SIAM/ASA J. Uncertainty Quantification
Vol. 2, pp. 647–669

An MCMC Algorithm for Parameter Estimation in Signals with Hidden Intermittent Instability*

Nan Chen†, Dimitrios Giannakis†, Radu Herbei‡, and Andrew J. Majda†

Abstract. Prediction of extreme events is a highly important and challenging problem in science, engineering, finance, and many other areas. The observed extreme events in these areas are often associated with complex nonlinear dynamics with intermittent instability. However, due to lack of resolution or incomplete knowledge of the dynamics of nature, these instabilities are typically hidden. To describe nature with hidden instability, a stochastic parameterized model is used as the low-order reduced model. Bayesian inference incorporating data augmentation, regarding the missing path of the hidden processes as the augmented variables, is adopted in a Markov chain Monte Carlo (MCMC) algorithm to estimate the parameters in this reduced model from the partially observed signal. However, direct application of this algorithm leads to an extremely low acceptance rate of the missing path. To overcome this shortcoming, an efficient MCMC algorithm which includes a pre-estimation of hidden processes is developed. This algorithm greatly increases the acceptance rate and provides the low-order reduced model with a high skill in capturing the extreme events due to intermittency.

Key words. hidden process, intermittency, stochastic parameterized model, data augmentation, MCMC algorithm, prediction skill

AMS subject classifications. 60G25, 60H30, 62F15, 62P12, 94A15

DOI. 10.1137/130944977

1. Introduction.

Prediction of extreme events is an important and challenging problem. Famously, in 1998, the Long Term Capital Management (LTCM) hedge fund was driven into the ground as a result of the ripple effect caused by the Russian government's debt default [38], but none of LTCM's forecast models were able to predict this event and its subsequent effects. Another example is a heat wave [31], which is a prolonged period of excessively hot weather. Albeit rare, severe heat waves are able to cause catastrophic crop damage and significant loses. However, due to the complexity and uncertainty of the climate system, precise prediction of heat waves remains unavailable. Furthermore, the El Niño Southern Oscillation (ENSO), which is a significant climate pattern, triggers extreme weather such as floods and droughts and seriously affects the agriculture and fishing industry of countries around the Pacific Ocean [7]. Here, imperfect knowledge of the mechanism of ENSO impedes the prediction of the occurrence and duration of these extremes accurately.

*Received by the editors November 12, 2013; accepted for publication (in revised form) September 2, 2014; published electronically November 11, 2014.
http://www.siam.org/journals/juq/2/94497.html

†Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (chennan@cims.nyu.edu, dimitris@cims.nyu.edu, jonjon@cims.nyu.edu). The research of the first, second, and fourth authors was partially supported by ONR-MURI Grant N00014-12-1-0912.

‡Department of Statistics, Ohio State University, Columbus, OH 43210 (herbei@stat.osu.edu). The research of the author was partially funded by the National Science Foundation award DMS-1209142.

Mathematically, extreme events observed in signals are often associated with intermittency due to complex nonlinear dynamics with instabilities. However, as mentioned in the heat wave and ENSO examples, the mechanisms that drive extreme events are typically complicated and unavailable from observations. That is, the actual dynamics consist of many hidden and unresolved processes. Due to such uncertainties in the dynamics of nature ("the perfect model"), it is natural to construct low-order models with adequate prediction skill, which are nevertheless able to reflect the salient features of observed extreme events, i.e., intermittency associated with hidden processes.

In the framework of these models, the observed signal corresponds to incomplete observation of the full state vector. Therefore, it is important to develop systematic methods of parameter estimation based on incomplete observations. The ultimate goal is to utilize these low-order models equipped with the estimated parameters for prediction. These are the topics of the present paper.

