} = \hat{l}_{(t)} + l^{\text{KF}}_{(t)}$;
5: Estimate the overall log-likelihood $L = \sum_t l^{*}_{(t)}$.

This section represented a discussion of a hierarchical model (HM) which includes two levels of models. The observation model is located at the top level of the hierarchy and is followed by the second level where the process model is located (Cressie and Wikle, 2015). Uncertainties around the hidden process are modelled by the process model via a probability distribution of the phenomenon of interest. The process model can be formed through sub-models. In addition, this section assumed that the static parameter $\boldsymbol{\theta}$ was fixed, but this is usually not the case and these two

levels typically depend on unknown parameters. Unknown parameters in a soil carbon dynamics model, for example, can be various quantities (e.g. decay rate of carbon in different pools) depend on a soil carbon model used by a user. The HM becomes a Bayesian Hierarchical Model (BHM) if the hierarchy contains the third level (underneath the observation model and the process model) which is specified by the joint probability distribution of all the unknown parameters (Allenby and Rossi, 2006; Berliner, 1996; Cressie and Wikle, 2015). In fact, this approach partitions the joint probability distribution over data, latent system states and model parameters into three conditional distributions which form well-defined hierarchical levels of the expression of uncertainties (Allenby and Rossi, 2006; Cressie and Wikle, 2015). A BHM can be represented mathematically as

$$p(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) = p(\mathbf{Y}, \mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{6}$$

where $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})$ is the observation model, $p(\mathbf{X}|\boldsymbol{\theta})$ is the process model, and $p(\boldsymbol{\theta})$ is the parameter model (Allenby and Rossi, 2006). In this context, inference on $\boldsymbol{\theta}$ and $\mathbf{X}$ through their joint posterior distribution which is proportional to equation (6) is of interest. When the posterior distribution is not tractable, it can be estimated by some numerical methods such as Markov chain Monte Carlo (MCMC) methods. To this end, in the following section, we introduce some sampling techniques based upon MCMC.

3 Particle marginal Metropolis-Hastings

When implementing a Bayesian approach, the purpose is to utilise the full joint posterior distribution over the set of unknown parameters, which are treated as random variables. In general for a model with parameter $\boldsymbol{\theta}$, latent state variable $\mathbf{X}$ and observed data $\mathbf{Y}$, the posterior distribution is defined as:

$$p(\mathbf{X}, \boldsymbol{\theta}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\mathbf{X}} \int_{\boldsymbol{\theta}} p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{X}}.$$

However, the denominator or "normalising constant" of the above posterior density requires the computation of typically intractable integrals. In these cases, we may proceed with sampling techniques based upon MCMC methods that do not require evaluation of the normalising constant. The Metropolis-Hastings (MH) algorithm (Andrieu et al., 2010) is an MCMC method for gener-

ating a sequence of correlated random samples from a probability distribution from which direct sampling is difficult. This sequence can be utilised to approximate the distribution or to estimate an expectation with respect to the posterior distribution (Andrieu et al., 2010).

Consider for the moment that we can analytically integrate out the state variables $\mathbf{X}$ and only inference on $\boldsymbol{\theta}$ is of interest. In that case, the expression for the posterior is:

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

Assume that $\boldsymbol{\theta}_{m-1}$ is the value of the Markov chain at iteration $m-1$ of the MCMC. A new parameter value $\boldsymbol{\theta}^*$ is proposed as the next value of the Markov chain according to a user-specified proposal distribution, $\boldsymbol{\theta}^* \sim Q(\cdot|\boldsymbol{\theta}_{m-1})$, conditional on the current value of the chain. The candidate $\boldsymbol{\theta}^*$ is accepted probabilistically based on the acceptance probability $\min(r, 1)$, where r is the below MH acceptance ratio:

$$r = \frac{p(\mathbf{Y}|\boldsymbol{\theta}^*)p(\boldsymbol{\theta}^*)Q(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}^*)}{p(\mathbf{Y}|\boldsymbol{\theta}_{m-1})p(\boldsymbol{\theta}_{m-1})Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})}; \tag{7}$$

The user is free to choose the proposal mechanism Q; if the Markov chain is run for enough iterations, it will converge to the desired posterior distribution. However, Q can have a significant impact on the finite-time efficiency of the MCMC (Robert, 2016). This proposal is often chosen to be symmetric, i.e. $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = Q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$, such as the normal distribution centred on the current value of the chain. Alternatively, it may be chosen to be asymmetric which is skewed towards larger values such as the log-normal distribution (Yildirim, 2012). The proposal distribution could also be selected to obey constraints on the parameter space, such as handling variance parameters that have all support placed on positive values.

The problem with this scheme is that in the case of non-linear and non-Gaussian state-space models, the likelihood $p(\mathbf{Y}|\boldsymbol{\theta})$ required in the acceptance ratio of the MH algorithm in equation (7) is often difficult to compute (Andrieu et al., 2010). The PMMH algorithm can be used to solve this problem.

3.1 Standard particle marginal Metropolis-Hastings

Particle filters give an unbiased estimate of the unknown likelihood for a given value of $\boldsymbol{\theta}$, which we denote as $\hat{p}(\mathbf{Y}|\boldsymbol{\theta})$ (Ala-Luhtala et al., 2016). The PMMH algorithm uses this unbiased likelihood

estimator in the MH ratio as a replacement to the exact likelihood. Even though only an unbiased estimator of the likelihood is used, the corresponding MCMC algorithms still converges to the true posterior distribution (Andrieu et al., 2010, 2009). The PMMH algorithm for estimating parameters is shown in Algorithm 5.

Algorithm 5 Particle marginal Metropolis-Hastings (PMMH) algorithm

1: Initialise $\boldsymbol{\theta}_0$;
2: **for** $m = 1, ..., M$ **do**
3: Draw $\boldsymbol{\theta}^*$ from proposal distribution $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})$;
4: Estimate the likelihood at $\boldsymbol{\theta}^*$ using a particle filter;
5: Compute the acceptance ratio:

$$r = \frac{\hat{p}(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}^*)p(\boldsymbol{\theta}^*)Q(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}^*)}{\hat{p}(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}_{m-1})p(\boldsymbol{\theta}_{m-1})Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})};$$

6: Accept $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}_m = \boldsymbol{\theta}^*$ with probability $\min(r, 1)$ otherwise $\boldsymbol{\theta}_m = \boldsymbol{\theta}_{m-1}$.

As in standard MCMC, an appropriate proposal distribution for generating each candidate parameter is required in the PMMH algorithm.

The number of particles used in the PFs should be chosen such that the variance of the log-likelihood ratio estimator $\hat{p}(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}^*)/\hat{p}(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}_{m-1})$ is approximately equal to one. This achieves a good compromise between MCMC mixing and computational complexity (Deligiannidis et al., 2018; Pitt et al., 2012). In addition, this variance is inversely proportional to the number of particles and increases linearly with the number of data T, therefore, the number of particles needed to keep the variance around one is $O(T^2)$ (Choppala et al., 2016; Deligiannidis et al., 2018). When the number of particles increases, the speed of computation decreases. To attempt to overcome this, we use an algorithm known as the correlated pseudo-marginal (CPM) method.

3.2 Correlated pseudo-marginal method

As mentioned earlier, the computational cost of the PMMH algorithm is $O(T^2)$ and this can be expensive for complex models and/or large datasets. The reason is that in the likelihood ratio of the PMMH algorithm, estimators in the numerator and denominator of the likelihood ratio are independent which increases the variance of the resulting ratio (Deligiannidis et al., 2018).

