CALIBRATION OF THE STOCHASTIC MULTICLOUD MODEL USING BAYESIAN INFERENCE

MICHÈLE DE LA CHEVROTIÈRE*, BOUALEM KHOUIDER†, AND ANDREW J. MAJDA‡

Abstract. The stochastic multicloud model (SMCM) was recently developed (Khouider, Biello, and Majda, 2010) to represent the missing variability in general circulation models due to unresolved features of organized tropical convection. This research aims at finding a robust calibration methodology for the SMCM to estimate key model parameters from data. We formulate the calibration problem within a Bayesian framework to derive the posterior distribution over the model parameters. The main challenge here is due to the likelihood function which requires solving a large system of differential equations (the Kolmogorov equations) as many times as there are data points, which is prohibitive both in terms of computation time and storage requirements. The most attractive numerical techniques to compute the transient solutions to large Markov chains are based on matrix exponentials, but none is unconditionally acceptable for all classes of problems. We develop a parallel version of a preconditioning technique known as the Uniformization Method, using the `PETSc` (Portable, Extensible Toolkit for Scientific Computation) suite of sparse matrix-vector operations. The parallel Uniformization Method allows for fast and scalable approximations of large sparse matrix exponentials, without sacrificing accuracy. Sampling of the high dimensional posterior distribution is achieved via the standard Markov Chain Monte Carlo. The robustness of the calibration procedure is tested using synthetic data produced by a simple toy climate model. A sensitivity study to the length of the data time series and to the prior distribution is presented, and a sequential learning strategy is also tested.

Key words. stochastic cumulus parameterization, inverse problem, Bayesian inference, large sparse matrix exponential, parallel Uniformization Method, high performance computing, `PETSc`, Monte Carlo Markov Chain, climate models

AMS subject classifications. 81T80, 65F30, 65Y05

1. Introduction.

General circulation models (GCMs) are mathematical models based on a careful discretization of the Navier–Stokes equations which are used to simulate the coupled circulation of the planetary oceans and atmosphere. Because of the finite resolution of numerical models, there are always physical processes and scales of motion that cannot be represented directly on the underlying grid. Inclusion of processes such as boundary layer fluxes, vertical mixing due to convection, turbulent mixing, formation of clouds and precipitation, and the interaction of clouds and radiation fluxes requires that the relevant subgrid scale processes be represented in terms of grid-level variables. The approximation of unresolved processes in terms of resolved variables is referred to as the *parameterization problem*. For atmospheric dynamics, the most important physical process that must be parameterized is moist convection in the tropics. In a typical GCM with grid spacing of 10 to 100 km, the cumulus updrafts and downdrafts are not resolved by the model grid. Cloud processes affect the climate system by regulating the radiation budget at the top of the atmosphere, by producing precipitation, and by transporting and redistributing water vapour in the atmosphere [3, 2].

*Department of Mathematics and Statistics, University of Victoria, PO BOX 3060 STN CSC, Victoria, BC, Canada V8W 3P4, Tel. 250-853-3293, Fax. 250-721-8962, `delachev@math.uvic.ca`

†Department of Mathematics and Statistics, University of Victoria, PO BOX 3060 STN CSC, Victoria, BC, Canada V8W 3P4, Tel. 250-721-7439, Fax. 250-721-8962, `khouider@math.uvic.ca`

‡ Department of Mathematics, Center for Atmosphere-Ocean Science, Courant Institute, and, Center for Prototype Climate Modeling, NYU Abu Dhabi Institute, New York University, 251 Mercer Street, New York, NY 10012 USA, Tel.: 212-998-3234, Fax: 212-995-3323, `jonjon@cims.nyu.edu`

However, state-of-the-art GCMs used for climate and mid-range weather predictions represent poorly the dominant features of atmospheric variability at intra-seasonal (40 to 60 days) and planetary and synoptic scales in the tropics, namely the Madden-Julian oscillation (MJO) and convectively coupled Kelvin waves (CCKWs) [24]. Both the MJO and CCKWs are believed to play a fundamental role in regulating the weather and climate in the tropics and extratropics [25].

There is a general consensus in the climate community that this deficiency is due mainly to the inadequate treatment of cumulus convection and the associated interactions across multiple temporal and spatial scales [25, 42, 24]. The search for new strategies for parameterizing the unresolved effects of tropical convection has been ongoing for the last few decades, starting with the moist convective adjustment idea of Manabe et al. in 1965 [31]. Later in 1974, Kuo introduced a closure based on the large-scale moisture convergence [27]. The same year, Arakawa and Schubert [3] formulated a parameterization based on the quasi-equilibrium assumption, which assumes a large separation between the subgrid- and grid-scales processes. While these models constitute the benchmark of parameterizations used in GCMs today (see for instance [46]), their purely deterministic closures do not capture the highly intermittent and organized tropical convection [36]. Since then, many improvements in GCMs came with the development of stochastic parameterizations [28, 29], which encode model uncertainties as probability distributions and derive convection statistics to represent the dynamics.

The stochastic multicloud model (SMCM) for tropical convection introduced by Khouider et al. [19] has shown to improve the intermittency in coherent structures that contribute to higher variability; when combined with a simple two-layer atmospheric climate model, the SMCM was shown to reasonably simulate tropical convection and improve the associated wave-like features [19, 10, 11]. However, the choice of key parameters in the SMCM is so far based solely on physical intuition and rough estimates obtained from idealized numerical simulations and/or ad hoc processing of observational data. Our main goal here is to develop a rigorous statistical method to infer some of these parameters systematically from observational and/or detailed cloud resolving model [17] time series using a Bayesian framework.

The SMCM is essentially a multi-dimensional Markov birth-death process [34] with immigration. Provided that reliable observational or numerical data exist, the main challenges of the Bayesian methodology for the SMCM include the efficient computation of the model likelihood function. The likelihood function involves solving a large system of Kolmogorov equations as many times as there are observed data, which is computationally prohibitive in a statistical inference setting. Several numerical techniques exist for computing transient solutions of large Markov chains, as detailed in Sidge [44], but none is satisfactory in all contexts. The Uniformization and Krylov-based Methods, which are based on the evaluation of matrix exponentials, are the front runners of such techniques. The Krylov-based Method, evidenced in the case studies [8, 5] due to its performance and robustness, has no reliable stopping criteria which makes it impractical to use in our case. By contrast, the convergence of the Uniformization Method is determined with *a priori error bounds*, and works really well in practice. Nonetheless, the method seems to suffer from numerical instability and performance degradation in some cases [8, 44]. Therefore, it remains interesting to see how it will perform for our inference problem where the likelihood function it approximates is computed so many times. On top of these problems, for large matrices (the size of our typical

problem is on the order $10^9 \times 10^9$), a careful implementation is needed to limit storage space and improve efficiency. To overcome these issues, we develop a parallel version of the Uniformization Method using sparse matrix–vector operations only, which is facilitated using the `PETSc` suite [4]. We use the standard Markov Chain Monte Carlo (MCMC) technique to sample the associated high dimensional Bayesian posterior distribution. The SMCM inference problem is an ideal benchmark test of the Uniformization Method's performance and reliability on a parallel platform.

As a proof of the inference methodology, we use synthetic data produced by the SMCM coupled to a toy GCM single column model [19]. We thus aim mainly at reproducing the assumed values of some key parameters of the SMCM from time series produced by the coupled toy GCM-SMCM model.

Will the inferred parameters be statistically close to the assumed values? How long and how sparse the time series should be in order to reasonably retrieve these parameters? If the toy GCM-SMCM is re-integrated using the inferred parameters instead, will the resulting time series have the same statistics as the original one? These are the key scientific questions we aim to answer here before we can use the Bayesian framework with real and/or cloud resolving model data.

The rest of the paper is organized as follows. In Section 2, we recall the main features of the SMCM and in Section 3, we present the Bayesian inference model highlighting both the numerical approximation of the likelihood function and the MCMC sampling strategy. The key validation results using synthetic data are presented in Section 4, where we attempt to answer the above questions. A concluding discussion is given in Section 5.

2. The Stochastic Multicloud Model. In this section we briefly review the dynamical and physical features of the SMCM parameterization that are relevant to the Bayesian set-up. A more complete discussion of the stochastic multicloud framework is found in the original papers [19, 22, 23, 20, 21, 10].

Radar and satellite data combined with local soundings and aircraft measurements etc. [16, 32] have provided strong evidence that large-scale tropical convective systems involve three main cloud populations: shallow/congestus, deep penetrative cumulus clouds, and stratiform clouds. Congestus cloud decks, with a vertical extent that does not exceed the freezing level (5 or 6 km), are followed by deep convective towers that extend to the top of the troposphere, which in turn are lagged by stratiform anvils in their dissipation phase (see Figure 2.1). Congestus clouds heat the lower troposphere due to condensational heating and induce upper troposphere cooling because of detrainment at their tops and by blocking long wave radiation from the surface. Deep convective towers dominate the core of the storm and are believed to be responsible for most of the tropical rainfall and provide the bulk heating for the whole tropospheric column. Stratiform anvil clouds finally heat the upper troposphere and cool the lower troposphere due to the evaporation of stratiform rain [23].

All three clouds are important components of the tropical convective cloud spectrum and are associated with trimodal distributions of heating profiles, divergence, cloud detrainment, and fractional cloudiness. Congestus clouds prevail in front of the wave where the atmosphere is dry and serve to moisten and precondition the environment for deep convection due to both detrainment of cloud water and lower-level large scale convergence of moisture induced by congestus precipitation [16].

Accordingly, the multicloud parameterization framework assumes three heating profiles associated with the main cloud types that characterize organized tropical convective systems.

The SMCM mimics the convection activity over a single GCM grid box using a square $n \times n$ lattice, in which individual lattice sites correspond to convective cells. Each lattice cell i is associated with a four-state Markov process $(Y_t^i)_{t>0}$ taking value 0, 1, 2, or 3, depending on whether it is clear sky, or occupied by a congestus, deep, or stratiform cloud, respectively. In the simple case where local interactions are ignored, all $N = n \times n$ stochastic processes $(Y_t^i)_{t>0}$ are independent and identically distributed [19]. This allows the derivation in a straightforward fashion of a coarse-grained birth-death Markov process that evolves efficiently the array of area fractions of the three cloud types, in each GCM grid box, without the detailed knowledge of the microscopic lattice configuration. A more general coarse-graining strategy that allows nearest neighbour interactions in the SMCM is presented in [18]. For the sake of simplicity, here we consider only the case without local interactions.