Let $\{u(t),\, t \geq 0\}$ be a process of interest. We characterize the intermittency and hidden processes in $\{u(t)\}$ using the stochastic parameterization extended Kalman filter (SPEKF) model [28], given by

$$du(t) = \big((-\gamma(t) + i\omega)u(t) + f(t) + b(t)\big)\, dt + \sigma_u\, dW_u(t), \tag{1.1a}$$

$$d\gamma(t) = -d_\gamma\big(\gamma(t) - \hat{\gamma}\big)\, dt + \sigma_\gamma\, dW_\gamma(t), \tag{1.1b}$$

$$db(t) = \big(-d_b + i\omega_b\big)\big(b(t) - \hat{b}\big)\, dt + \sigma_b\, dW_b(t). \tag{1.1c}$$

In SPEKF models, the process $u(t)$ described in (1.1a) is driven by the stochastic damping $\gamma(t)$ and stochastic forcing correction $b(t)$, both of which are specified as Ornstein-Uhlenbeck (OU) processes as in (1.1b) and (1.1c). Here, $W_u(t), W_\gamma(t)$, and $W_b(t)$ are independent Brownian motions and all the parameters $\omega, \omega_b, d_\gamma, d_b, \hat{\gamma}, \hat{b}, \sigma_u, \sigma_\gamma$, and σ_b are scalars. Physically, the variable $u(t)$ in (1.1a) represents one of the resolved modes (i.e., observable) in the turbulent signal, while $\gamma(t)$ and $b(t)$ are hidden processes. In particular, $\gamma(t)$ and $b(t)$ are surrogates for the nonlinear interaction between $u(t)$ and other unobserved modes in the perfect model. This nonlinear system was first introduced in [14, 15] for filtering multiscale turbulent signals with hidden instabilities and has been used for filtering and prediction in the presence of model error [3, 4, 5, 6, 16, 23, 30]. The intermittency of the observed variable $u(t)$ is mainly a consequence of the sign switching of the hidden variable $\gamma(t)$, alternating between positive and negative phases, during which the process $\{u(t)\}$ switches between stable and unstable regimes. The strength of this intermittency also depends on the forcing correction $b(t)$.

In this paper, as a process model, we consider a simplified version of the unforced SPEKF model,

$$du(t) = -\gamma(t)u(t)\, dt + \sigma_u\, dW_u(t), \tag{1.2a}$$

$$d\gamma(t) = -d_\gamma\big(\gamma(t) - \hat{\gamma}\big)\, dt + \sigma_\gamma\, dW_\gamma(t), \tag{1.2b}$$

which is complex enough to incorporate intermittency and includes the hidden process $\gamma(t)$. Our primary interest is *the prediction skill* of the simplified SPEKF model (1.2), and in particular the ability to generate intermittency. This problem is tackled by first estimating the parameters $\boldsymbol{\theta} = (d_\gamma, \hat{\gamma}, \sigma_\gamma, \sigma_u)$ based on discrete-time observations of the $\{u(t)\}$ process. Consequently, the prediction step is performed in a probabilistic framework.

We use a Bayesian approach, under which the estimation step is based on exploring the posterior distribution of $\boldsymbol{\theta}$ conditionally on observations of the $\{u(t)\}$ process, via a Markov chain Monte Carlo (MCMC) approach. As we show in the next sections, the likelihood function is intractable in this case, and thus standard techniques are unavailable. We suggest an innovative approach based on data augmentation [39] and treat the unobserved path of the $\{\gamma(t)\}$ processed as missing data.

Roberts and Stramer [34] first applied this idea successfully for parameter estimation in a nonlinear diffusion model, handling certain technical issues using reparameterization [24] to circumvent the singularity of the dominating measures for diffusion with different diffusion coefficients. Other related work and improved algorithms for diffusion models are found in [1, 8, 9, 11, 13, 25, 32, 35, 36, 37]. However, extra effort is required to handle intermittency and unresolved processes. Various updating strategies have been proposed in [12, 18, 19, 20, 21, 22] in the context of well-behaved multivariate diffusion models.

As mentioned above, given that $u(t)$ may switch between different instability regimes, we expect that an off-the-shelf MCMC approach for estimating $\boldsymbol{\theta}$ may fail. To that end, we suggest an innovative sampling strategy involving a preconditioning step which restricts the proposed paths of the missing process $\gamma(t)$ to pass near a discrete set of pre-estimated values at the observation times. In this manner, the new sampler is able to explore efficiently and with sufficient accuracy the joint posterior distribution for the parameters and missing path, despite the fact that $\gamma(t)$ is infinite-dimensional. Throughout, we exploit the fact that (1.2a) defines a conditional Gaussian process given the path of $\{\gamma(t)\}$. We demonstrate the efficiency of the new sampler and predictive skill of the resulting simplified SPEKF models in perfect-model experiments and experiments with model error.