The CPM method is proposed by Deligiannidis et al. (2018) to correlate the estimators of the likelihood ratio of the PMMH algorithm in order to reduce the variance of the resulting ratio. These estimators are correlated through correlating the auxiliary random numbers used to obtain these

estimators. Assume that the unbiased estimator of the intractable likelihood $p(\mathbf{Y}|\boldsymbol{\theta})$ is $\hat{p}(\mathbf{Y}|\boldsymbol{\theta}, U)$, where U is the $\mathcal{U}$-valued auxiliary random variables that are used to obtain the likelihood estimator. Given a fixed set of random numbers U, all calculations in the particle filter are deterministic. This includes generating samples from the transition density and the resampling step. Denote the state of the joint Markov chain as $(\boldsymbol{\theta}_{m-1}, U_{m-1})$. The idea of the CPM method is to propose a new value $(\boldsymbol{\theta}^*, U^*)$ such that the random numbers U^* are close to U_{m-1}, then we might expect the likelihood estimates $\hat{p}(\mathbf{Y}|\boldsymbol{\theta}_{m-1}, U_{m-1})$ and $\hat{p}(\mathbf{Y}|\boldsymbol{\theta}^*, U^*)$ to be correlated. More specifically, the proposal density of the CPM algorithm is $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})Z(U^*|U)$ where Z is the density derived from the proposal $U^* = \tau U + \sqrt{1-\tau^2}\xi$. Here, $\xi \sim N(\mathbf{0}, \boldsymbol{I})$ and $\tau \in (-1, 1)$ (Deligiannidis et al., 2018). The MH acceptance ratio of this method is:

$$r = \frac{\hat{p}(\mathbf{Y}|\boldsymbol{\theta}^*, U^*)p(\boldsymbol{\theta}^*)Q(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}^*)}{\hat{p}(\mathbf{Y}|\boldsymbol{\theta}_{m-1}, U_{m-1})p(\boldsymbol{\theta}_{m-1})Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})}.$$

The value of the Markov chain at iteration m is $(\boldsymbol{\theta}^*, U^*)$ with probability $\min(r, 1)$, otherwise it is $(\boldsymbol{\theta}_{m-1}, U_{m-1})$. The CPM algorithm is presented in Algorithm 6.

Algorithm 6 Correlated pseudo-marginal (CPM) algorithm

1: Initialise $\boldsymbol{\theta}_0$;
2: **for** $m = 1, ..., M$ **do**
3: Sample $\boldsymbol{\theta}^* \sim Q(.|\boldsymbol{\theta}_{m-1})$;
4: Sample $\xi \sim N(\mathbf{0}, \boldsymbol{I})$ and set $U^* = \tau U_{m-1} + \sqrt{1-\tau^2}\xi$;
5: Compute the estimator $\hat{p}(\mathbf{Y}|\boldsymbol{\theta}^*, U^*)$ using Algorithm 7
6: Compute the acceptance ratio:

$$r = \frac{\hat{p}(\mathbf{Y}|\boldsymbol{\theta}^*, U^*)p(\boldsymbol{\theta}^*)Q(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}^*)}{\hat{p}(\mathbf{Y}|\boldsymbol{\theta}_{m-1}, U_{m-1})p(\boldsymbol{\theta}_{m-1})Q(\boldsymbol{\theta}^*|\Theta_{m-1})};$$

7: Accept $(\boldsymbol{\theta}^*, U^*)$ with probability $\min(r, 1)$ otherwise, output $(\boldsymbol{\theta}_{m-1}, U_{m-1})$

For the CPM method to be effective, it is crucial that the likelihood estimators in the numerator and denominator of the MH ratio be highly correlated. Therefore, we require a particle filter that processes the random numbers in such a manner that the likelihood estimates are similar as possible when slightly perturbing the random numbers. The particle filter with a given set of random numbers is shown in Algorithm 7. In the resampling step, systematic resampling is used to resample the sorted particles using normalised importance weights W_t. The quality of a sample generated with an MCMC algorithm can be checked by some MCMC diagnostic. In what follows, two popular diagnostic checks are presented.

3.3 MCMC diagnostics

For the sampling algorithms described to yield valid posterior distributions for statistical inference, it is necessary to perform some simple diagnostic checks. We use two popular diagnostic checks: (i) trace plots; and (ii) the Gelman and Rubin's convergence diagnostic statistic, $\hat{R}$ (Gelman and Rubin, 1992) both of which can be used to assess whether the MCMC samples have reached a stationary distribution and are "mixing" adequately. Trace plots are a simple, visual approach for assessing convergence of the Markov process. Convergence can be diagnosed more formally through the Gelman and Rubin's convergence diagnostic. In this approach, samples from multiple chains are compared to assess whether the output from each chain is indistinguishable, and this can be considered to be the case when the scale reduction factor is less than 1.2.

Algorithm 7 Particle filter with fixed random numbers

1: Sample $U_{(j^*)} \sim N(0,1)$ and $V_{(i^*)} \sim N(0,1)$ for all $j^* \in \{1,\ldots,TN\}$ and $i^* \in \{1,\ldots,T\}$;
2: Sample $X^k_{(1)} \sim p(.|U_{1:N}, \boldsymbol{\theta})$ for all $k \in \{1,\ldots,N\}$;
3: **for** $t = 1, ...,$ *T-1* **do**
4: Sort the collection $\{X^1_{(t)},\ldots,X^N_{(t)}\}$;
5: Compute importance weights $w^k_{(t)}$ and log-likelihoods $\hat{l}_{(t)} = \log\left(\frac{\sum_k w^k_{(t)}}{N}\right)$ for $k \in \{1,\ldots,N\}$;
6: Sample $X^k_{(t)}$ based on systematic resampling using random values $V_{1:T}$ and normalised weights $W^k_{(t)}$ for $k \in \{1,\ldots,N\}$;
7: Set $X^k_{(t+1)}$ as a sample from $p(.|X^k_{(t)}, U_{Nt+1:N(t+1)}, \boldsymbol{\theta})$ for $k \in \{1,\ldots,N\}$;
8: Estimate the overall log-likelihood $L = \sum_t \hat{l}_{(t)}$.

4 Case study

The Tarlee soil carbon model in Clifford et al. (2014) is used as a case study to illustrate the proposed methods, but the methods can be applied to other soil carbon state space models too. In the following, a background of the Tarlee site is provided in detail.

4.1 The Tarlee soil carbon model

Tarlee was an agricultural research experiment site located 80 km north of Adelaide, South Australia and was established in 1977 as a long term field experiment examining the impact of management practices on agricultural productivity (Skjemstad and Spouncer, 2003). The site is dominated

by winter rainfall with an average of 355 mm from April to October and the soil is classified as a hard-setting red-brown earth with sandy loam texture (Clifford et al., 2014). Long-term monitoring data from this site has been used for studying changes to soil carbon stocks under different rotations of cereal, pasture and fallow (Clifford et al., 2014; Luo et al., 2010; Skjemstad et al., 2004). The data and model motivating this study are those used in Clifford et al. (2014).

The Tarlee data used by Clifford et al. (2014) and in this study included those acquired from three agricultural management treatments: the first was wheat cropping for grain each year from 1979 - 1987 and 1990 - 1996, wheat for hay in 1988 and 1989, and the year 1997 being fallow; the second consisted of wheat cropping for grain and fallow rotations between 1979 - 1988 and 1990 - 1996, wheat for hay in 1989, and an additional year of fallow in 1997; the third consisted of alternating years of wheat and pasture from 1979 to 1987, pasture and wheat for hay rotations in 1988 and 1989, respectively, wheat for grain and pasture rotations from 1990 - 1996, with a final year of fallow in 1997. For years in which a field was non-fallow, the total dry matter (TDM; above-ground) was measured (t/ha) and in years where wheat was grown for grain, the total grain yield (t/ha) was also measured. In addition, total organic carbon (TOC) was measured for each treatment in the years 1979, 1985 and 1997.

Clifford et al. (2014) utilised a carbon model that included four latent variables which we will denote as X_C, X_G, X_W, and X_P. These represent the masses of SOC, total grain dry matter, total wheat dry matter, and total pasture dry matter, respectively, that evolve over time. To differentiate these latent variables from their measured values, we will use the random variables Y_C, Y_G, Y_W, and Y_P to denote the corresponding observations. In what follows, the process model of the soil carbon model used by Clifford et al. (2014) is presented.