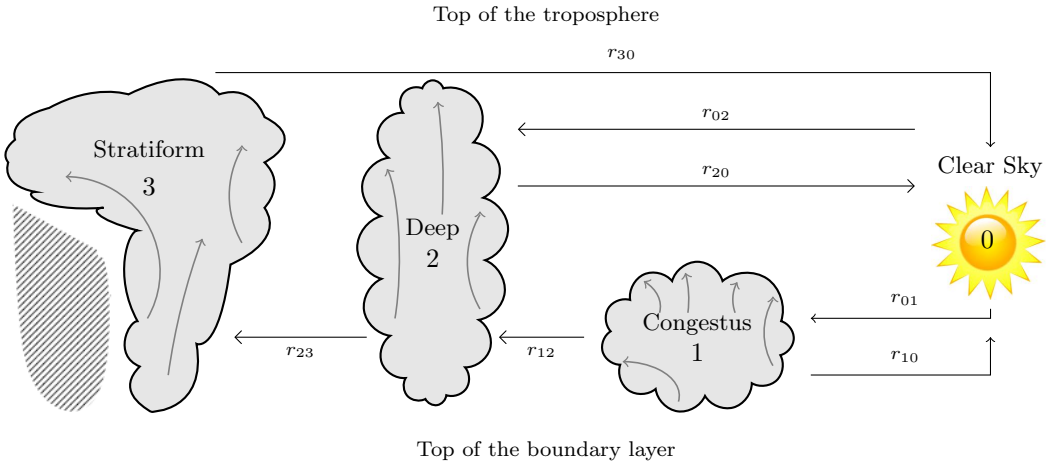


FIG. 2.1. *A cartoon of the three cloud types showing congestus, deep convective, and a decaying deep convective tower with a lagging large stratiform anvil, with stratiform rain falling into a dry region below it where it eventually evaporates and cools the environment (hatched area). The probability transition rates between the different clouds and clear sky state are given as functions r_{kl} of the large-scale variables C, C_l, and D.*

FIG. 2.2. *$n \times n$ Lattice Cloud Model*

0	1	3	2	2	1
2	2	1	0	0	3
0	0	3	0	0	3
1	3	1	0	1	3
2	2	1	0	0	2
1	0	3	1	0	3

GCM grid box: $\mathcal{O}(100km)$

TABLE 2.1
Large-scale variables

C	convective available potential energy (CAPE)
C_l	low-level CAPE
D	middle troposphere dryness (or moisture)

In the case without local interactions, each process $(Y_t^i)_{t>0}$ evolves by allowing transitions between states, based on some intuitive interaction rules that depend only on the large-scale resolved variables, in accordance with observations of cloud dynamics in the tropics. These large-scale variables are the convective available potential energy (CAPE) integrated over the whole troposphere (C), the convective available potential energy integrated over the lower troposphere (C_l), and the dryness of the midtroposphere (D). The interaction rules between the different cloud types and the environment are summarized as follows [19]:

1. A clear site turns into a congestus site with high probability if low level CAPE is positive and the middle troposphere is dry;
2. A congestus or clear sky turns into a deep convective site with high probability if CAPE is positive and the middle troposphere is moist;
3. A deep convective site turns into a stratiform site with high probability;
4. All three cloud types decay naturally to clear sky at some fixed rate;
5. All other transitions are assumed to have negligible probability.

These rules are formalized by the probability transition rates r_{kl} and the associated time scales τ_{kl} listed in Table 2.2, in terms of the activation function

$$\Gamma(x) = \left\{1 - e^{-x} \text{ if } x > 0,\ 0 \text{ otherwise}\right\}.$$

TABLE 2.2
Transition rates and timescales in the stochastic parameterization.

Cloud Transition	Probability Transition Rate	Cloud Transition Timescale (hours)
Formation of congestus	$r_{01} = \frac{1}{\tau_{01}}\Gamma(C_l)\Gamma(D)$	$\tau_{01} = 1$
Decay of congestus	$r_{10} = \frac{1}{\tau_{10}}\Gamma(D)$	$\tau_{10} = 1$
Conversion of congestus to deep	$r_{12} = \frac{1}{\tau_{12}}\Gamma(C)\big(1 - \Gamma(D)\big)$	$\tau_{12} = 0.25$
Formation of deep	$r_{02} = \frac{1}{\tau_{02}}\Gamma(C)\big(1 - \Gamma(D)\big)$	$\tau_{02} = 3$
Conversion of deep to stratiform	$r_{23} = \frac{1}{\tau_{23}}$	$\tau_{23} = 3$
Decay of deep	$r_{20} = \frac{1}{\tau_{20}}\big(1 - \Gamma(C)\big)$	$\tau_{20} = 2$
Decay of stratiform	$r_{30} = \frac{1}{\tau_{30}}$	$\tau_{30} = 5$
Activation Function	$\Gamma(x) = 1 - e^{-x}, \quad x \geq 0$	

Note from Assumption 5, that we have $r_{03} = r_{13} = r_{21} = r_{31} = r_{32} = 0$. Each one of

the N Markov chains can be simulated with an accept-reject Monte Carlo algorithm, using the fact that the sojourn time in any state of a Markov process is an exponential random variable. However this is very costly. As already stated, a much cheaper, coarse-grained version of the model is constructed out of the cloud populations

$$N_c^t = \sum_{i=1}^{N} \mathbb{1}_{\{Y_t^i=1\}}, \quad N_d^t = \sum_{i=1}^{N} \mathbb{1}_{\{Y_t^i=2\}}, \quad N_s^t = \sum_{i=1}^{N} \mathbb{1}_{\{Y_t^i=3\}},$$

where N_c^t, N_d^t, and N_s^t are the lattice total number of congestus, deep, and stratiform at time t, and $\mathbb{1}$ is the indicator function. By conservation of the total number of sites, the population of clear sky sites at time t is given by $N_{cs}^t = N - N_c^t - N_d^t - N_s^t$, where N is the size of the lattice. The evolution of the cloud populations N_c, N_d, N_s effectively constitutes a birth-death process with immigration. Note that the actual area fractions (cloud cover on a GCM grid box) are given by $\sigma_c = N_c/N$, $\sigma_d = N_d/N$, $\sigma_s = N_s/N$.

We now briefly describe the three-dimensional birth-death process $(\boldsymbol{X}_t)_{t>0}$, $\boldsymbol{X}_t = (N_c^t, N_d^t, N_s^t)$, and give its associated probability distributions in Section 3. Let $\boldsymbol{i} = (i_1, i_2, i_3)$ and $\boldsymbol{j} = (j_1, j_2, j_3)$ be triplets of non-negative integers in the range space $\mathcal{S}$ of $(\boldsymbol{X}_t)_{t>0}$. Then the conditional probability that at time t the congestus, deep and stratiform populations are respectively j_1, j_2, and j_3, given that at time $t = 0$ there were i_1 congestus, i_2 stratiform, and i_3 deep clouds is denoted by

$$P_{\boldsymbol{ij}}(t) = P\{\boldsymbol{X}_t = \boldsymbol{j} | \boldsymbol{X}_0 = \boldsymbol{i}\}.$$

The transition probabilities $P_{\boldsymbol{ij}}(t)$ satisfy the initial condition

$$P_{\boldsymbol{ij}}(0) = \delta_{i_1 j_1} \cdot \delta_{i_2 j_2} \cdot \delta_{i_3 j_3},$$

where $\delta_{\alpha\beta}$ is the Kronecker delta, and we assume that $P_{\boldsymbol{ij}}(t)$ are differentiable functions of t for $t > 0$.

Transitions $\boldsymbol{i} \to \boldsymbol{j}$ can occur as single births (with the clear sky population N_{cs} losing one site), deaths (with the clear sky population N_{cs} gaining one site), or immigrations (when a congestus becomes a deep cloud, or a deep cloud becomes a stratiform), via rates related to those described in Table 2.2. We introduce furthermore the standard unit vectors $\boldsymbol{\varepsilon_1} = (1,0,0)$, $\boldsymbol{\varepsilon_2} = (0,1,0)$, and $\boldsymbol{\varepsilon_3} = (0,0,1)$. Then the admissible transitions from state $\boldsymbol{i}$ are given as the following model postulates (transition in parentheses) [19]:

$$\begin{aligned}
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} - \boldsymbol{\varepsilon_1} + \boldsymbol{\varepsilon_2} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{12}h + o(h) \quad \text{(congestus to deep)},\\
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} - \boldsymbol{\varepsilon_2} + \boldsymbol{\varepsilon_3} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{23}h + o(h) \quad \text{(deep to stratiform)},\\
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} - \boldsymbol{\varepsilon_1} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{10}h + o(h) \quad \text{(congestus to clear sky)},\\
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} - \boldsymbol{\varepsilon_2} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{20}h + o(h) \quad \text{(deep to clear sky)},\\
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} - \boldsymbol{\varepsilon_3} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{30}h + o(h) \quad \text{(stratiform to clear sky)},\\
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} + \boldsymbol{\varepsilon_1} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{01}h + o(h) \quad \text{(clear sky to congestus)},\\
P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} + \boldsymbol{\varepsilon_2} | \boldsymbol{X}_t = \boldsymbol{i}\} &= R_{02}h + o(h) \quad \text{(clear sky to deep)},
\end{aligned} \tag{2.1}$$

where h is a small increment, and $R_{kl} = i_k r_{kl}$ $(k, l = 0, 1, 2, 3)$, with the rates r_{kl} depending upon the exogenous factors C, C_l, D as listed in Table 2.2. Moreover, the probability of a transition other than those listed in (2.1) in a time interval $(t, t+h)$ is $o(h)$, and the probability of no transition in $(t, t+h)$ is

$$P\{\boldsymbol{X}_{t+h} = \boldsymbol{i} | \boldsymbol{X}_t = \boldsymbol{i}\} = 1 - h\big(R_{10} + R_{01} + R_{02} + R_{12} + R_{20} + R_{23} + R_{30}\big) + o(h). \tag{2.2}$$

3. The Bayesian Inference Model. The SMCM calculates the evolution of the cloud populations $\boldsymbol{x} = (N_c, N_d, N_s)$[1] constrained by the large-scale atmospheric state $\boldsymbol{u} = (C, C_l, D)$. We label the corresponding sequence of observations $x_1, x_2, x_3, \ldots$ and $u_1, u_2, u_3, \ldots$ by $\boldsymbol{x}_t$ and $\boldsymbol{u}_t$, respectively. The parameterization includes seven numerical inputs (or parameters), namely the cloud convective timescales (see Table 2.2), which we stack in the vector

$$\boldsymbol{\theta} = (\tau_{01}, \tau_{10}, \tau_{12}, \tau_{02}, \tau_{23}, \tau_{20}, \tau_{30}).$$

The relationship between the parameters, large-scale variables, and model output can be represented by the mapping $\boldsymbol{x}_t = g(\boldsymbol{\theta}, \boldsymbol{u}_t)$, where g is a function that represents the SMCM coupled to a climate model. The parameters, or inputs $\boldsymbol{\theta}$, can be 'tuned' to find $\boldsymbol{\theta^*}$, the 'best' input configuration, so that the model reproduces some observed data well. The role of data is to help us *learn* which parameter values best simulate the climate we know. Learning which choice of parameter values lead the given model to best reproduce the climate is the process of *calibration*. The SMCM calibration problem should be viewed as an *inverse problem*: given the SMCM/climate model (represented by the function g), the climate (cloud) data $\boldsymbol{x}_t$, and the large-scale external factors $\boldsymbol{u}_t$, find the best parameter input values $\boldsymbol{\theta^*}$ so that $\boldsymbol{x}_t^* = g(\boldsymbol{\theta^*}, \boldsymbol{u}_t)$ is statistically close to $\boldsymbol{x}_t$.