The rest of the paper is organized as follows. In section 2, we describe the Bayesian inference approach via data augmentation. For the inference procedure, we are implementing a standard MCMC algorithm as well as the approach of [22] and discuss their drawbacks in the context of the model described in (1.2). Our proposed approach is presented in section 3. Section 4 includes three numerical tests for parameter estimation and prediction skill with the new algorithm, the first of which is in a perfect model setting, while the other two deal with model error. Concluding remarks are given in section 5.

2. Preliminaries.

We consider the simplified SPEKF model (1.2) and assume that we observe the process $\{u(t)\}$ at a collection of discrete time points $0 = t_0 < t_1 < \cdots < t_n = T$. Let $\boldsymbol{U} = (U_0, U_1, \ldots, U_n)$ with $U_i = u(t_i)$. Our goal is to explore the posterior probability distribution,

$$p(\boldsymbol{\theta} \,|\, \boldsymbol{U}) \propto p(\boldsymbol{\theta}) p(\boldsymbol{U} \,|\, \boldsymbol{\theta}), \tag{2.1}$$

where $p(\boldsymbol{\theta})$ is the prior distribution on $\boldsymbol{\theta}$ and $p(\boldsymbol{U} \,|\, \boldsymbol{\theta})$ is the likelihood function. We note that, abusing notation, we use the generic notation $p(\cdot)$ to denote the probability density/conditional probability density function for the relevant quantities. Since the processes $u(\cdot)$ and $\gamma(\cdot)$ are coupled in a nonlinear way and $\gamma(\cdot)$ itself is a stochastic process, the likelihood function $p(\boldsymbol{U} \,|\, \boldsymbol{\theta})$ is not available in closed form. Therefore, we adopt a data augmentation approach [34]. Let $\gamma^{mis} = \{\gamma^{mis}(t),\ 0 \le t \le T\}$ represent the unobserved full path of $\gamma(\cdot)$, and consider the augmented state space $(\boldsymbol{\theta}, \gamma^{mis})$. The distribution (2.1) is replaced by the augmented

posterior distribution

$$p(\boldsymbol{\theta}, \gamma^{mis} \mid \boldsymbol{U}) \propto p(\boldsymbol{\theta}, \gamma^{mis}, \boldsymbol{U}) \tag{2.2a}$$

$$= p(\boldsymbol{\theta}) p(\gamma^{mis}, \boldsymbol{U} \mid \boldsymbol{\theta})$$

$$= p(\boldsymbol{\theta}) p(\gamma^{mis} \mid \boldsymbol{\theta}) p(\boldsymbol{U} \mid \gamma^{mis}, \boldsymbol{\theta}), \tag{2.2b}$$

where all the corresponding densities are viewed with respect to appropriate dominating measures; see [34]. If one is able to explore the probability model specified by (2.2b), the desired distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{U})$ can be determined by marginalizing over the auxiliary variable γ^{mis}. In light of the Markov property, the likelihood function can be decomposed recursively as follows:

$$p(\boldsymbol{U} \,|\, \gamma^{mis}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(U_i \,|\, U_{i-1}, \gamma^{mis}_{[t_{i-1}, t_i]}, \boldsymbol{\theta}),$$

where $\gamma^{mis}_{[t_{i-1}, t_i]}$ represents the path of γ^{mis} in the interval $[t_{i-1}, t_i]$. Conditionally on γ^{mis} and $U_0 = u(t_0)$, the dynamics of $\{u(t)\}$ are well understood and the pathwise solution of (1.2a) is given by

$$u(t) = \rho(t, t_0) u(t_0) + \rho(t, t_0) \int_{t_0}^{t} \sigma_u \rho^{-1}(s, t_0) \, dW_u(s), \tag{2.3}$$

where

$$\rho(t, t_0) = \exp\left(- \int_{t_0}^{t} \gamma^{mis}(s) \, ds \right). \tag{2.4}$$