4.2 Process model

In our process model of soil carbon dynamics, $X_{(t)}$ which is presented in equation (1) includes the masses of SOC in various pools, and the masses of carbon from different plant matter that can enter the soil which are presented in equations (8)-(11). It is important to highlight that equations (9) and (11) differ slightly from those presented in Clifford et al. (2014) which contained

a typographical error. The process models for time t and field indexed by i are:

$$\log(X^i_{C(t)}) = \log(X^i_{C(t-1)}e^{-K\Delta t} + I^i_{C(t)}) + \eta^i_{(t)}, \quad \eta^i_{(t)} \sim N(0, \sigma^2_\eta); \tag{8}$$

$$X^i_{G(t)} \sim LN(\mu_G + \rho_G(\log(X^i_{G(t-1)}) - \mu_G), \sigma^2_G); \tag{9}$$

$$X^i_{W(t)} \sim LN(\log h + \log(x^i_{G(t)}), \sigma^2_W); \quad \text{and} \tag{10}$$

$$X^i_{P(t)} \sim LN(\mu_P + \rho_P(\log(X^i_{P(t-1)}) - \mu_P), \sigma^2_P); \tag{11}$$

where ρ_G and ρ_P are auto-regressive parameters for the evolution of grain TDM and pasture TDM respectively; Δt (= 1 year) is the time step and h is the harvest index (the ratio of wheat TDM to grain TDM). The normal distribution is denoted by $N(\mu_1, \sigma^2_1)$ with mean parameter μ_1 and variance parameter σ^2_1, and $LN(\mu_2, \sigma^2_2)$ denotes the log-normal distribution with mean parameter μ_2 and variance parameter σ^2_2 for a log transformation of the random variable. The model treats soil carbon as a single pool which has exponential decay rate K, assumed to be the same in each field. In addition, $X^i_{W(t)}$ is defined conditional on $X^i_{G(t)} = x^i_{G(t)}$. Although the total wheat dry matter contains the total grain dry matter i.e. $X^i_{G(t)} \leq X^i_{W(t)}$, these processes are separated since grain yield is usually measured. The mass of carbon input to each field in each year, $I^i_{C(t)}$, is modelled as a function of the management practice, which is:

$$I^i_{C(t)} = \begin{cases} c(X^i_{W(t)} - X^i_{G(t)}) + cr_W X^i_{W(t)} & \text{Wheat for Grain} \\ cpX^i_{W(t)} + cr_W X^i_{W(t)} & \text{Wheat for Hay} \\ cX^i_{P(t)} + cr_P X^i_{P(t)} & \text{Pasture} \\ cpX^i_{P(t)} + cr_P X^i_{P(t)} & \text{Pasture for Hay} \\ 0 & \text{Fallow} \end{cases}$$

where p is the proportion of the crop left above-ground after harvest, r_P and r_W denote the root-to-shoot ratios (in terms of TDM) for pasture and wheat crops, respectively. The input function $I^i_{C(t)}$ includes carbon that enters into the soil from plant-matter that is already below-ground (i.e., roots) and from plant-matter that remains above-ground after harvesting, $c(X^i_{W(t)} - X^i_{G(t)})$, in which, c is the carbon content of dry plant matter and the above-ground plant-matter biomass is denoted by $(X^i_{W(t)} - X^i_{G(t)})$. Given a vector of parameters for the model, denoted by $\boldsymbol{\theta}$ that includes all parameters presented in Table 1, the joint process model of the four independent processes (8)-(11)

at time t and field i is a discrete-time Markov chain whose transition density can be written as:

$$p(X^i_{(t)}|X^i_{(t-1)},\boldsymbol{\theta}) = p(X^i_{C(t)}|I^i_{C(t)},X^i_{C(t-1)},\boldsymbol{\theta}) \times p(X^i_{G(t)}|X^i_{G(t-1)},\boldsymbol{\theta}) \\ \times p(X^i_{W(t)}|X^i_{G(t)},\boldsymbol{\theta}) \times p(X^i_{P(t)}|X^i_{P(t-1)},\boldsymbol{\theta});$$

where $X^i_{(t)} = (X^i_{C(t)}, X^i_{G(t)}, X^i_{W(t)}, X^i_{P(t)})$ for $i \in \{1,2,3\}$. Assuming that the three fields are independent, the overall transition density is therefore:

$$p(X_{(t)}|X_{(t-1)},\boldsymbol{\theta}) = \prod_{i=1}^{3} p(X^i_{(t)}|X^i_{(t-1)},\boldsymbol{\theta});$$

where $X_{(t)} = (X^1_{(t)}, X^2_{(t)}, X^3_{(t)})$.

The next section is allocated to describe the observation model of the soil carbon model utilised by Clifford et al. (2014).

4.3 Observation model

Observation models for the measured data are constructed by conditioning on the unobserved variables X_C, X_G, X_W, and X_P. At time t in field i, Clifford et al. (2014) used the following models to account for measurement error:

$$Y^i_{C(t)}|X^i_{C(t)} = x^i_{C(t)} \sim LN(\log(x^i_{C(t)}), \sigma^2_{\epsilon C}); \tag{12}$$

$$Y^i_{G(t)}|X^i_{G(t)} = x^i_{G(t)} \sim LN(\log(x^i_{G(t)}), \sigma^2_{\epsilon G}); \tag{13}$$

$$Y^i_{W(t)}|X^i_{W(t)} = x^i_{W(t)} \sim LN(\log(x^i_{W(t)}), \sigma^2_{\epsilon W}); \quad \text{and} \tag{14}$$

$$Y^i_{P(t)}|X^i_{P(t)} = x^i_{P(t)} \sim LN(\log(x^i_{P(t)}), \sigma^2_{\epsilon P}); \tag{15}$$

where $Y^i_{C(t)}$, $Y^i_{G(t)}$, $Y^i_{W(t)}$, and $Y^i_{P(t)}$ are the measurement data of the SOC, grain dry matter, wheat dry matter, and pasture dry matter, respectively. All data in field i at time t is denoted by $Y^i_{(t)} = (Y^i_{C(t)}, Y^i_{G(t)}, Y^i_{W(t)}, Y^i_{P(t)})$. The joint observation model of the independent measurements (12)-(15) at time t and field i, conditioning on unobserved variable vector $X^i_{(t)}$ and parameter $\boldsymbol{\theta}$ is:

$$p(Y^i_{(t)}|X^i_{(t)},\boldsymbol{\theta}) = p(Y^i_{C(t)}|X^i_{C(t)},\boldsymbol{\theta}) \times p(Y^i_{G(t)}|X^i_{G(t)},\boldsymbol{\theta}) \\ \times p(Y^i_{W(t)}|X^i_{W(t)},\boldsymbol{\theta}) \times p(Y^i_{P(t)}|X^i_{P(t)},\boldsymbol{\theta});$$

and the overall observation model across all three fields is therefore:

$$p(Y_{(t)}|X_{(t)}, \boldsymbol{\theta}) = \prod_{i=1}^{3} p(Y^i_{(t)}|X^i_{(t)}, \boldsymbol{\theta});$$

where $Y_{(t)} = (Y^1_{(t)}, Y^2_{(t)}, Y^3_{(t)})$.

4.4 Parameter model

To quantify the uncertainty in parameters and predictions, Clifford et al. (2014) adopted a Bayesian approach for model fitting. To conduct a Bayesian analysis, we place a prior distribution on the unknown model parameter $\boldsymbol{\theta}$. It represents our prior knowledge about the ranges of likely values that parameters can take. The prior is often constructed as a set of independent univariate probability density functions, whose joint density can then be easily constructed by simple multiplication. For the Tarlee model and data, we used the same prior distributions as Clifford et al. (2014), which are defined for 22 independent parameters of the observation model, process model as well as the initial masses of SOC in the fields at the commencement of the study (1978). For ease of access, we re-present the priors defined by Clifford et al. (2014) here in Table 1.

4.5 Estimating the posterior distribution for the SOC model

Changes in soil carbon stocks over time are important to both government agencies and farmers. One of the aims of this paper is to quantify the uncertainty surrounding important soil attributes such as the amount of SOC that was added through the management practices, and estimating the parameters driving the sequestration of carbon. To estimate the posterior distribution $p(X_C, \boldsymbol{\theta}|\boldsymbol{Y})$, we draw samples from $p(\boldsymbol{X}, \boldsymbol{\theta}|\boldsymbol{Y})$ as it can be decomposed to $p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{Y}) \times p(\boldsymbol{\theta}|\boldsymbol{Y})$ and preserve the components related to the SOC process X_C and its parameters $\boldsymbol{\theta}$. Here $\mathbf{Y}$ represents all available data and $\mathbf{X}$ represents all state variables at all times and all fields. Because the resulting likelihood in the model is intractable, the CPM algorithm (6) is applied to the model to draw samples of parameters from $p(\boldsymbol{\theta}|\boldsymbol{Y})$. For simplicity, for our motivating model, we use the same proposal distributions as in Clifford et al. (2014), which are reproduced in Table 2.