In the inversion process, uncertainties arise from imperfect and finite data, and from the assumption that the model is a true representation of the climate dynamics. Data uncertainty propagates through the reductive model to give us uncertainty on the parameters $\boldsymbol{\theta}$. Additionally, inverse problem are often ill-posed in the sense that various parameters $\boldsymbol{\theta}$ can relate to the same input data set, or that the parameters $\boldsymbol{\theta}$ may not depend continuously on the data.

The Bayesian statistical approach provides a solution by formulating a complete probabilistic description of the unknowns and uncertainties given the data. It incorporates the initial information and residual uncertainty about the model parameters $\boldsymbol{\theta}$ into a *prior* distribution $\pi(\boldsymbol{\theta})$, which is then updated by a model *likelihood* function $f(\boldsymbol{x}_t|\boldsymbol{\theta})$ to formulate a *posterior* distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x}_t)$ of the model parameters given the data [38]. Hence it does not find a single best-fit parameter values configuration $\boldsymbol{\theta^*}$ but a distribution of solutions $\pi(\boldsymbol{\theta}|\boldsymbol{x}_t)$, informed by the data $\boldsymbol{x}_t$.

The inversion of probabilities is given by *Bayes' Theorem* and finds the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x}_t)$ as a consequence of the two antecedents $\pi(\boldsymbol{\theta})$ and $f(\boldsymbol{x}_t|\boldsymbol{\theta})$ [38]:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_t) = \frac{f(\boldsymbol{x}_t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{x}_t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

Note that $\pi(\boldsymbol{\theta}|\boldsymbol{x}_t)$ is actually proportional to the distribution of $\boldsymbol{x}_t$ conditional upon $\boldsymbol{\theta}$, i.e. the likelihood, multiplied by the prior distribution on $\boldsymbol{\theta}$:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_t) \propto f(\boldsymbol{x}_t|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Conditioning further on $\boldsymbol{u}_t$ we obtain:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_t, \boldsymbol{u}_t) \propto f(\boldsymbol{x}_t|\boldsymbol{u}_t, \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{3.1}$$

[1] Here observations of the random variable $\boldsymbol{X}$ are written in lower case.

In this Bayesian context, the large-scale variables $\boldsymbol{u}_t$ are seen as covariates to the model and we may think of them as having Dirac delta distributions.

We now consider series of observations $\boldsymbol{x}_{1:T} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ and $\boldsymbol{u}_{1:T} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)$ of length T. By conditioning on past events, the likelihood is effectively factorized into a product of T *one-step transition likelihoods*:

$$f(\boldsymbol{x}_{1:T}|\boldsymbol{u}_{1:T}, \boldsymbol{\theta}) = \prod_{t=1}^{T} f_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_{t-1}, \boldsymbol{\theta}). \tag{3.2}$$

Here we resorted to the Markov property of the probabilistic multicloud model to exclude old events from the representation.

3.1. Numerical Approximation of the Likelihood Function. The transition probability matrix of the continuous time birth-death process $(\boldsymbol{X}_t)_{t>0}$ defined by the probabilities (2.1)-(2.2) solves the so-called system of Kolmogorov *backward* differential equations [9]:

$$\begin{aligned}\frac{dP_{\boldsymbol{ij}}(t)}{dt} &= R_{12}P_{\boldsymbol{i}-\boldsymbol{\varepsilon_1}+\boldsymbol{\varepsilon_2},\boldsymbol{j}}(t) + R_{23}P_{\boldsymbol{i}-\boldsymbol{\varepsilon_2}+\boldsymbol{\varepsilon_3},\boldsymbol{j}}(t) + R_{10}P_{\boldsymbol{i}-\boldsymbol{\varepsilon_1},\boldsymbol{j}}(t)\\ &\quad + R_{20}P_{\boldsymbol{i}-\boldsymbol{\varepsilon_2},\boldsymbol{j}}(t) + R_{30}P_{\boldsymbol{i}-\boldsymbol{\varepsilon_3},\boldsymbol{j}}(t) + R_{01}P_{\boldsymbol{i}+\boldsymbol{\varepsilon_1},\boldsymbol{j}}(t) + R_{02}P_{\boldsymbol{i}+\boldsymbol{\varepsilon_2},\boldsymbol{j}}(t)\\ &\quad - \big(R_{12} + R_{23} + R_{10} + R_{20} + R_{30} + R_{01} + R_{02}\big)P_{\boldsymbol{ij}}(t),\end{aligned}$$

with the initial conditions $P_{\boldsymbol{ij}}(0) = \delta_{\boldsymbol{ij}}$. Let $P = \{P_{\boldsymbol{ij}}(t)\} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ be the matrix of transition probability functions. We may cast the Kolmogorov system in its matrix form:

$$\begin{aligned}P'(t) &= R(\boldsymbol{u}_t, \boldsymbol{\theta})P(t),\\ P(0) &= Id,\end{aligned} \tag{3.3}$$

where Id is the identity matrix of order $|\mathcal{S}|$, and $R \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ is the matrix of transition rates R_{kl} (the infinitesimal generator of the birth-death process). On the time interval $[t, t + \Delta t]$, during which the large-scale variables $\boldsymbol{u}_t$ may be assumed constant, the solution to the system (3.3) is approximately given by

$$P(s) = \exp\big[R(\bar{\boldsymbol{u}}_t, \boldsymbol{\theta})s\big], \quad s \in [t, t + \Delta t] \tag{3.4}$$

for some fixed model parameter values $\boldsymbol{\theta}$ and constant values $\bar{\boldsymbol{u}}_t$ of the large-scale variables. The one-step transition likelihoods in (3.2) are merely the density functions associated with the probability matrix entries (3.4), as functions of $\boldsymbol{\theta}$. The computation of the likelihood function (3.2) in full requires repeated solves of the forward problem, more precisely $T - 1$ large matrix exponentials for an observed sample of length T.

Large Matrix Exponential. How large the matrix R is depends on the size of the cloud lattice. Given a lattice of size N, the stochastic process $(\boldsymbol{X}_t)_{t>0}$ evolves in a finite space $\mathcal{S} \subset \mathbb{N}^3$, where $\mathcal{S}$ is the set of all ordered triplets of nonnegative integers (a, b, c) satisfying the relation $a+b+c \leq N$. The geometrical domain associated with $\mathcal{S}$ is illustrated in Figure 3.1(a).

The state space $\mathcal{S}$ is countable, so we may find an ordering formula $\phi : \mathcal{S} \to \mathbb{N}$ for the triples in $\mathcal{S}$ (in practice, ϕ is needed to construct the large matrix R incrementally). One

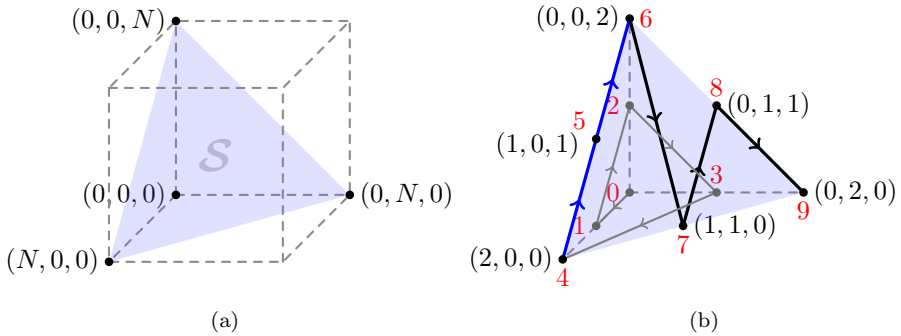


FIG. 3.1. *(left) The state space $\mathcal{S}$ of the immigration-birth-death process $(\boldsymbol{X}_t)_{t>0}$ is the subset of $\mathbb{N}^3$ of all ordered triples of nonnegative integers (a, b, c) that lie below and on the plane $a + b + c = N$. (right) Sequence of triples given by the counting function ϕ (in red) for $N = 2$. The stride $b = 0$ for the plane $d = 2$ is colored in blue.*

such injection is given by the mapping

$$\begin{aligned}\phi(d,b,c) &= \underbrace{\sum_{s=0}^{d-1}\sum_{b=0}^{s}(s-b+1)}_{\text{rank given plane } d} + \underbrace{\sum_{s=0}^{b-1}(d-s+1)}_{\text{rank given stride } b} + c, \\ &= \frac{d^3}{6} + \frac{d^2}{2} + \frac{d}{3} + db - \frac{b^2}{2} + \frac{3b}{2} + c, \end{aligned} \tag{3.5}$$

for which we set $a + b + c = d$ for $0 \le d \le N$, $0 \le b \le d$, and $0 \le c \le d - b$. The counting increment is determined first by planes, then by strides within those planes. The function ϕ maps each triple in $\mathcal{S}$ to a counting order (address). An illustration of the mapping is given in Figure 3.1(b). We may now write the likelihood (3.2) as

$$\begin{aligned} f(\boldsymbol{x}_{1:T}|\boldsymbol{u}_{1:T},\boldsymbol{\theta}) &= \prod_{t=1}^{T} f_{t-1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{u}_{t-1},\boldsymbol{\theta}) \\ &= \prod_{t=1}^{T} \mathbb{1}_{\{\phi(\boldsymbol{d}^{t-1},\boldsymbol{N}_d^{t-1},\boldsymbol{N}_s^{t-1})\}}{}^{*} \exp[R(\boldsymbol{u}_{t-1},\boldsymbol{\theta})h]\mathbb{1}_{\{\phi(\boldsymbol{d}^{t},\boldsymbol{N}_d^{t},\boldsymbol{N}_s^{t})\}}, \end{aligned} \tag{3.6}$$

where $\boldsymbol{x}_t = (N_c^t, N_d^t, N_s^t)$, $d^t = N_c^t + N_d^t + N_s^t$, h is the sampling time interval, and $\mathbb{1}_{\{\phi(\cdot)\}}$ is a vector in $\mathbb{R}^{|\mathcal{S}|}$ that has 1 at the index corresponding to $\phi(\cdot)$, and 0's everywhere else. Here $\mathbb{1}_{\{\phi(\cdot)\}}{}^{*}$ is the transpose of $\mathbb{1}_{\{\phi(\cdot)\}}$, and ϕ is given by (3.5).