Note that conditionally on $u(t_0)$ and γ^{mis}, the variate $u(t)$ has a Gaussian distribution with the mean and variance given, respectively, by

$$\begin{aligned} \mu(t; t_0) &= \rho(t, t_0) u(t_0), \\ \Sigma(t; t_0) &= \rho(t, t_0)^2 \int_{t_0}^{t} \sigma_u^2 \rho^{-2}(s, t_0) \, ds. \end{aligned} \tag{2.5}$$

Therefore, each term in the product in (2.3) becomes

$$p(U_i | U_{i-1}, \gamma^{mis}_{[t_{i-1}, t_i]}, \boldsymbol{\theta}) = \phi\big(U_i; \mu(t_i; t_{i-1}), \Sigma(t_i; t_{i-1})\big), \tag{2.6}$$

where $\phi(x; m, v)$ is the Gaussian PDF with mean m and variance v evaluated at x. Therefore, the conditional probability $p(\boldsymbol{U} \mid \gamma^{mis}, \boldsymbol{\theta})$ in (2.3) is obtained by capturing all the $p(U_i \mid U_{i-1}, \gamma^{mis}_{[t_{i-1}, t_i]}, \boldsymbol{\theta})$ with i from 1 to n.

We now focus on $p(\gamma^{mis} \,|\, \boldsymbol{\theta})$, which is viewed as the Radon–Nikodym derivative of the measure induced by (1.2b) with respect to the Wiener measure scaled by the diffusion coefficient σ_γ. Given that two such dominating measures (with different σ_γ) are singular, a direct MCMC implementation will result in a reducible algorithm; see [34]. Therefore, we introduce

a change of variable to normalize the diffusion coefficient σ_γ in the governing equation (1.2b). Setting $\alpha(t) = \gamma(t)/\sigma_\gamma$, the governing equation (1.2b) can be rewritten as

$$d\alpha(t) = -\frac{d_\gamma}{\sigma_\gamma}\left(\sigma_\gamma \alpha(t) - \hat{\gamma}\right) dt + dW_\gamma(t). \tag{2.7}$$

Note that the $\{\alpha(t)\}$ process will have constant quadratic variation on the interval $[0, T]$, and thus we now avoid the singularity of the corresponding dominating measures. We also note that this reparameterization will have to be reflected in (2.6) as well.

Let $a(\alpha, t, \boldsymbol{\theta}) = -(d_\gamma/\sigma_\gamma)\left(\sigma_\gamma \alpha(t) - \hat{\gamma}\right)$. Using the Girsanov formula [33],

$$p(\alpha^{mis} \,|\, \boldsymbol{\theta}) \propto L(\alpha^{mis}; \boldsymbol{\theta}), \tag{2.8}$$

where

$$L(\alpha^{mis}; \boldsymbol{\theta}) = \exp\left(\int_0^T a(\alpha, t, \boldsymbol{\theta})\, d\alpha_t\right) \exp\left(-\frac{1}{2}\int_0^T a(\alpha, t, \boldsymbol{\theta})^2\, dt\right). \tag{2.9}$$

With these in mind, a standard Metropolis–Hastings algorithm for exploring the posterior distribution (2.2a) will alternate between updating $\boldsymbol{\theta}$ conditionally on γ^{mis} and $\boldsymbol{U}$ and updating γ^{mis} conditionally on $\boldsymbol{\theta}$ and $\boldsymbol{U}$. Since neither of these conditional distributions are available in closed form, in our implementation we use Metropolis–Hastings updates. Since $\boldsymbol{\theta}$ is low-dimensional, designing a proposal distribution which will perform well is achievable either using a univariate or multivariate update strategy. We find that a deterministic scan univariate updating strategy performs on par with a more sophisticated adaptive approach. However, designing a good proposal distribution for the $\alpha^{mis}(\cdot)$ component turns out to be a very challenging task. Since the prior distribution for $\alpha^{mis}(\cdot)$ is an OU process, we experimented simulating OU paths with very limited success. It is well known that independent proposal distributions for Metropolis–Hastings algorithms are severely inefficient [25, 34]. Our best implementation resulted in acceptance rates below 0.1%, indicating that the sampler fails to explore the posterior distribution properly.