Since the model is a combination of linear and non-linear state variables, the marginal likelihood of non-linear state variable X_C must be estimated using PFs. The marginal likelihoods of the linear-Gaussian state variables can also be computed by PFs, but can be computed more efficiently by the

Parameter	Prior	Type
$X^1_{C(1978)}$	Truncated-normal$(40, 10^2, lower = 0)$	Uninformative
$X^2_{C(1978)}$	Truncated-normal$(40, 10^2, lower = 0)$	Uninformative
$X^3_{C(1978)}$	Truncated-normal$(40, 10^2, lower = 0)$	Uninformative
K	LN$(-2.71, (0.127)^2)$	Informative
c	N$(0.45, (0.01)^2)$	Informative
r_W	N$(0.5, (0.067)^2)$	Informative
r_P	N$(1, (0.125)^2)$	Informative
p	Beta$(89.9, 809.1)$	Informative
h	LN$(0.825, (0.36)^2)$	Weakly Informative
μ_G	N$(0.42, (1.18)^2)$	Weakly Informative
μ_W	N$(1.24, (1.12)^2)$	Weakly Informative
μ_P	N$(1.41, (1.81)^2)$	Weakly Informative
ρ_G	Uniform$(-1, 1)$	Uninformative
ρ_P	Uniform$(-1, 1)$	Uninformative
σ^2_η	Inv-gamma$(0.001, 0.001)$	Uninformative
σ^2_G	Inv-gamma$(0.001, 0.001)$	Uninformative
σ^2_W	Inv-gamma$(0.001, 0.001)$	Uninformative
σ^2_P	Inv-gamma$(0.001, 0.001)$	Uninformative
$\sigma^2_{\epsilon C}$	0.025	Fixed
$\sigma^2_{\epsilon G}$	0.023	Fixed
$\sigma^2_{\epsilon W}$	0.133	Fixed
$\sigma^2_{\epsilon P}$	0.067	Fixed

Table 1: Prior distributions of parameters in the model.

Parameter	Proposal
K	N$(K, 0.001^2)$
c	Truncated-normal$(c, 0.005^2, lower = 0, upper = 1)$
r_W	Truncated-normal$(r_W, 0.05^2, lower = 0)$
r_P	Truncated-normal$(r_P, 0.05^2, lower = 0)$
p	Truncated-normal$(p, 0.005^2, lower = 0, upper = 1)$
h	Truncated-normal$(h, 0.05^2, lower = 0)$
μ_G	N$(K, 0.05^2)$
μ_P	N$(K, 0.05^2)$
ρ_G	Truncated-normal$(\rho_G, 0.05^2, lower = -1, upper = 1)$
ρ_P	Truncated-normal$(\rho_P, 0.1^2, lower = -1, upper = 1)$
σ^2_η	Truncated-normal$(\sigma^2_\eta, 0.001^2, lower = 0)$
σ^2_G	Truncated-normal$(\sigma^2_G, \frac{\sigma^2_G}{20^2}, lower = 0)$
σ^2_W	Truncated-normal$(\sigma^2_W, 0.001^2, lower = 0)$
σ^2_P	Truncated-normal$(\sigma^2_P, 0.1^2, lower = 0)$

Table 2: Proposal distributions used in the CPM method.

KF that will yield exact likelihoods for these state components. The KF is applicable in the Tarlee model by using a log transformation, i.e. $\tilde{X} = \log(X)$ and $\tilde{Y} = \log(Y)$, the state-space model formed by the sub-model involving $\{\tilde{X}_G, \tilde{X}_W, \tilde{X}_P, \tilde{Y}_G, \tilde{Y}_W, \tilde{Y}_P\}$ forms a linear-Gaussian state-space

model. Since $\tilde{X}_{W(t)}$ depends on $\tilde{X}_{G(t)}$ and the structure of the state-space model is an auto-regressive model, we rewrite it based on $\tilde{X}_{G(t-1)}$ so that:

$$\tilde{X}^i_{W(t)} \sim N(\log h + \mu_G + \rho_G(\tilde{X}^i_{G(t-1)} - \mu_G), \sigma^2_W + \sigma^2_G).$$

The state-space model of $\tilde{X}^i_{(t)} = [\tilde{X}^i_{G(t)}, \tilde{X}^i_{W(t)}, \tilde{X}^i_{P(t)}]$, $i \in \{1, 2, 3\}$ and its corresponding observation is:

$$\begin{aligned}
\tilde{X}^i_{(t)} &= \boldsymbol{A}\tilde{X}^i_{(t-1)} + \boldsymbol{B} + \epsilon^*_{(t)}, \quad \epsilon^*_{(t)} \sim N(\boldsymbol{0}, \boldsymbol{Q}^*); \\
\tilde{Y}^i_{(t)} &= \boldsymbol{C}\tilde{X}^i_{(t)} + \nu^*_{(t)}, \quad \nu^*_{(t)} \sim N(\boldsymbol{0}, \boldsymbol{R}^*);
\end{aligned}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are:

$$\boldsymbol{A} = \begin{bmatrix} \rho_G & 0 & 0 \\ \rho_G & 0 & 0 \\ 0 & 0 & \rho_P \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \mu_G(1-\rho_G) \\ \mu_G(1-\rho_G) + \log(h) \\ \mu_P(1-\rho_P) \end{bmatrix}, \text{ and } \quad \boldsymbol{C} = \boldsymbol{I}_{3\times 3}.$$

Also,

$$\boldsymbol{Q}^* = \begin{bmatrix} \sigma^2_G & \sigma^2_G & 0 \\ \sigma^2_G & \sigma^2_W + \sigma^2_G & 0 \\ 0 & 0 & \sigma^2_P \end{bmatrix}, \quad \boldsymbol{R}^* = \begin{bmatrix} \sigma^2_{\epsilon G} & 0 & 0 \\ 0 & \sigma^2_{\epsilon W} & 0 \\ 0 & 0 & \sigma^2_{\epsilon P} \end{bmatrix}.$$

Thus, the RBPF algorithm (4) is applicable to estimate the marginal likelihood of the non-linear state component as well as compute the exact marginal likelihoods of linear state components.

The KF and PF filters are reused to draw a sample of the state process from $p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{Y})$. To obtain this sample from the posterior distribution X_C, we sample one of the trajectories based on the final particle weights at time T and then trace this particles lineage backwards through time in order to obtain a complete sample path $(X^*_{C(1)}, \ldots, X^*_{C(T)})$. The algorithm of drawing a single trajectory X^*_C recursively is shown in Algorithm 8.

where b^*_t is the index of $X_{C(t)}$ at time t, *Multinomial* refers to a multinomial distribution that gives an index $k \in \{1, \ldots, N\}$ proportional to the probability $W^k_{(t)}$, and a^k_t is the index of the kth sampled particle in the resample step of the particle filter algorithm at time t.

Quantifying the uncertainty of our estimate can be done in many ways, such as through the

Algorithm 8 Sampling from the posterior distribution X_C

1: Draw $b_T^* \sim Multinomial(W_{(T)}^1, \ldots, W_{(T)}^N)$;
2: Set $X_{C(T)}^* = X_{C(T)}^{b_T^*}$;
3: **for** $t = T, \ldots, 2$ **do**
4: Set $b_{t-1}^* = a_t^{b_t^*}$;
5: Set $X_{C(t-1)}^* = X_{C(t-1)}^{b_{t-1}^*}$;

estimated probability of functionals of interest or a 95% credible interval. We can infer the mass of SOC added over a 20-year period from sampling from the posterior distribution using MCMC. We work with 10 chains in the CPM method, each initialised with a randomly sampled parameter vector and then run for 50,000 iterations with the first 5,000 discarded as burn-in. Therefore, the total posterior sample size is 450,000. The posterior distribution $p(\boldsymbol{X}, \boldsymbol{\theta}|\boldsymbol{Y})$, is represented by M samples $\{(X^m, \boldsymbol{\theta}^m) : m = 1, ..., M\}$ drawn from $p(\boldsymbol{X}, \boldsymbol{\theta}|\boldsymbol{Y})$ so that these samples can be used to estimate the posterior expectation of any function $\varphi(\boldsymbol{X}, \boldsymbol{\theta})$.

$$\mathbf{E}(\varphi(\boldsymbol{X}, \boldsymbol{\theta})|\boldsymbol{Y}) \approx \frac{1}{M} \sum_{m=1}^{M} \varphi(X^m, \boldsymbol{\theta}^m).$$

The only limitation of the accuracy of such estimates is the sample size, but for sufficiently large M the error is negligible.

5 Results

We use two particle filters APF and BPF in the RBPF algorithm for the non-linear part of the model. To avoid confusion, we define them as RB-APF and RB-BPF, respectively. In what follows, we compare the performances of three algorithms BPF, RB-APF and RB-BPF in terms of the precision of estimating the log-likelihood. The particle filter that has the smallest variance of the estimated log-likelihood for a given number of particles will be used in the MCMC to obtain the full posterior of parameters and latent states. Finally, we present the estimation of state variables, model parameters as well as computing the convergence diagnostic of model parameters.