From the counting formula (3.5), we can get the dimension of the infinitesimal generator R, i.e. the cardinality of $\mathcal{S}$, as the last element in the counting sequence:

$$|\mathcal{S}| = \phi(N, N, 0) + 1 = \frac{N^3}{6} + N^2 + \frac{11}{6}N + 1 = \dim(R).$$

We find that $\dim(R) = \mathcal{O}(N^3)$, so the size and memory requirements of R become prohibitively large with the dimensions of the cloud lattice. For instance, for a 10×10 lattice

($N = 100$) the size of R is 176,851, which requires about 70 Mb of storage in the compressed standard `PETSc` AIJ format. The memory requirement escalates to about 50 Gb for a 30×30 lattice ($N = 900$) and correspondingly, $\dim(R) = 122,311,651$. But although R is large, its density (fraction of non-zero elements) is quite low: $4.52 \times 10^{-3}\%$ for $N = 100$ and $6.54 \times 10^{-6}\%$ for $N = 900$. Sparse matrix compression alone is however not sufficient as indicated by the memory storage numbers above, so further scalability is gained using distributed memory. We achieved considerable performances (in matrix assembly and operations) using distributed sparse `PETSc` [4] matrices and related routines.

Uniformization Method. We are led to consider the problem of computing the exponential of a large sparse matrix. Since the matrix exponential is dense even when the matrix R is sparse, the computation of $\exp(Rt)$ in full (based on matrix-matrix operations) remains possible only when R is relatively small. These include Padé-type or matrix decompositions [35]. For large-scale problems, the family of series methods, based on matrix-vector products, are preferable [44]. In Markov chain modelling the use of Jensen's *Uniformization Method* [14, 44] is widespread, but an alternative technique based on *Krylov subspaces* [44, 43] seems to rival in performance. We implemented both methods in parallel and compared their applicability and performance in extensive numerical tests that will be reported in the future. The Krylov-based Method is an iterative technique which comes with an important caveat: there exists no practical and satisfactory stopping criterion for the method. While the performance results in [44] are obtained by fixing the size of the Krylov basis beforehand, Saad's [41] residual-based stopping criterion used in [5, 8] fails the convergence test in our case. So we resorted to the Uniformization Method for which there exists an established stopping criterion, and which revealed to be both reliable and accurate in our series of numerical tests.

The Uniformization Method is based on the partial Taylor series expansion of the matrix exponential:

$$\boldsymbol{w}(t) := \exp(Rt)\boldsymbol{e_j} \approx \sum_{k=0}^{p} \frac{t^k}{k!} R^k \boldsymbol{e_j}, \tag{3.7}$$

where $\boldsymbol{e_j}$ is the standard canonical vector in $\mathbb{R}^{|\mathcal{S}|}$. However, because R is essentially nonnegative (the diagonal elements of R are negative and the off-diagonal elements are nonnegative), a direct use of (3.7) leads to severe roundoff errors in finite floating point arithmetic due to catastrophic cancellation. For numerical stability, the series expansion (3.7) is combined with the preconditioner $Q = \frac{1}{\alpha} R + Id$, where $\alpha = \max_i |R_{ii}|$. It follows that Q is a stochastic matrix, that is its entries satisfy $\sum_j Q_{ij} = 1$, and $Q_{ij} \geq 0$. Then the truncated approximation

$$\boldsymbol{w_p}(t) := e^{-\alpha t} \sum_{k=0}^{p} \frac{(\alpha t)^k}{k!} Q^k \boldsymbol{e_j}, \tag{3.8}$$

involves nonnegative terms only and is numerically stable. Using the fact that $\|Q\|_\infty = 1$, it is easy to show that the error of the approximation (3.8) is such that [14, 44]

$$\|\boldsymbol{w}(t) - \boldsymbol{w_p}(t)\|_\infty \leq 1 - e^{-\alpha t} \sum_{k=0}^{p} \frac{(\alpha t)^k}{k!}. \tag{3.9}$$

The a priori bound (3.9) on the series truncation error is used as a stopping criterion for the Uniformization Method: If ϵ_{tol} is a prescribed error tolerance, then truncate the series at order p_ϵ, where p_ϵ is the smallest integer that satisfies

$$1 - e^{-\alpha t} \sum_{k=0}^{p_\epsilon} \frac{(\alpha t)^k}{k!} \leq \epsilon_{tol} \Leftrightarrow \sum_{k=0}^{p_\epsilon} \frac{(\alpha t)^k}{k!} \geq e^{\alpha t}(1 - \epsilon_{tol}).$$

As such, the Uniformization Method is easy to implement numerically. We used the `PETSc` [4] parallel suite for distributed sparse matrix–vector products. A tolerance error $\epsilon_{tol} = 10^{-10}$ was set for all numerical results reported herein.

3.2. Posterior Sampling Using Monte Carlo Markov Chain Algorithm. In the Bayesian paradigm, all inferences about the model parameters $\boldsymbol{\theta}$ are carried out based on the posterior distribution (equations (3.1) and (3.6))

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:T}, \boldsymbol{u}_{1:T}) \propto \pi(\boldsymbol{\theta}) \prod_{t=1}^{T} \mathbb{1}_{\{\boldsymbol{\phi}(\boldsymbol{d}^{t-1}, \boldsymbol{N}_{\boldsymbol{d}}^{t-1}, \boldsymbol{N}_{\boldsymbol{s}}^{t-1})\}}{}^{*} \exp[R(\boldsymbol{u}_{t-1}, \boldsymbol{\theta})h] \mathbb{1}_{\{\boldsymbol{\phi}(\boldsymbol{d}^{t}, \boldsymbol{N}_{\boldsymbol{d}}^{t}, \boldsymbol{N}_{\boldsymbol{s}}^{t})\}},$$

for a suitable choice of the prior $\pi(\boldsymbol{\theta})$. Bayesian point-estimators of interest, like the posterior mean $\mathbb{E}_\pi[\boldsymbol{\theta}] = \int_\Theta \boldsymbol{\theta}\, \pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:T}, \boldsymbol{u}_{1:T}) d\boldsymbol{\theta}$, require to evaluate an integral on a parameter space Θ of dimension 7. A classical approximation method for complex multidimensional integrals is the Monte Carlo Markov Chain (MCMC) technique [40]. The underlying idea of MCMC is to construct a Markov chain in $\boldsymbol{\theta}$ with ergodic (stationary) distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:T}, \boldsymbol{u}_{1:T})$, which is guaranteed under the conditions of irreducibility and aperiodicity of the chain. Starting with some initial state $\boldsymbol{\theta}^{(0)}$, we simulate M transitions under this Markov chain and record the observed values $\boldsymbol{\theta}^{(j)}$, $j = 0, \ldots, M$. If $\mathbb{E}_\pi|\boldsymbol{\theta}| < \infty$, then the ergodic sample average $\hat{\boldsymbol{\theta}} = \frac{1}{M+1} \sum_{j=0}^{M} \boldsymbol{\theta}^{(j)}$ converges almost surely to $\mathbb{E}_\pi[\boldsymbol{\theta}]$ by the Markov chain strong law of large numbers (SLLN, see [6]). We implemented a standard Metropolis within Gibbs sampler (also known as component-wise Metropolis Hastings) which breaks down the 7-dimensional target $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:T}, \boldsymbol{u}_{1:T})$ into simpler, one dimensional, targets. For all MCMC computations, we used a truncated normal distribution $\mathcal{TN}(0, \sigma^2, 0, +\infty)$ proposal, with scaling parameter σ calibrated so as to obtain an optimal acceptance rate of $1/4$, as recommended by Roberts et al. (2004) in the case of high dimensional models. For technical details on the algorithm, see [39, 6]. The code for our MCMC posterior simulator is written in the MPI C programming language, and runs on the WestGrid Nestor Cluster.

4. Validation Using Synthetic Data. We ran validation tests for our Bayesian procedure (and MCMC posterior simulator) using synthetic data. These 'fake-data check' consist of 5 steps.

1. Fix the input cloud timescale parameters values at the values reported in Table 2.2.
2. Run the coupled toy GCM-SMCM model with those input parameters for a given size of the cloud lattice. The model outputs the cloud population (i.e. cloud cover) time series, and large-scale variable time series.
3. Isolate the stationary cloud population and large-scale variable time series to be used as "synthetic" observed time series for the Bayesian procedure.
4. Run the Bayesian procedure using the synthetic time series obtained in Step 3 and selected priors on the cloud timescale parameters. The Bayesian procedure outputs distributions on the parameters, from which point estimates are calculated.

5. Compare the inferred parameter values with the "true" input values selected in Step 1.

Arguably the choice of priors has a significant impact on the Bayesian inference. We work under the assumption that we have little to no prior information on the timescale parameters, which motivates our choice of a diffuse normal and uniform priors, respectively.

We ran the consistency check described above for synthetic time series of increasing lengths, in the goal of testing the sensitivity of our inference method to observing more data. The results of these tests are reported in Section 4.2 for the normal and uniform priors, while the synthetic data are first presented in Section 4.1.

A second consistency check consists of carrying the inference in a sequential manner, using small successive batches of data, as opposed to a large contiguous one. This technique, commonly called *sequential learning*, becomes quite handy in the cases when only multiple short samples of real data are available. We discuss our results on sequential learning in Section 4.3 in the case of a normal prior.