We note that in the paragraphs above, we describe an approach which aims to explore the infinite-dimensional space of the sample paths of the process $\alpha^{mis}(\cdot)$. While the Metropolis–Hastings formulation is straightforward, designing efficient algorithms is far from it. The main setback is that it is extremely difficult to design efficient proposal distributions over infinite-dimensional spaces. Evidently, any computer implementation will require some kind of finite-dimensional representation for $\alpha(\cdot)$, $u(\cdot)$ as well as a discrete time approximation for all the intractable integrals present above. A natural idea is to design and simulate an algorithm which will explore the finite-dimensional corresponding to some discrete-time approximation for the process described in (1.2). However, it is well documented [21, 34] that such approximating algorithms become increasingly inefficient as the discretization gets finer.

In the body of work of Golightly and Wilkinson, notably [22] and the references therein, one finds an extremely general MCMC approach for exploring the posterior distribution, such as (2.2a). Their approach is based on a very fine time discretization strategy and a data augmentation approach for the missing paths of both the observed and unobserved variables. This results in an algorithm exploring an extremely high-dimensional state space, where it is

imperative to perform multivariate updates in order to speed up convergence. Their strategy is to update overlapping missing paths of duration $2\Delta t$. Similarly, Chib, Pitt, and Shephard [10] suggest an alternative approach for Bayesian inference for discretely observed multivariate diffusions. Their approach is to update the unobserved paths using a proposal distribution which is constructed to be close in some sense to the target. Such algorithms have great generality and have proved useful in a variety of situations, as their authors demonstrate. In this paper we put forward a simpler strategy which in our case performs very efficiently. Rather than using a general MCMC strategy, we develop an efficient algorithm specifically for the SPEKF model. As we explain below, our algorithm leverages the conditional Gaussian structure and updates the missing paths of only the unobserved variable, thus reducing the computational cost. Moreover, it operates in simultaneous blocks of length $2\Delta t$, potentially improving the mixing properties. We have experimented with a basic implementation of the Golightly and Wilkinson approach, which produced reasonable estimates for the simplified SPEKF model parameters in (1.2), but at a significantly higher computational cost owing to the sampling of the $u(\cdot)$ paths. This issue would be further compounded in the applications of section 4 involving a stochastic phase that requires an additional sampling of the imaginary part of the $u(\cdot)$ signal.

We introduce the idea of *preconditioning* by informing the paths of the $\alpha^{mis}(\cdot)$. This can be done either by incorporating additional prior information or through the available observations. To that end we take full advantage of several important characteristics of the SPEKF model (1.2). In the absence of any additional prior information about the unresolved process $\alpha^{mis}(\cdot)$, we suggest a *mean stochastic model* (MSM) and use it to inform the unobserved paths. Critical to our approach is the observation that (1.2) defines a conditional Gaussian process, thus avoiding the need to carry out an ellaborate time discretization scheme for the $\{u(t)\}$ process. The necessary details are given in the next section.

3. Approximate MCMC via preconditioning. To address the deficiencies of a standard MCMC approach as described above, we put forward a new algorithm which preconditions the proposed missing path γ^{mis} by imposing soft constraints on $\gamma(t)$ at the observation times t_i. This is in contrast to the typical approach, see [22] and the references therein, where one would construct clever proposing mechanisms for the unobserved paths. In essence, the preconditioning procedure replaces the samples drawn from the posterior distribution $p(\boldsymbol{\theta}, \gamma^{mis} \mid \boldsymbol{U})$ in (2.2b) by samples from the distribution $p(\boldsymbol{\theta}, \gamma^{mis}, \boldsymbol{\gamma}^{pre} \mid \boldsymbol{U})$, where $\boldsymbol{\gamma}^{pre}$ is a set of pre-estimated values for $\gamma(t)$ at t_i. That is, instead of targeting the posterior distribution for the process model in (1.2), our algorithm targets the posterior distribution for a conditional SPEKF model where the $\gamma(t)$ process is restricted to pass through the pre-estimated endpoints. This procedure may introduce a bias in the sampled parameters, but at the same time greatly improves the efficiency of the proposal distribution over the infinite-dimensional $\gamma(t)$ paths. Since SPEKF models have been designed from the outset as surrogate models for complex partially observed processes, it is acceptable to incur a small bias in the estimated parameters if the predictive skill (which is the ultimate goal in this context) is high. Indeed, as we demonstrate in section 4, simplifed SPEKF models with parameters estimated via our proposed approximate MCMC strategy have high predictive skill in both perfect- and imperfect-model scenarios.