5.1 Comparison of particle filters

We use BPF, RB-APF and RB-BPF to estimate the log-likelihood of the state variables at a point estimate of the parameter based on approximate univariate posterior modes (obtained from the

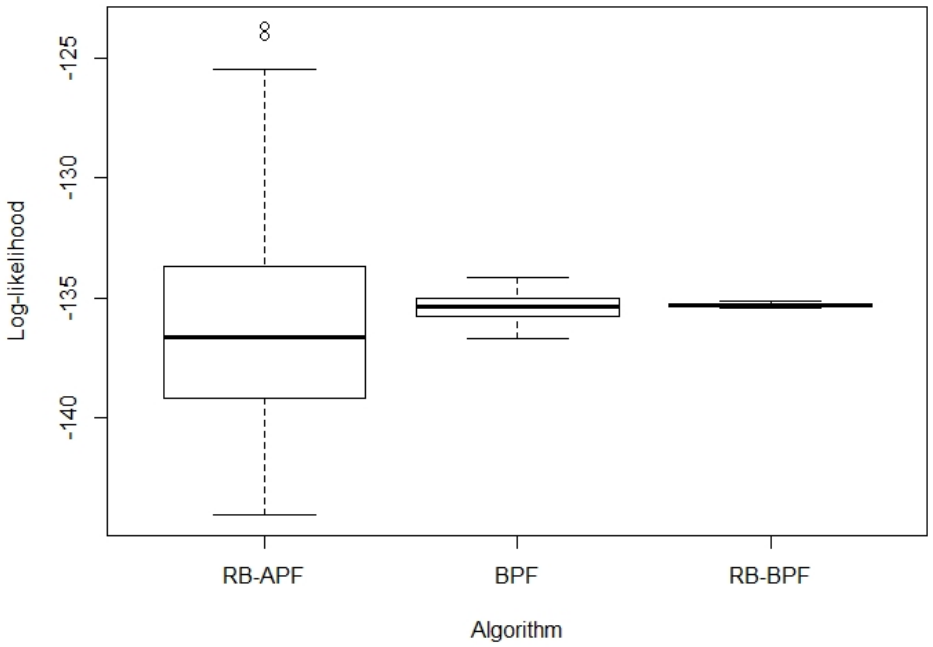

Figure 2: The box plots showing the precision of the estimated log-likelihoods at the posterior mode

results of Clifford et al. (2014)) and compared their results to select the appropriate one for using in the PMMH algorithm. In the case of the BPF, the linear and Gaussian structure of some of the state variables are not taken advantage of, whereas for the RB-BPF and RB-APF we take advantage of being able to compute some components of the log-likelihood analytically. To this end, we did an experiment to estimate the log-likelihoods of the measurement variables with 1000 particles and repeat it independently 100 times to see the precision of estimation based on each of the aforementioned particle filters. Table 3 shows the standard deviations of the estimated log-likelihoods of the measurement variables and Figure 2 shows the box plots of them. As shown in Table 3, the performance of the RB-BPF algorithm is far better than two other algorithms.

Method	Standard deviation	N
RB-APF	4.8	1000
BPF	0.54	1000
RB-BPF	**0.07**	**1000**

Table 3: The standard deviations of the estimated log-likelihoods based on RB-APF, BPF and RB-BPF algorithms with 1000 particles at the posterior mode.

5.2 Inference on the hidden process model

To perform inference on the mass of SOC, we chose the RB-BPF to estimate the state variables and likelihoods required in the MH acceptance ratio on the basis that it has greater precision at estimating log-likelihoods than the BPF and the RB-APF. As mentioned in section 3.1, the number of particles should be chosen such that the variance of the estimated log-likelihoods is around one to have a good compromise between MCMC mixing and computational complexity. In this case, for getting the number of particles, the standard deviation of the estimated log-likelihood is chosen to be 0.5 and we compared the speed of computation of the RB-BPF algorithm with the BPF algorithm which is applied in Clifford et al. (2014). The results are presented in Table 4. The RB-BPF requires about 50 times fewer particles compared to the BPF and is about 3 times faster.

Method	Time (Secs)	N
BPF	16.3	1060
RB-BPF	**5.6**	**20**

Table 4: The number of particles (N) for estimating log-likelihoods based on BPF and RB-BPF algorithms at the posterior mode.

Consider the function $\varphi(\boldsymbol{X}, \boldsymbol{\theta})$ which is described in section 4.5, as the change in SOC to field i between 1978 and following year t:

$$\varphi(\boldsymbol{X}, \boldsymbol{\theta}) = X^i_{C(t)} - X^i_{C(1978)};$$

and a common Bayes estimate of this function is:

$$\hat{\varphi}(\boldsymbol{X}, \boldsymbol{\theta}) = \mathbf{E}(X^i_{C(t)} - X^i_{C(1978)} | \boldsymbol{Y}).$$

As it is shown in Figure 3, the amount of SOC is reduced in fields 1 and 2 while it is increased in field 3 and the averages of the SOC change from 1978 in fields 1, 2, and 3 are -7.48, -7.11, and 2.06, respectively, with negative values denoting that fields 1 and 2 are expected to lose carbon over a 20-year period.

The 95% credible intervals for φ can be achieved by computing the 2.5^{th} and the 97.5^{th} percentiles of the posterior distribution and define them as the lower and upper limits of the interval, respectively. These 95% credible intervals for the amount of carbon added to the three fields in Tarlee are $(-20.68, 4.92)$, $(-19.93, 5.07)$, and $(-11.38, 14.56)$ for fields 1, 2, and 3, respectively.

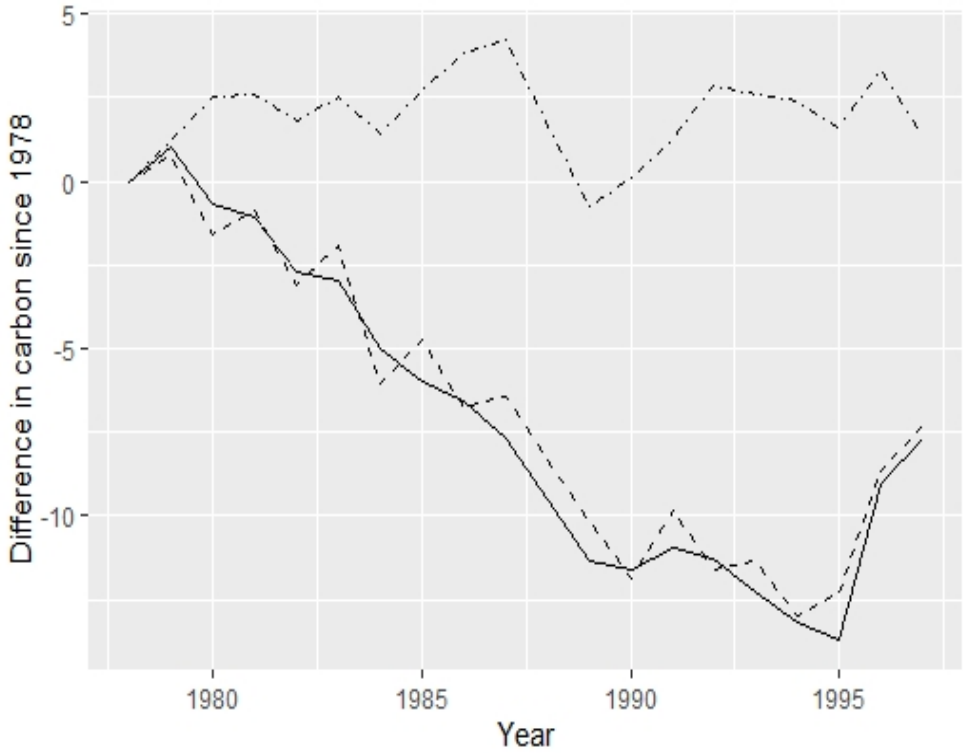

Figure 3: The expected difference of the SOC in each year from 1978, estimated based on the RB-BPF. The changes in fields 1, 2, and 3 are indicated by solid line, dashed line, and dot-dash line, respectively.