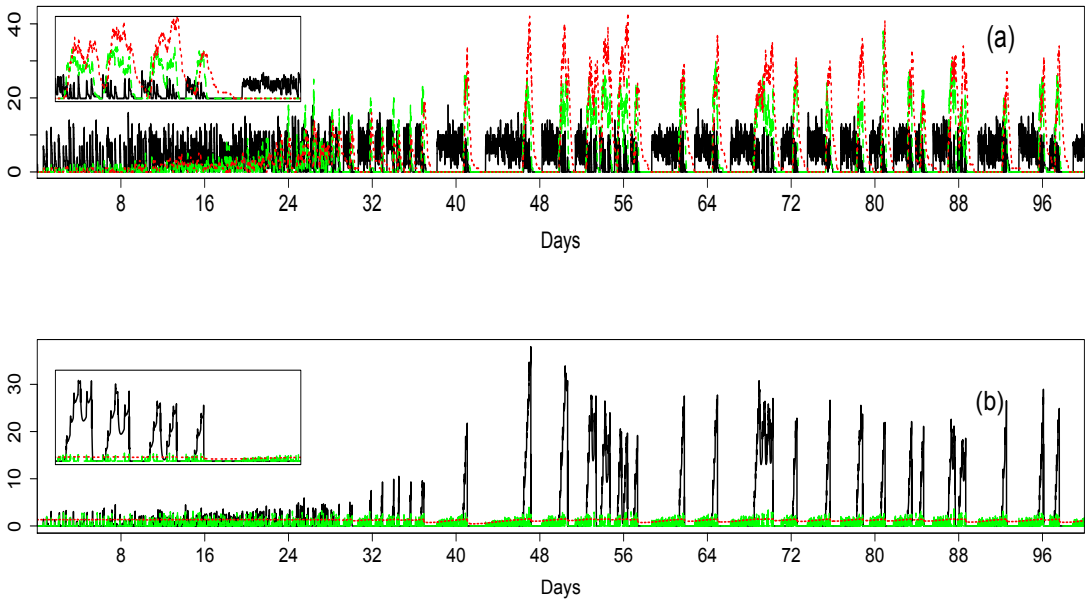


FIG. 4.1. *(a) (Main) Synthetic SMCM time series of the cloud populations of congestus N_c (black solid), deep N_d (green dashed) and stratiform N_s (red dotted) using the "true" parameter values of Table 2.2. (Inset) Interval from day 52 to day 61 in the equilibrium regime used for the validation. (b) Same as in (a) but for the large-scale variables C (black solid), C_l (green dashed), and D (red dotted).*

4.1. Synthetic Data. The synthetic data used for the validation were generated using the SMCM described in Section 2 coupled with a simple atmospheric climate model (toy GCM) [19], using the cloud timescale parameter values listed in Table 2.2, and a 10×10 cloud lattice ($N = 100$).

The simple one–column toy GCM, as described in details in [19], consists of a set of ordinary differential equation that are systematically coupled to the area fractions σ_c, σ_d, σ_s of congestus, deep, and stratiform clouds, predicted by the SMCM. More precisely, the ODE system describes the evolution of four key thermodynamic variables: the potential temperature components associated with the first and second baroclinic modes, θ_1 and θ_2, the midtroposphere specific humidity, q, and the boundary layer equivalent potential temperature, θ_{eb}. In turn, the large–scale factors that affect directly the dynamics of the SMCM, namely C, C_l, and D, are functions of θ_1, θ_2, q, and θ_{eb}. The interested reader is referred to [19] for more details on the toy GCM and its coupling to the SMCM.

The data consist of the cloud populations and large-scale variables time series shown in Figure 4.1, each totalling 28,800 sample points with a 5 minute sampling time interval, which corresponds to 100 days in real time. For the validation purposes, we use a shorter 9-day interval (between days 52 and 61, approximately 2500 sample points) taken from the radiative–convective equilibrium regime, as shown in the insets of Figure 4.1.

4.2. Sensitivity to the Length of the Time Series and Prior Specification. We first run the validation tests using a 7-dimensional truncated normal $\mathcal{TN}_7(10, 10\,Id_7, 0, +\infty)$ prior with mean 10 and covariance matrix $10 \cdot Id_7$, where Id_7 is the 7-dimensional identity matrix. The normal is restricted to the positive real half space to reflect the fact that the timescales are nonnegative values. The choice of this distribution is motivated by the need of having a weakly informative prior in the sense that (1) the prior mean of 10 is chosen sufficiently far from the true values in an attempt to demonstrate Bayesian learning towards the true values, but not too far to mitigate the limited amount of data information (increasing the amount of data T is cost prohibitive) and (2), the variance is large enough to have a significant amount of prior uncertainty, but not too large to reflect some degree of prior belief.

We used T = 100, 500, 1000, 1500, 2000, and 2500 contiguous observations of the synthetic time series, and ran our MCMC posterior simulator. In physical time, these numbers of observations approximately correspond to 8.3 hrs, 41.7 hrs, 3 days 11 hrs, 5 days 5 hrs, 6 days 23 hrs, and 8 days 16 hrs, respectively.

The Bayes estimates (mean, standard deviation, percentiles) and Monte Carlo standard error (MCSE) [33] for T = 100, 500, and 2500 are shown in Table 4.1 for all 7 parameter marginal posterior distributions. The MCSE and Bayes estimates are calculated using the `MCMCpack R` package [33], after the burn-in portions of the chains have been removed. Several convergence diagnostics were used to ensure that the chains have reached equilibrium. First we used an ensemble of well-dispersed chains in parallel and compared their performances, a paramount tool recommended by Robert and Casella (2010) when assessing convergence to stationarity. This was facilitated by parallel processing on the WestGrid Nestor cluster. We also monitored convergence to stationarity and convergence of averages using graphical and statistical tests provided by the `coda` package in `R`.

Also reported in Table 4.1 are the MCMC sample sizes, approximate runtimes, and number of processors used. As discussed in Section 3.1, the forward problem is computationally expensive, and increasing the number T of observations results in much longer computing time, even on a greater number of processors and for a smaller MCMC sample size.

We highlight a few results. First we note that we expect the prior to dominate the posterior when the number T of observations used is too small, and the likelihood to domi-

nate the posterior when a sufficiently large number of observations is used. When only 100 observations are used, the parameters τ_{01} and τ_{10} are recovered with good enough accuracy (0.934 and 0.894, respectively; their true value is 1.0 for both), while the estimated mean and standard deviation of the remaining parameters are essentially those of the prior. This seems to correlate well with data, as much of the activity happens mainly between congestus and clear skies during that lapse of time; the 100 observations used for that validation correspond to the first 1/25 of the time series shown in the inset of Figure 4.1, and covers only a small section of the first onsets of deeps and stratiforms. When 500 observations are used, all parameters except τ_{20} roughly recover their true value within one standard deviation not exceeding 21%. When 2500 observations are used, τ_{20}'s true value is almost recovered within a standard deviation of 19%, while the remaining parameters' relative error do not exceed 7%.

TABLE 4.1

Validation results for a normal prior, based on $T = 100$, 500, and 2500 observations. MCMC sample size (excluding burn in), approximate total runtime, and number of processors used in parentheses. Mean, SD, and MCSE are posterior mean, standard deviation, and Monte Carlo standard error, respectively. 2.5% and 97.5% are posterior percentiles.

	True Value	Mean	SD (MCSE)	2.5%	97.5%
$T = 100$ (100,000, 24 days, 16 cores)					
τ_{01}	1	0.934	0.1727 (0.003028)	0.6547	1.328
τ_{10}	1	0.894	0.1534 (0.001701)	0.6450	1.243
τ_{12}	.25	9.779	2.9458 (0.395798)	3.9921	15.590
τ_{02}	3	6.370	2.8332 (0.137016)	1.9763	12.572
τ_{23}	3	7.584	3.2871 (0.222591)	1.7847	14.201
τ_{20}	2	10.088	3.0369 (0.067716)	4.1766	16.048
τ_{30}	5	10.004	3.0264 (0.108871)	4.1161	15.884
$T = 500$ (25,293, 42 days, 16 cores)					
τ_{01}	1	1.0116	0.09669 (0.001547)	0.8422	1.2204
τ_{10}	1	0.9111	0.11283 (0.002064)	0.7178	1.1558
τ_{12}	.25	0.3206	0.06582 (0.001334)	0.2198	0.4776
τ_{02}	3	3.4709	0.39302 (0.007823)	2.8028	4.3413
τ_{23}	3	2.9576	0.27455 (0.005086)	2.4747	3.5561
τ_{20}	2	7.4629	2.91743 (0.121603)	2.8282	13.7671
τ_{30}	5	4.9184	0.46225 (0.007028)	4.0957	5.8955
$T = 2500$ (12,695, 42 days, 72 cores)					
τ_{01}	1	1.0019	0.04285 (0.0009227)	0.9198	1.0882
τ_{10}	1	0.9821	0.05310 (0.0009740)	0.8869	1.0968
τ_{12}	.25	0.2411	0.02347 (0.0004343)	0.2005	0.2915
τ_{02}	3	3.0110	0.18012 (0.0032739)	2.6712	3.3749
τ_{23}	3	2.9295	0.14277 (0.0026691)	2.6643	3.2199
τ_{20}	2	2.5256	0.48401 (0.0096552)	1.7842	3.6602
τ_{30}	5	5.3193	0.26249 (0.0052716)	4.8276	5.8481

In Figure 4.2, we compare the marginal posterior densities for all seven parameters using from $T = 500$ to 2500 observations, by increments of 500. It is interesting to see the

convergence to a limiting posterior distribution, as we increase the number of observations. More precisely, as T increases, the marginal posterior distributions concentrate close to the parameter true values, and we see a progressive reduction and then a sudden stagnation of the variance; for all parameters, the distribution curves for $T = 2000$ and 2500 match closely. This suggests that although all parameters were reproduced accurately, some uncertainty remains, independently of the length of the data time series.

To test the effect of this 'sample uncertainty' on the climate model dynamics, we now verify whether the inferred values of the cloud timescale parameters do reproduce the climate features we started with. For this, we run the SMCM coupled with a toy GCM as described in Section 4.1, with the inferred values reported in Table 4.1 for $T = 2500$. We then compare the resulting new time series, shown in Figure 4.3, with the original ones (see Figure 4.1). In Table 4.2, we report some statistics (mean and standard deviation) for the cloud populations and main climate thermodynamic variables of interest (time series not shown), both for the original and the new (inferred) time series. From these statistics, and by inspection of the two time series, we can ascertain that we recover the climate mean and variability with high fidelity. The fact that the climate dynamics are reproduced accurately, in spite of the uncertainty in some inferred parameters such as τ_{20}, suggests that the stochastic climate model itself is not very sensitive to those parameters. This is in fact, the more meaningful test that the Bayesian methodology has to pass and it passed it successfully.

TABLE 4.2

Mean and standard deviation (SD) for the time series of the cloud populations (shown in Figures 4.1 and 4.3) and climate thermodynamic variables (time series not shown), which were obtained (1) using the parameter values reported in Table 2.2 (original), and (2) using the Bayes posterior mean values reported in Table 4.1, for $T = 2500$ *(inferred). The climate thermodynamic variables shown are:* θ_1 *and* θ_2 *the potential temperatures associated with the first and second baroclinic modes,* θ_{eb} *the equilibrium temperature, and* q *the moisture.*

	(1) *Original* Mean (SD)	(2) *Inferred* Mean (SD)
Cloud Populations		
Congestus	3.583 (3.625)	3.267 (3.484)
Deep	3.088 (6.371)	3.456 (6.640)
Stratiform	5.214 (8.744)	6.334 (9.598)
Climate Thermodynamic Variables		
θ_1	9.759e-02 (8.538e-02)	9.368e-02 (7.910e-02)
θ_2	-5.641e-02 (6.306e-02)	-5.310e-02 (5.742e-02)
θ_{eb}	-5.604e-02 (1.256e-01)	-5.196e-02 (1.158e-01)
q	4.326e-02 (4.608e-02)	4.354e-02 (4.299e-02)

In order to investigate the sensitivity of the Bayesian model to prior belief specification, we selected a second prior, the 7-dimensional uniform distribution with large support $\mathcal{U}_7(0, 30)$. This choice of prior can represent an inference context in which there is no a priori information about the parameters. The validation results using the uniform prior were comparable to those obtained using the normal prior except, as might be expected, when only 100 data were used. In the case $T = 100$ and for a MCMC sample of size 92,000, we obtained posterior means (standard deviations) of 0.9026 (0.1641), 0.8654 (0.1418), 22.2945 (3.3827), 4.0160 (2.65), 3.7721 (3.3077), 16.327 (8.0798) and 15.381 (8.1068) for the param-

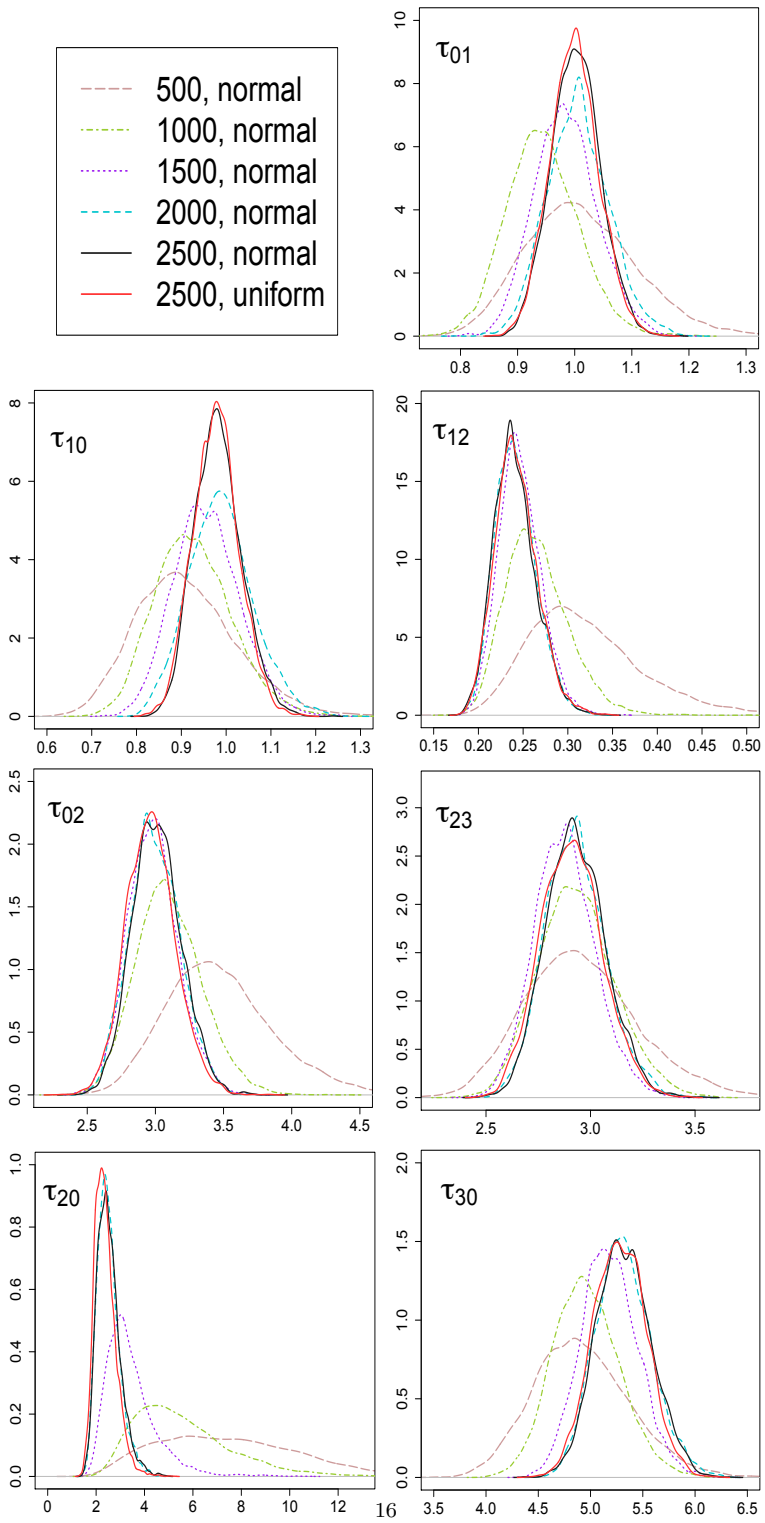
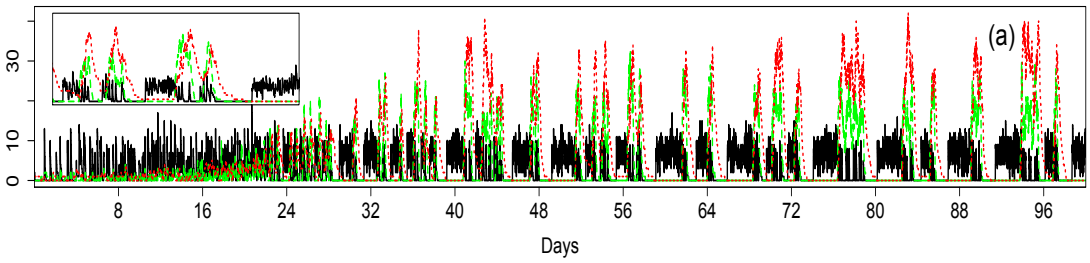
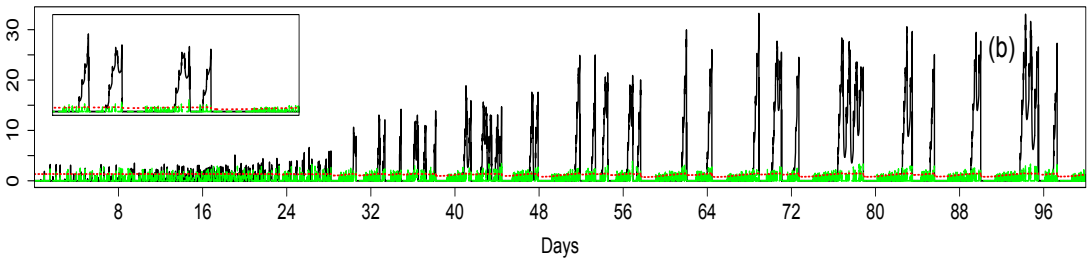


FIG. 4.2. *Comparison of marginal posterior densities using from* $T = 500$ *to* 2500 *observations, by increments of* 500*, for a normal prior. Also shown (in red) is the marginal posterior density for the uniform prior using* $T = 2500$.



(a) Cloud population time series.



(b) Covariates time series.

FIG. 4.3. *(a) (Main) SMCM time series of the cloud populations of congestus N_c (black solid), deep N_d (green dashed) and stratiform N_s (red dotted) using the inferred values of the cloud timescale parameters in Table 4.1 for $T = 2500$, for a normal prior. (Inset) Smaller interval covering day 52 through day 61. (b) but for the large-scale variables C (black solid), C_l (green dashed), and D (red dotted).*

eters in the same order as that given in Table 4.1. We note in passing that the true value of 1 of the parameters τ_{10} and τ_{01} is recovered within one standard deviation of roughly 17%, which is consistent with the values obtained under the normal prior, and supports the evidence that the first 100 cloud observations provide information mainly about the transitions between clear sky and congestus. The posterior means for τ_{20} and τ_{30} (16.327 and 15.381, respectively) are far from their true values (2 and 5, respectively), and closer to the midpoint (15) of the large support of the uniform prior. The posterior mean of 22.2945 for τ_{12} is somewhat off, but the behaviour of the MCMC simulator may be counterintuitive on a high dimensional space in the sense that one should not expect a posterior mean somewhat between the midpoint of the uniform prior and the true value for each parameter. Also, as noted above, the first 100 observations constitute the preconditioning phase dominated by congestus activity and only very few –under sampled– intermittent deep convection events are produced. The validation results for larger T values are not reported here because they are not statistically different from those obtained under the normal prior (see Table 4.1), but the posterior marginal distributions for $T = 2500$ are shown in Figure 4.2 (red solid lines) together with the ensemble of posterior marginals obtained under the normal prior.

These results indicate that the outcome of the inference is statistically the same irrespective of the choice of the prior, provided there is sufficient data. This constitutes a formal verification of the correctness of our MCMC sampler. In particular, it verifies that our posterior simulator is robust to the choice of prior, and that it responds normally to the input data, i.e. it outputs a posterior that is dominated by the prior when the data do not provide enough information about the parameters.

4.3. Sequential Bayesian Inference. We end this validation study by looking at how the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{x}_{1:T}, \boldsymbol{u}_{1:T})$ evolves as we sequentially update the posterior by new data points $\boldsymbol{x}_t$, $\boldsymbol{u}_t$. Such an incremental procedure is justified by the fact that the data are assumed to be Markovian. For instance, for $\boldsymbol{x}_{1:3} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$, and if we drop $\boldsymbol{u}_{1:3}$ for simplicity:

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) &\propto f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= f_2(\boldsymbol{x}_3|\boldsymbol{x}_2, \boldsymbol{\theta}) \times f_1(\boldsymbol{x}_2|\boldsymbol{x}_1, \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &= \text{Likelihood of } \{\boldsymbol{x}_3\} \times \text{posterior having observed } \{\boldsymbol{x}_2, \boldsymbol{x}_1\}\end{aligned}$$

More generally, the posterior having observed $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)$ can be used as a 'prior' for the remaining data $(\boldsymbol{x}_{K+1}, \ldots, \boldsymbol{x}_N)$. In the Bayesian framework, this is equivalent to observing the sequence of data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ all at once. This is particularly useful if the inference is done online while data is gradually available, or in cases where only short, discontinuous data series are available. This is the case for example when data is gathered from various sources, e.g. various observation sites. We perform the sequential Bayesian inference following the main steps below.

1. Run the Bayesian model with the first 100 observations (*Sequence 1*), using a $\mathcal{TN}_7(10, 10\,Id_7, 0, +\infty)$ prior. Obtain a posterior distribution over the model parameters. (This was done in Section 4.2. See validation results of Table 4.1 for $T = 100$.)
2. Use the method of moments approach to fit a multivariate normal distribution to the posterior samples.
3. Use the fitted distribution obtained in 2. as the prior for computing the posterior distribution from the next successive 100 observations (*Sequence 2*).
4. Iterate Steps 2 and 3 three more times until 5 sequences of 100 observations have been used for the sequential inference.
5. Compare the resulting posterior distribution with the posterior obtained by observing the 500 contiguous observations all at once.