We plot the 2.5^{th}, 25^{th}, 50^{th}, 75^{th}, and 97.5^{th} percentiles for the SOC process and its forecast from 1998 to 2018 under the same management strategies of each field based on all 450,000 draws from the posterior distribution in Figure 4. The posterior model is run forward from 1998 to 2018 to forecast SOC stocks. As there are no available observations for that period, the uncertainty around the SOC inputs is large and consequently, the uncertainty on SOC levels for each field is large. The management strategies that are used in fields 1, 2, and 3 are "Wheat-Wheat", "Wheat-Fallow", and "Wheat-Pasture", respectively. It is noteworthy to mention that the proposed methods are more efficient computationally than the methods used by Clifford et al. (2014) to estimate the amount of carbon added between 1978 and 1997 which is shown in Figure 4.

Figure 5 highlights the histogram of samples drawn from the posterior distributions of model parameters which presented in Table 1, and compares them with the prior distributions of the model parameters to see what we have learned about them. Some priors are hardly visible when they are plotted over the histograms because they are uninformative and much has been learned about those parameters from the data. For example, the initial SOC conditions of each field, the harvest index, the mean of the pasture yield, and different variance parameters in Figure 5 show good evidence of divergence between prior and posterior distributions. Furthermore, the posterior distributions from the CPM method based on the RB-BPF approach agree with those

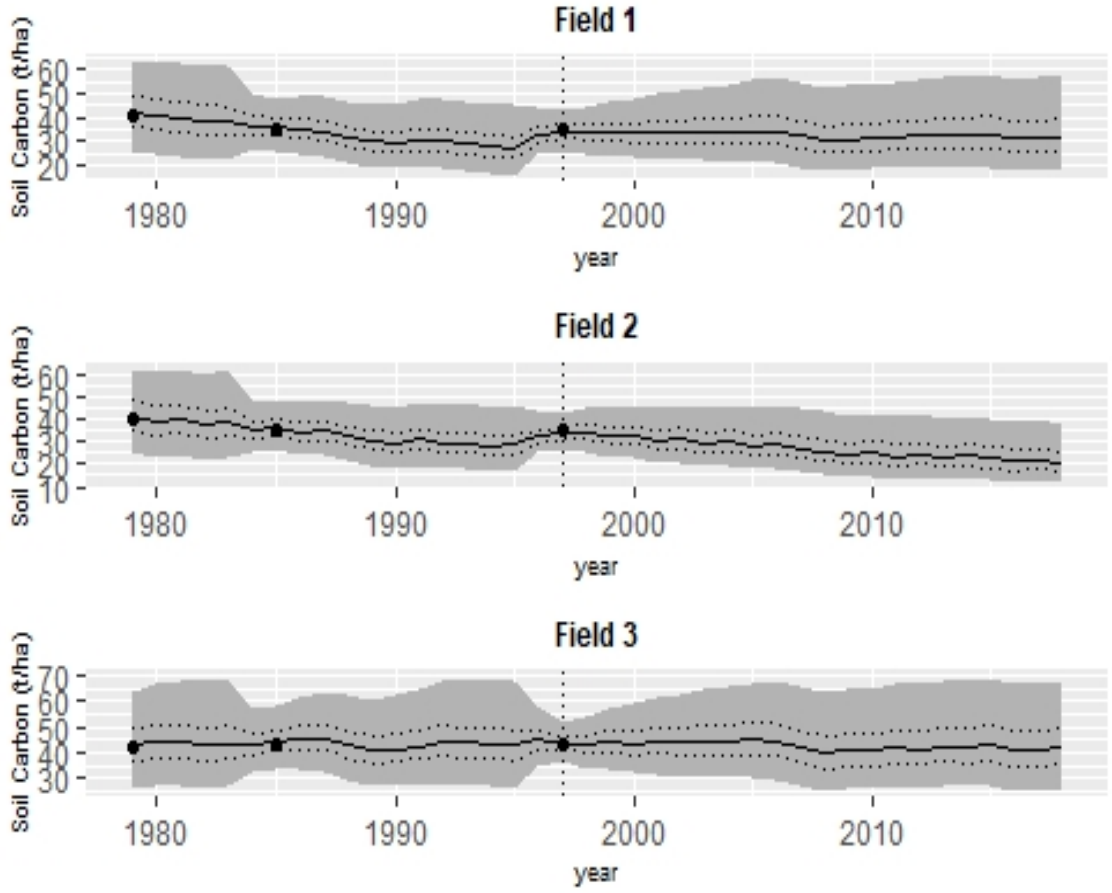

Figure 4: Soil carbon dynamics and its forecast from 1998 onwards under the same management strategies in the three Tarlee fields, estimated by the RB-BPF algorithm. The gray shaded part is the area between the 2.5^{th} percentiles and the 97.5^{th} percentiles for the SOC process. The 25^{th} percentiles and the 75^{th} percentiles for the SOC process are indicated by the dashed lines. The median carbon mass is indicated by the solid line and the measured SOC values are indicated by filled dots.

from Clifford et al. (2014) which estimated from the PMMH method based on the BPF. The advantage of using these new methods over the methods that are used in Clifford et al. (2014) is the significant computational efficiency of them to replicate the results in Clifford et al. (2014).

If the RB-BPF is used in both CPM and PMMH methods with the same sample size for estimating the likelihood ratio, the number of particles required by the CPM and PMMH methods in each MCMC iteration are 20 and 55, respectively. The elapsed running times of the CPM and PMMH methods for one chain with 50,000 samples based on the BPF algorithm are 156 and 210 minutes, respectively. This illustrate the better computational efficiency of the CPM method than the PMMH method for a given choice of particle filter. Also the elapsed running time of the PMMH method based on the BPF method which is used in Clifford et al. (2014) and the CPM method based on the RB-BPF method for one chain with 50,000 samples are about 50 and 210 minutes, respectively. The elapsed time for running the CPM method based on the RB-APF algorithm is not reported here, as it took too long to converge to a stationary distribution.

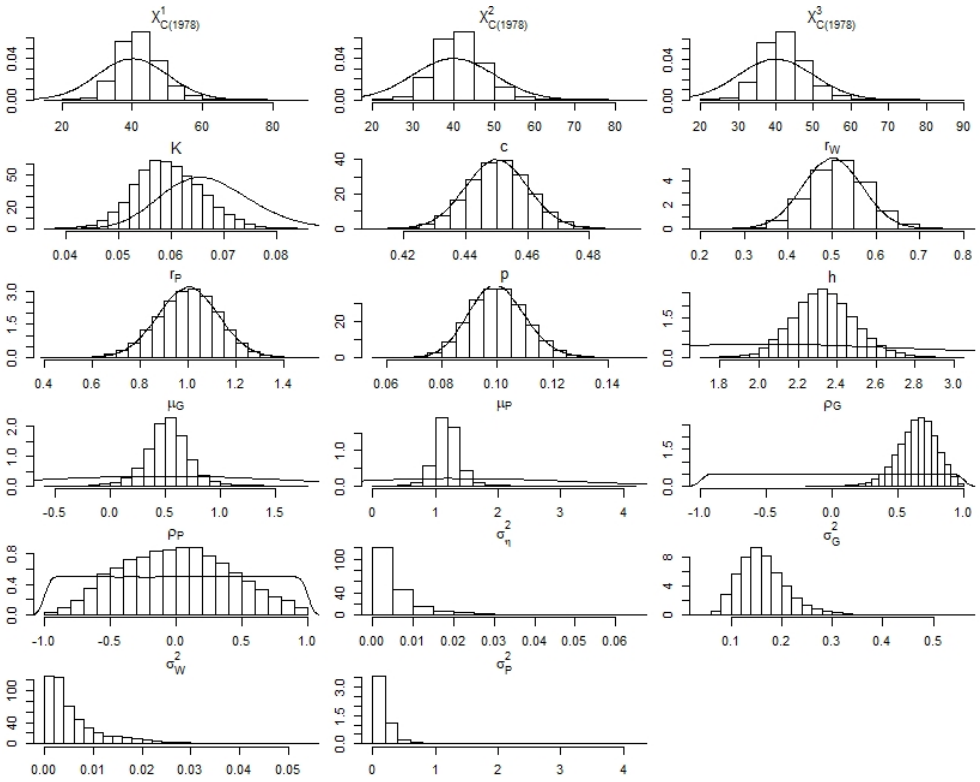

Figure 5: The marginal posterior distribution (histogram) based on the RB-BPF method for 17 model parameters and their prior distribution (grey density).

We use the multivariate effective sample size divided by the number of particles as the overall measure of performance to compare the statistical efficiency of the CPM method based on the RB-BPF algorithm and what is used in Clifford et al. (2014). The multivariate effective sample size divided by the number of particles for the CPM method based on the RB-BPF algorithm and the PMMH method based on the BPF algorithm are 2.57 and 0.067, respectively. This illustrates that the CPM method based on the RB-BPF algorithm outperforms the PMMH method based on the BPF algorithm in terms of statistical efficiency.