The sequential learning strategy described here should not be confused with sequential Monte Carlo samplers, a collection of algorithms that build on importance sampling methods [40]. Here, the fitting of the Bayesian posterior sample to a parameterized distribution (Step 2) introduces some degree of uncertainty. In fact, the sequential inference strategy failed in the case of an initial $\mathcal{U}_7(0, 30)$ prior, due to the large misfit to a normal introduced after observing the first 100 data (see Section 4.2 for the validation results under the uniform prior in the case of $T = 100$).

The method was however successful in the case of an initial $\mathcal{TN}_7(10, 10\,Id_7, 0, +\infty)$ prior, as it can be seen from the sequences of updated marginalized posteriors shown in Figure 4.4. As more and more sequences of data are observed, the posteriors shift towards and concentrate on the parameter true values. In some cases, the initial high variance reduces

significantly after one or two observed sequences only. The statistics of the marginalized posteriors for Sequence 5 are summarized in Table 4.3.

As shown in Figure 4.4, all posterior marginals obtained after observing Sequence 5 (dashed green) match closely those obtained when all 500 data are observed at once (solid black), except in the case of the parameter τ_{12}. For that specific parameter, the posterior mean after having observed Sequence 5 (dashed green) is equal to 0.2688, which is closer to the true value of 0.25 than the posterior mean of 0.3206 obtained when all 500 data are observed at once (see Table 4.1). There is also less uncertainty about the posterior mode in the sequential case (posterior standard deviation of 0.03152 compared to 0.06582). It is interesting to see how the batch-wise technique captures the true value τ_{12} better than when the data are used all at once, which might suggest that Step 2 of the sequential learning strategy enhances the inference results in some way.

TABLE 4.3

Sequential Bayesian inference validation results for Sequence 5, for an initial $\mathcal{TN}_7(10, 10\,Id_7, 0, +\infty)$ prior. MCMC sample size (excluding burn in) is 150,424. Mean, SD, and MCSE are posterior mean, standard deviation, and Monte Carlo standard error, respectively. 2.5% and 97.5% are posterior percentiles.

	True Value	Mean	SD (MCSE)	2.5%	97.5%
τ_{01}	1	1.0155	0.09064 (0.0005093)	0.8388	1.194
τ_{10}	1	0.9094	0.09471 (0.0006049)	0.6286	1.183
τ_{12}	.25	0.2688	0.03152 (0.0002034)	0.2077	0.331
τ_{02}	3	3.5928	0.40934 (0.0035545)	2.7982	4.398
τ_{23}	3	2.9869	0.29188 (0.0019786)	2.4218	3.562
τ_{20}	2	7.4548	2.68044 (0.0389249)	2.6357	12.935
τ_{30}	5	4.7373	0.42297 (0.0026518)	3.9285	5.585

5. Discussion. A Bayesian method for learning some parameters for the stochastic multicloud model (SMCM) for organized tropical convection of Khouider et al. [19] is presented and validated here using synthetic data. The SMCM is in essence a three dimensional birth-death process with immigration whose population species track the time evolution of the area fractions of three cloud types, congestus, deep, and stratiform, that are observed to characterize tropical convective systems [16, 32]. The SMCM is based on gas model lattice overlaid over each GCM grid box. Each lattice site is either occupied by one of the three cloud types or is a clear sky site. Lattice sites switch between the four possible states according to intuitive probability rules motivated by observations. This results in probability transition rates which depend exclusively on the large-scale variables through some prescribed functions of exogenous factors represented by the potential for convection (CAPE) and middle tropospheric humidity, modulated by timescales. While the functionals are educated guesses that take the form of Arrhenius activation functions, the timescales are essentially free parameters whose values are very uncertain. In the past, intuitive values have been used satisfactorily in the case of idealized simulations of convectively coupled gravity waves [10, 11, 30] and rough estimates were obtained by a simple matching of the equilibrium distribution of the Markov process to observed mean area fractions [37]. Yet, the accurate estimation of the parameters from observation and/or detailed cloud resolving data remains an important step forward in order to effectively use the SMCM for the

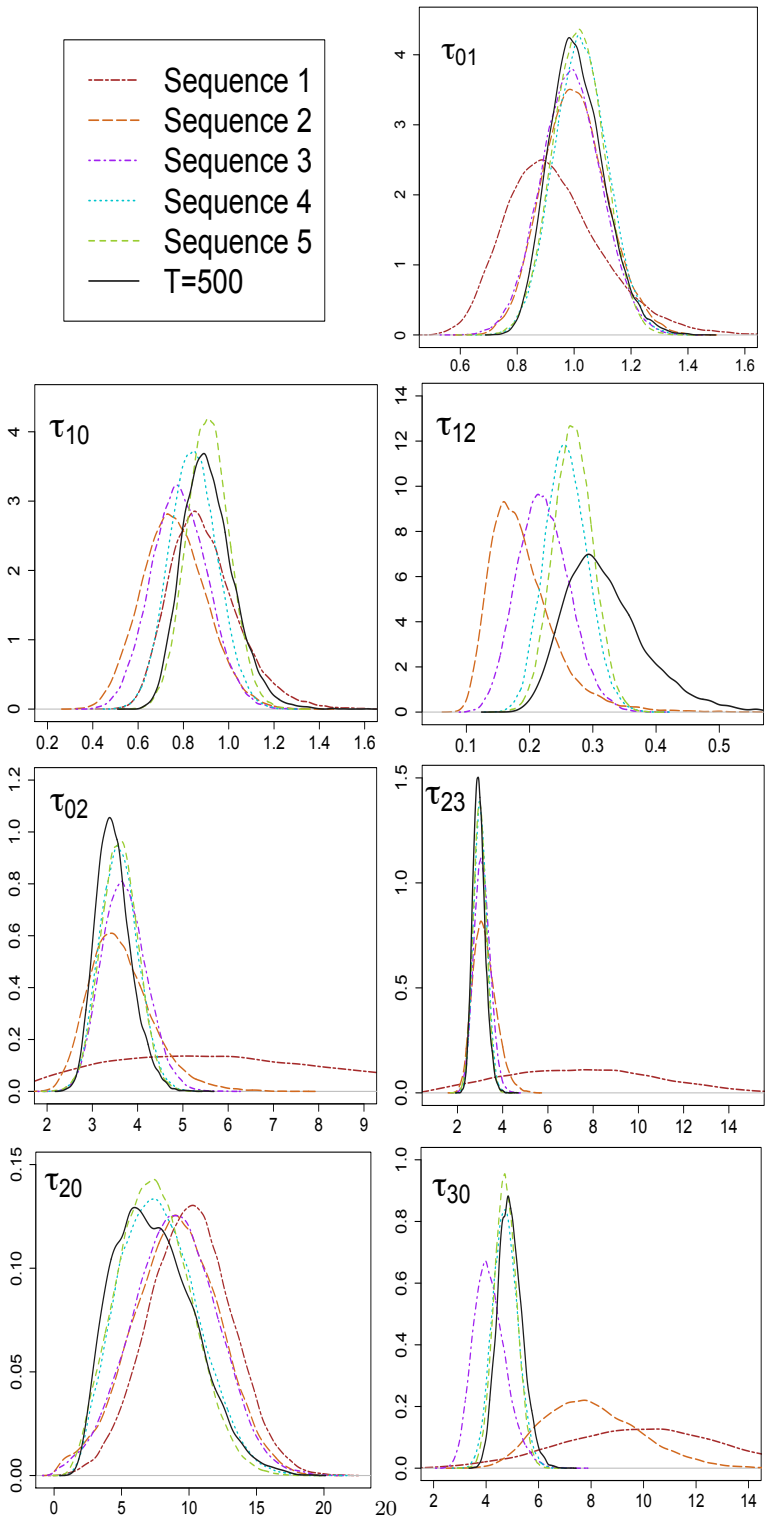


FIG. 4.4. *Marginal posterior densities obtained using 5 contiguous sequences of 100 observations each (Sequences 1-5), when the posterior of the previous sequence is used as the prior for the next, and for a initial* $\mathcal{TN}_7(10, 10\, Id_7, 0, +\infty)$ *prior. Also shown are the marginal posterior densities when 500 observations are used all at once* ($T = 500$). *The Sequence 1 distribution is missing for the parameter* τ_{12} *as it lies far from the true value.*

parameterization of organized convection in operational climate models.

The main challenge in using the Bayesian approach to infer parameters for the SMCM from data resides in the computation of the likelihood function which involves the computation of the transition probability matrix for each time step of the data time series, containing both the actual cloud area fractions and the associated exogenous factors. According to the Kolmogorov backward equations, the transition probability matrix is given by the exponential of the infinitesimal generator of the Markov chain, which is, in this case, of very high dimension but also very sparse. In this work, we took advantage of the Uniformization preconditioning methodology and a highly parallel software (`PETSc`) to approximate the exponential matrix with an acceptable accuracy and a meaningful efficiency. Moreover, we use the Markov Chain Monte Carlo technique to sample the high dimensional posterior distribution which adds another layer of computational complexity. However, overall, the Bayesian approach remains very competitive compared to the pure sampling of the conditional transition probabilities as done in [7] or to a method based on clustering analysis [13]. While such methods can be advantageous when there is a complete lack of physical intuition about the functional dependance of the transition probabilities on the exogenous factors for the former and about the actual cloud types and their structural properties or both for the latter, they both necessitate highly dense and very large time series. The lack of good quality observations and numerical simulations with such properties compounded with the associated sheer computational cost inhibit us from using effectively such methods.

Here the Bayesian approach is successfully tested with synthetic data generated by an idealized single column toy GCM coupled to the SMCM. At first the coupled toy GCM-SMCM model was run with prescribed parameters. The output time series of cloud area fractions and exogenous factors are then fed to the Bayesian algorithm to infer back some of the SMCM parameters, namely, the (seven) transition timescales. The Bayesian method is tested with two different choices of prior, a weakly informative normal prior centred far away from the true values and an uninformative uniform prior, and is shown to be robust to prior specification when enough observations are provided (about 500). With a moderately sized time series, around $T = 500$ observations (corresponding to 1.7 days in physical time), most of the timescales were reproduced with some accuracy–to within one standard deviation of about 21%, except for the transition of deep to clear sky parameter τ_{20} which remains highly inaccurate at this level. With the relatively higher number of observations $T = 2500$ the parameter τ_{20} is recovered within a standard deviation of 19%, while all other 6 parameters are closely recovered within one standard deviation of about 7% (see Table 4.1). Interestingly, the convergence tests reported in Figure 4.2 for the marginal distributions with $T = 500, 1000, 1500, 2000, 2500$ show a systematic convergence in the beginning and then a sudden stagnation towards a limiting posterior distribution which seems to suggest the existence of an upper bound or a maximum knowledge which can be gained from data in terms of the parameter values. As a consistency check we rerun the coupled toy GCM-SMCM model with the newly inferred parameters (i.e. the inferred means) and compared the statistics of the resulting climate variables to their original counterparts. The results reported in Table 4.2 and Figure 4.3 demonstrate that despite the systematic errors committed by the inferred parameters, the coupled model reproduces the original climate statistics quite accurately, in terms of both the stochastic area fractions and the large-scale dynamical variables. This in essence indicates the level to which the coupled toy GCM-SMCM model is actually sensitive to these parameters. It is clearly less sensitive

to some parameters such as τ_{20} than it is to others. This is in fact very good news, and explains in some sense why the SMCM is so successful in previous studies, when only rough or intuitive estimates of these parameters were used [10, 11, 37, 30].

Moreover, the Bayesian methodology is tested for sequential learning, which consists of cutting the available time series into a sequence of small non-overlapping segments of 100 observations. The Bayesian algorithm is then sequentially applied on each segment and at each step the prior is taken to be a Gaussian distribution with mean and variance given by those learned from the previous step. It is found that the sequential learning performed reasonably well, compared to using all the available data at once (here 500 observations), when the initial step prior is also a Gaussian but not when the initial prior is uniform. Such sequential learning will be useful in more practical applications if for instance discontinuous time series are used, such as in the case of multiple observation sites or the case of on and off observation periods, making the intervals between some consecutive observations too large to satisfy the stationarity assumption.

Although, the Bayesian algorithm is developed here for the transition timescales, it remains very flexible. It can be easily extended to estimate other parameters of the SMCM such as the adimensionalization prefactors $CAPE_0$ and T_0 used in the definition of the exogenous factors C, C_l and D, respectively, to rescale the actual measurement of CAPE, low-level CAPE, and mid-tropospheric dryness (see [19, 10, 11, 37]). More importantly it can be extended in a straightforward fashion to learn from data the actual interaction potential for the SMCM with local interactions presented recently in [18]. Such local interactions are important for the self-organization of convection due to local processes such as gravity currents, cold pools, sea breezes, and the diurnal cycle. The main causes of the initiation of the Madden-Julian oscillation (MJO), over the Indian, remain a matter of a heated debate in the tropical meteorology community. While the initiation of successive MJOs are more or less elucidated as being due to dry Kelvin waves that are excited by previous MJO events [1], the case of primary MJOs remain largely an unsolved problem. The Bayesian method for the SMCM presented here can help illustrate the degree of self-organization of convection during the initiation of primary MJOs using data from the Dynamics of the MJO field campaign in order to understand the dilemma of the deepening of convection due to congestus moistening [45, 12, 15, 26].

Acknowledgement. This research has been enabled by the use of computing resources provided by WestGrid and Compute/Calcul Canada. The authors would like to thank Belaid Moa from Compute Canada for his HPC assistance and help, and Farouk Nathoo for useful discussions on the topic of statistical inference. The research of BK is partially supported by the Natural Sciences and Engineering Council of Canada.

REFERENCES

[1] R. S Ajayamohan, B. Khouider, and A. J. Majda, *Realistic Madden-Julian Oscillation initiation and dynamics in a general circulation model*, Geophy. Res. Lett., (2013), p. Submitted.

[2] A. Arakawa, *The cumulus parameterization problem: Past, present, and future*, Journal of Climate, 17 (2004), pp. 2493–2525.

[3] A. Arakawa and W. H. Schubert, *Interaction of a cumulus cloud ensemble with the large-scale environment, Part I*, Journal of the Atmospheric Sciences, 31 (1974), pp. 674–701.

[4] Satish B., J. Brown, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, *PETSc Web page*, 2013. http://www.mcs.anl.gov/petsc.

[5] M. A. Botchev, *A short guide to exponential Krylov subspace time integration for Maxwell's equations*, Internal Report, Department of Applied Mathematics, University of Twente, (2012).
[6] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, Taylor & Francis US, 2011.
[7] J. Dorrestijn, D.T. Crommelin, J.A. Biello, and S.J. Boing, *A data-driven multicloud model for stochastic parameterization of deep convection*, Phil. Trans. R. Soc. A, (2013), p. In press.
[8] F. Dulat, J.-P. Katoen, and V. Y. Nguyen, *Model checking Markov chains using Krylov subspace methods: an experience report*, in Computer Performance Engineering, Springer, 2010, pp. 115–130.
[9] W. Feller, *On the theory of stochastic processes, with particular reference to applications*, in Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability. August 13-18, 1945 and January 27-29, 1946. Statistical Laboratory of the University of California, Berkeley. Berkeley, Calif.: University of California Press, 1949. 501 pp. Editor: Jerzy Neyman, p. 403-432, vol. 1, 1949, pp. 403–432.
[10] Y. Frenkel, A. J. Majda, and B. Khouider, *Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example*, Journal of the Atmospheric Sciences, 69 (2012), pp. 1080–1105.
[11] ———, *Stochastic and deterministic multicloud parameterizations for tropical convection*, Climate Dynamics, (2013), pp. 1–25.
[12] C. Hohenegger and B. Stevens, *Preconditioning deep convection with cumulus congestus*, J. Atmos. Sci., 70 (2013), pp. 448–464.
[13] I Horenko, *Nonstationarity in multifactor models of discrete jump processes, memory and application to cloud modeling*, J. Atmos. Sci., 68 (2011), pp. 1493–1506.
[14] A. Jensen, *Markoff chains as an aid in the study of Markoff processes*, Scandinavian Actuarial Journal, 1953 (1953), pp. 87–91.
[15] R. Johnson and P. E. Ciesielski, *Structure and properties of Madden-Julian oscillations deduced from DYNAMO sounding arrays*, Journal of the Atmospheric Sciences, (2013), p. In press.
[16] R. H. Johnson, T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert, *Trimodal characteristics of tropical convection*, Journal of climate, 12 (1999), pp. 2397–2418.
[17] M. F. Khairoutdinov, S. K. Krueger, C.-H. Moeng, P. A. Bogenschutz, and D. A. Randall, *Large-eddy simulation of maritime deep tropical convection*, Journal of Advances in Modeling Earth Systems, 1 (2009).
[18] B. Khouider, *A coarse-grained stochastic particle interacting system for tropical convection*, Communications in Mathematical Sciences, In press (2013).
[19] B. Khouider, J. Biello, and A. J. Majda, *A stochastic multicloud model for tropical convection*, Communications in Mathematical Sciences, 8 (2010), pp. 187–216.
[20] B. Khouider and A. J. Majda, *Multicloud convective parametrizations with crude vertical structure*, Theoretical and Computational Fluid Dynamics, 20 (2006), pp. 351–375.
[21] ———, *A simple multicloud parameterization for convectively coupled tropical waves. Part I: Linear analysis*, Journal of the atmospheric sciences, 63 (2006), pp. 1308–1323.
[22] ———, *Equatorial convectively coupled waves in a simple multicloud model*, Journal of the Atmospheric Sciences, 65 (2008), pp. 3376–3397.
[23] ———, *Multicloud models for organized tropical convection: Enhanced congestus heating*, Journal of the Atmospheric Sciences, 65 (2008), pp. 895–914.
[24] B. Khouider, A. J. Majda, and S. N. Stechmann, *Climate science in the tropics: waves, vortices and PDEs*, Nonlinearity, 26 (2013), p. R1.
[25] B. Khouider, A. St-Cyr, A. J. Majda, and J. Tribbia, *The MJO and convectively coupled waves in a coarse-resolution GCM with a simple multicloud parameterization*, Journal of the Atmospheric Sciences, 68 (2011), pp. 240–264.
[26] V. V. Kumar, C. Jakob, A. Protat, P. T. May, and L. Davies, *The four cumulus cloud modes and their progression during rainfall events: A C-band polarimetric radar perspective*, Journal of Geophysical Research: Atmospheres, (2013), p. In press.
[27] H.-L. Kuo, *Further studies of the parameterization of the influence of cumulus convection on large-scale flow*, Journal of the Atmospheric Sciences, 31 (1974), pp. 1232–1240.
[28] J. W.-B. Lin and J. D. Neelin, *Toward stochastic deep convective parameterization in general circulation models*, Geophysical research letters, 30 (2003).
[29] A. J. Majda and B. Khouider, *Stochastic and mesoscopic models for tropical convection*, Proceedings of the National Academy of Sciences, 99 (2002), pp. 1123–1128.

[30] B. Majda, A. J. Khouider and Y. Frenkel, *Effects of rotation and mid-troposphere moisture on organized convection and convectively coupled waves*, Climate Dynamics, (2013), p. Submitted.
[31] S. Manabe and J. Smagorinsky, *Simulated climatology of a general circulation model with a hydrologic cycle*, Monthly Weather Review, 95 (1967).
[32] B. Mapes, S. Tulich, J. Lin, and P. Zuidema, *The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves?*, Dynamics of atmospheres and oceans, 42 (2006), pp. 3–29.
[33] A. D. Martin, K. M. Quinn, and J. H. Park, *Mcmcpack: Markov Chain Monte Carlo in R*, Journal of Statistical Software, 42 (2011), pp. 1–21.
[34] C.J. Mode, *Some multi-dimensional birth and death processes and their applications in population genetics*, Biometrics, (1962), pp. 543–567.
[35] C. Moler and C. Van Loan, *Nineteen dubious ways to compute the exponential of a matrix*, SIAM review, 20 (1978), pp. 801–836.
[36] T. N. Palmer, *A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models*, Quarterly Journal of the Royal Meteorological Society, 127 (2001), pp. 279–304.
[37] K. Peters, C. Jakob, L. Davies, B. Khouider, and A. Majda, *Stochastic behaviour of tropical convection in 1 observations and a multicloud model*, J. Atmos. Sci., (2013), p. In press.
[38] C. Robert, *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Springer, 2007.
[39] C. Robert and G. Casella, *Introducing Monte Carlo Methods with R*, Springer, 2010.
[40] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, vol. 319, Citeseer, 2004.
[41] Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 209–228.
[42] J. F. Scinocca and N. A. McFarlane, *The variability of modeled tropical precipitation*, Journal of the atmospheric sciences, 61 (2004), pp. 1993–2015.
[43] R. B. Sidje, *Expokit: a software package for computing matrix exponentials*, ACM Transactions on Mathematical Software (TOMS), 24 (1998), pp. 130–156.
[44] R. B. Sidje and W. J. Stewart, *A numerical study of large sparse matrix exponentials arising in Markov chains*, Computational Statistics & Data Analysis, 29 (1999), pp. 345–368.
[45] M. Waite and B. Khouider, *The deepening of tropical convection by congestus preconditioning*, J. Atmos. Sci., 67 (2010), p. 2601?2615.
[46] G. J. Zhang and N. A. McFarlane, *Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model*, Atmosphere-Ocean, 33 (1995), pp. 407–446.