The trace plots of 14 model parameters are presented in Figure 6 of the Appendix. The results of the Gelman and Rubin's convergence diagnostic statistics for 14 model parameters of the Tarlee data set based on the RB-BPF method are shown in Table 5. Since the point estimate of $\hat{R}$ is less than 1.2, this indicates that the MCMC samples have converged a stationary distribution and are mixing adequately.

The information gain observed between prior and posterior distributions in Figure 5 can be used to formulate prior distributions for future studies in Tarlee or similar soil types and land management.

6 Discussion and conclusions

In this paper, we have revisited the use of Bayesian statistical methods for model fitting and uncertainty quantification for the soil carbon model presented in Clifford et al. (2014). A major advantage of the Bayesian hierarchical modelling (BHM) framework used in that work and here is that it allows us to think conditionally and critically about the process, the parameters, and the data. This is a natural way to decompose the larger statistical problem into manageable components that can be communicated more easily within multi-disciplinary science teams. A drawback of using these Bayesian approaches when fitting models to data is that they can be computationally burdensome. However, in this work, we have shown that advanced Bayesian methods can drastically increase the computational efficiency of such modelling activities compared to those employed in Clifford et al. (2014). In particular, we demonstrate that the Rao-Blackwellised bootstrap particle filter (RB-BPF) offers substantial efficiency gains (in terms of computation time and precision of estimating the log-likelihood) compared to the standard bootstrap particle filter (BPF) and Rao-Blackwellised auxiliary particle filter (RB-APF). In terms of the precision of estimating the likelihood for a fixed number of particles, the number of particles required and computation time, the RB-BPF algorithm greatly outperforms the RB-APF, and the BPF algorithm which was used in Clifford et al. (2014) (see Tables 3 and 4). This RB-BPF algorithm will therefore be useful for building decision support systems that allow farmers and other land managers to rapidly obtain predictions of soil carbon stocks into the future, based on past measurements using standard computer hardware. In particular, efficient Bayesian statistical methods such as the RB-BPF are likely to be beneficial to land-holders interested in selling carbon credits for carbon sequestered in the soil. Since Bayesian methods allow for the estimation of probabilities associated with meeting sequestration targets, they are therefore useful for assessing the financial risks and rewards that might be associated with entering into such contracts.

Another important aspect of this paper is in introducing existing advanced Bayesian methods to soil scientists and other environmental scientists working with similar classes of models. We build on the ideas introduced in Clifford et al. (2014) to show the benefits of using advanced particle filters rather than the standard bootstrap particle filter. We show how they can improve the speed of computation and the precision of estimation which can be useful in the case of having a large data set for representing knowledge about model parameters, model process dynamics and measurement errors. It is important that these three sources of uncertainty are acknowledged jointly when making

predictions about soil carbon stocks, otherwise a land-owner might unwittingly enter into a contract to sequester carbon that has higher risk than anticipated. Figure 4 demonstrates that when few observations of soil carbon are available, there can be considerable uncertainty around what the soil carbon stock is and that frequent monitoring might be necessary to obtain predictions with enough certainty to make the sale of carbon credits a low-risk venture. In addition, the Bayesian modelling approach makes it possible for a regulator to set limits on the volume of carbon credits they will allow a land holder to sell on the basis of how achievable that is deemed to be.

In this study, the one-pool model of Clifford et al. (2014) is used to illustrate the performance and benefits of advanced Bayesian methods and the correlated pseudo-marginal (CPM) method over the standard BPF and standard pseudo-marginal (i.e. particle marginal Metropolis-Hastings (PMMH)), respectively. The importance of these methods will be more obvious in the case of: (i) more complex models i.e. models with multiple carbon pools (higher state-space dimension); and (ii) with longer datasets. Both the performance of the standard BPF and PMMH methods decrease as these two model attributes are increased in size. Further research should consider how these advanced Bayesian methods can also be used to identify the most parsimonious model for predicting soil carbon dynamics. For example, the optimal number of carbon pools to include could be considered as a model selection problem to provide a good fit to the data whilst avoiding unnecessary model complexity model complexity.

7 Data availability

Data set can be accessed online at:
https://doi.org/10.4225/08/54F0786D6D923.

8 Acknowledgments

We would like to thank CSIRO for providing the data used in this study. MJD was supported by QUT-CSIRO Digital Agriculture Scholarship and a CSIRO Digital Agriculture Top-Up Scholarship. CD was supported by an Australian Research Council Discovery Project (DP200102101). We gratefully acknowledge the computational resources provided by QUT's High Performance Computing (HPC) and Research Support Group.

References

B. Ahrens, M. Reichstein, W. Borken, J. Muhr, S. E. Trumbore, and T. Wutzler. Bayesian calibration of a soil organic carbon model using Δ14c measurements of soil organic carbon and heterotrophic respiration as joint constraints. *Biogeosciences*, 11(8):2147–2168, 2014.

J. Ala-Luhtala, N. Whiteley, K. Heine, and R. Piché. An introduction to twisted particle filters and parameter estimation in non-linear state-space models. *IEEE Transactions on Signal Processing*, 64(18):4875–4890, 2016.

Greg M Allenby and Peter E Rossi. Hierarchical bayes models. *The handbook of marketing research: Uses, misuses, and future advances*, pages 418–440, 2006.

C. Andrieu, G. O. Roberts, et al. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.

C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

L. M. Berliner. Hierarchical bayesian time series models. In *Maximum entropy and Bayesian methods*, pages 15–22. Springer, 1996.

J. M. Cable, K. Ogle, A. P. Tyler, M. A. Pavao-Zuckerman, and T. E. Huxman. Woody plant encroachment impacts on soil carbon and microbial processes: results from a hierarchical bayesian analysis of soil incubation data. *Plant and Soil*, 320(1-2):153–167, 2009.

C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

P. Choppala, D. Gunawan, J. Chen, M. N. Tran, and R. Kohn. Bayesian inference for state space models using block and correlated pseudo marginal methods. *arXiv preprint arXiv:1612.07072*, 2016.

D. Clifford, D. Pagendam, J. Baldock, N. Cressie, R. Farquharson, M. Farrell, L. Macdonald, and L. Murray. Rethinking soil carbon modelling: a stochastic approach to quantify uncertainties. *Environmetrics*, 25(4):265–278, 2014.

N. Cressie and C. K. Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.

G. Deligiannidis, A. Doucet, and M. K. Pitt. The correlated pseudomarginal method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):839–870, 2018.

A. Doucet and A. M Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3, 2009.

A. Doucet, N. De Freitas, K. Murphy, and S. Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 176–183. Morgan Kaufmann Publishers Inc., 2000.

E. T. Elliott, K. Paustian, and S. D. Frey. Modeling the measurable or measuring the modelable: A hierarchical approach to isolating meaningful soil organic matter fractionations. In *Evaluation of soil organic matter models*, pages 161–179. Springer, 1996.

P. Fearnhead and H. R Künsch. Particle filters and data assimilation. 2018.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

N. J. Gordon, D. J. Salmond, and A. F. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE proceedings F (radar and signal processing)*, volume 140, pages 107–113. IET, 1993.

M. S. Grewal. Kalman filtering. In *International Encyclopedia of Statistical Science*, pages 705–708. Springer, 2011.

Chao Huang, Hong S He, Todd J Hawbaker, Yu Liang, Peng Gong, Zhiwei Wu, and Zhiliang Zhu. A coupled modeling framework for predicting ecosystem carbon dynamics in boreal forests. *Environmental Modelling & Software*, 93:332–343, 2017.

G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis. Analysis and improvement of the consistency of extended Kalman filter based SLAM. In *2008 IEEE International Conference on Robotics and Automation*, pages 473–479. IEEE, 2008.

D. S. Jenkinson, P. B. S. Hart, J. H. Rayner, and L. C. Parry. Modelling the turnover of organic matter in long-term experiments at rothamsted. *INTECOL Bulletin*, 15:1–8, 1987.

A. M. Johansen and A. Doucet. A note on auxiliary particle filters. *Statistics & Probability Letters*, 78(12):1498–1504, 2008.

J. W. Jones, J. Koo, J. B. Naab, W. M. Bostick, S. Traore, and W. D. Graham. Integrating stochastic models and in situ sampling for monitoring soil carbon sequestration. *Agricultural Systems*, 94(1):52–62, 2007.

J. Juston, O. Andrén, T. Kätterer, and P. Jansson. Uncertainty analyses for calibrating a soil carbon balance model to agricultural field trial data in Sweden and Kenya. *Ecological Modelling*, 221(16):1880–1888, 2010.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

J. Koo, W. M. Bostick, J. B. Naab, J. W. Jones, W. D. Graham, and A. J. Gijsman. Estimating soil carbon in agricultural systems using ensemble Kalman filter and DSSAT-Century. *Transactions of the ASABE*, 50(5):1851–1865, 2007.

J. H. Kotecha and P. M. Djuric. Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2592–2601, 2003.

H. R. Künsch. Particle filters. *Bernoulli*, 19(4):1391–1403, 2013.

D. Li Liu, K. Y. Chan, and M. K. Conyers. Simulation of soil organic carbon under different tillage and stubble management practices using the rothamsted carbon model. *Soil and Tillage Research*, 104(1):65–73, 2009.

D. Li Liu, K Y. Chan, M. K. Conyers, G. Li, and G. J. Poile. Simulation of soil organic carbon dynamics under different pasture managements using the RothC carbon model. *Geoderma*, 165 (1):69–77, 2011.

J. S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association*, 93(443):1032–1044, 1998.

L. Ljung. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1):36–50, 1979.

Z. Luo, E. Wang, and O. J. Sun. Soil carbon change and its responses to agricultural practices in Australian agro-ecosystems: a review and synthesis. *Geoderma*, 155(3):211–223, 2010.

Z. Luo, E. Wang, Q. Shao, M. K. Conyers, and D. Li Liu. Confidence in soil carbon predictions undermined by the uncertainties in observations and model parameterisation. *Environmental modelling & software*, 80:26–32, 2016.

H. Orlande, M. Colaço, G. Dulikravich, F. Vianna, W. da Silva, H. da Fonseca, and O. Fudym. Tutorial 10 kalman and particle filters. *Advanced Spring School: Thermal measurements and inverse techniques*, 5:1–39, 2011.

W. J. Parton, J. WB Stewart, and C. V. Cole. Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry*, 5(1):109–131, 1988.

K. I. Paul, P. J. Polglase, and G. P. Richards. Sensitivity analysis of predicted change in soil carbon following afforestation. *Ecological Modelling*, 164(2-3):137–152, 2003.

M. Pitt, R. Silva, P. Giordani, and R. Kohn. Auxiliary particle filtering within adaptive metropolis-hastings sampling. *arXiv preprint arXiv:1006.1914*, 2010.

M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

M. K. Pitt, R. dos Santos Silva, P. Giordani, and R. Kohn. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2): 134–151, 2012.

J. Post, F. F. Hattermann, V. Krysanova, and F. Suckow. Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environmental Modelling & Software*, 23(2):125–138, 2008.

C. P. Robert. The metropolis-hastings algorithm. *arXiv preprint arXiv:1504.01896*, 2016.

F. Robertson and D. Nash. Limited potential for soil carbon accumulation using current cropping practices in victoria, australia. *Agriculture, Ecosystems & Environment*, 165:130–140, 2013.

J. Sanderman and J. A. Baldock. Accounting for soil carbon sequestration in national inventories: a soil scientist's perspective. *Environmental Research Letters*, 5(3):034003, 2010.

J. Sanderman and A. Chappell. Uncertainty in soil carbon accounting due to unrecognized soil erosion. *Global Change Biology*, 19(1):264–272, 2013.

N. Senapati, P. Smith, B. Wilson, J. B. Yeluripati, H. Daniel, P. Lockwood, and S. Ghosh. Projections of changes in grassland soil organic carbon under climate change are relatively insensitive to methods of model initialization. *European Journal of Soil Science*, 64(2):229–238, 2013.

R. H. Shumway and D. S. Stoffer. Time series analysis and its applications. *Studies In Informatics And Control*, 9(4):375–376, 2000.

J. Skjemstad and L. Spouncer. NCAS calibration and verification data. v1. *CSIRO. Data Collection.*, 2003. doi: https://doi.org/10.4225/08/54F0786D6D923.

J. O. Skjemstad, L. R. Spouncer, B. Cowie, and R. S. Swift. Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. *Soil Research*, 42(1):79–88, 2004.

F. E. Stamati, N. P. Nikolaidis, and J. L. Schnoor. Modeling topsoil carbon sequestration in two contrasting crop production to set-aside conversions with RothC–calibration issues and uncertainty analysis. *Agriculture, Ecosystems & Environment*, 165:190–200, 2013.

Y. Wan, E. Lin, W. Xiong, Y. Li, and L. Guo. Modeling the impact of climate change on soil organic carbon stock in upland soils in the 21st century in china. *Agriculture, ecosystems & environment*, 141(1-2):23–31, 2011.

C. K. Wikle and L. M. Berliner. A bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):1–16, 2007.

I. Yildirim. Bayesian inference: Metropolis-Hastings sampling. *Department of Brain and Cognitive Sciences, University of Rochester, Rochester*, 2012.

A MCMC diagnostics of the results presented in section 5

This section presents the trace plots and the Gelman and Rubin's convergence diagnostics for 14 model parameters of the Tarlee data set based on the RB-BPF method in Figure 6 and Table 5, respectively.

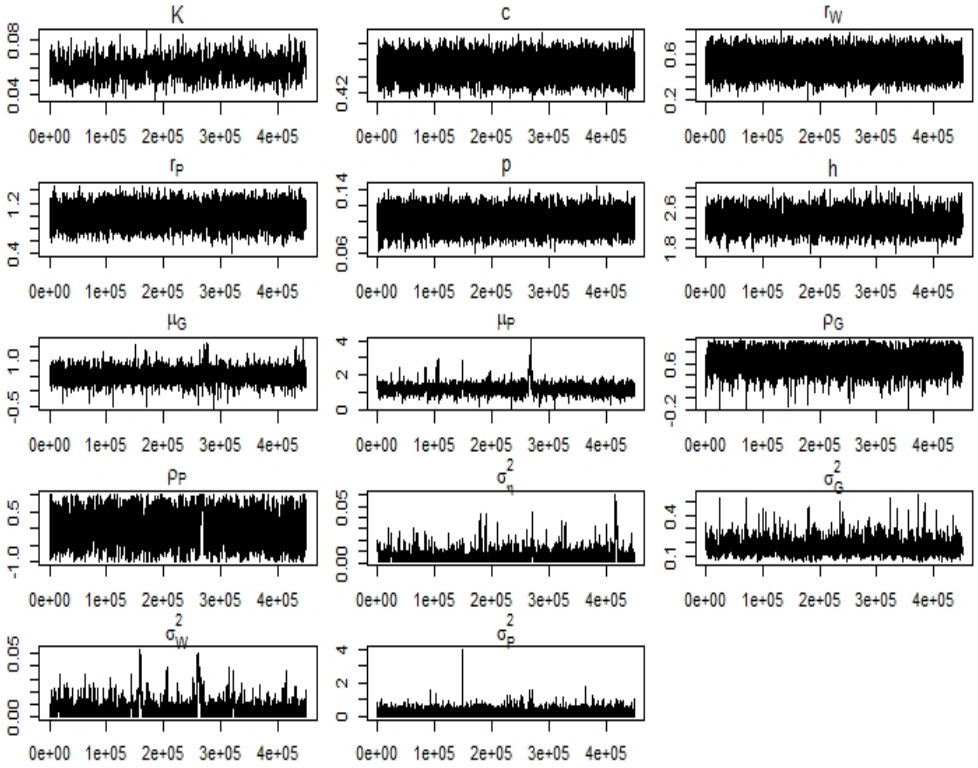

Figure 6: The trace plots-based on the RB-BPF method of 14 model parameters that presented in Table 2.

Parameter	$\hat{R}$	Upper C.I. bound on $\hat{R}$
K	1.01	1.02
c	1	1
r_W	1	1
r_P	1	1
p	1	1
h	1	1.01
μ_G	1.01	1.01
μ_P	1.11	1.17
ρ_G	1.01	1.01
ρ_P	1.01	1.03
σ^2_η	1.04	1.06
σ^2_G	1.01	1.02
σ^2_W	1.05	1.09
σ^2_P	1.08	1.09

Table 5: The Gelman and Rubin's convergence diagnostic, $\hat{R}$ calculated for the 14 model parameters presented in Table 2. Since the point estimate of $\hat{R}$ for each parameter is less than 1.2, the MCMC samples can be considered to have reached a stationary distribution and are mixing adequately.