3.1. Pre-estimating endpoints for the missing process. In a first step, carried out offline, we estimate "plausible" values for the process $\{\gamma(t)\}$ at the observation times t_i. These values are computed by assuming that the observed signal $u(t)$ over the interval $[0, T]$ is governed by an MSM

$$du(t) = -\gamma^b u(t)\, dt + \sigma_u^b\, dW_u(t). \tag{3.1}$$

The role of this model is to estimate "background" damping and noise parameters, γ^b and σ_u^b, respectively. The parameter σ_u^b can in fact be estimated consistently using the quadratic variation of the process $\{u(t)\}$, as long as the observations are not too sparse. Let

$$\hat{\sigma}_u^b = \sqrt{\frac{1}{T}\sum_{i=1}^{N}(U_i - U_{i-1})^2}$$

be a background estimate of σ_u. Given $\hat{\sigma}_u^b$, we use a maximum likelihood estimate for γ^b, based on the background likelihood function

$$p_{msm}(\boldsymbol{U} \,|\, \gamma^b) = \prod_{i=1}^{n} p(U_i \,|\, U_{i-1}, \gamma^u),$$

which is available in closed form, given that (3.1) is an OU process. Let

$$\hat{\gamma}^b = \text{argmax}_{\gamma^b \in \ldots} p_{msm}(\boldsymbol{U} \,|\, \gamma^b). \tag{3.2}$$

Consequently, we assume that in each subinterval $[t_{i-1}, t_i]$ the $u(t)$ process is governed by a local MSM,

$$du(t) = -\bar{\gamma}_i u(t)\, dt + \hat{\sigma}_u^b\, dW_u(t), \tag{3.3}$$

where $\hat{\sigma}_u^b$ is fixed and estimated as above, and $\bar{\gamma}_i$ is estimated as the maximizer of the local posterior distribution,

$$p(\bar{\gamma}_i \mid U_{i-1}, U_i) \propto p(\bar{\gamma}_i)\, p(U_i \mid U_{i-1}, \bar{\gamma}_i). \tag{3.4}$$

Here, we assume that the prior density $p(\bar{\gamma}_i)$ is Gaussian with mean $\hat{\gamma}^b$ estimated from (3.2) and a large variance c_γ^b. Our choice is motivated by two reasons. First, we have limited prior knowledge on $\bar{\gamma}_i$ and therefore the prior distribution should reflect that uncertainty. In addition, we do expect that some of the estimated $\bar{\gamma}_i$ values are negative to reflect the intermittent instability in the process $\{u(t)\}$. We note that the local posterior distribution (3.4) is Gaussian, and the maximizer of $p(\bar{\gamma}_i \mid U_{i-1}, U_i)$ can be computed directly.

Let

$$\widehat{\bar{\gamma}}_i = \text{argmax}\; p(\bar{\gamma}_i \mid U_{i-1}, U_i)$$

be the estimate of $\bar{\gamma}_i$ over the interval $[t_{i-1}, t_i]$. The average values $\Gamma_i = (\widehat{\bar{\gamma}}_{i-1} + \widehat{\bar{\gamma}}_i)/2$ in two neighboring subintervals $[t_{i-1}, t_i]$ and $[t_i, t_{i+1}]$ are taken as *preconditioning values* of the endpoints of γ^{mis} at t_i for $i = 1, \ldots, n-1$. For the first and last endpoint ($i = 0$ and $i = n$), we set $\Gamma_i = \widehat{\bar{\gamma}}_i$. An illustration of this procedure is provided in Figure 1.

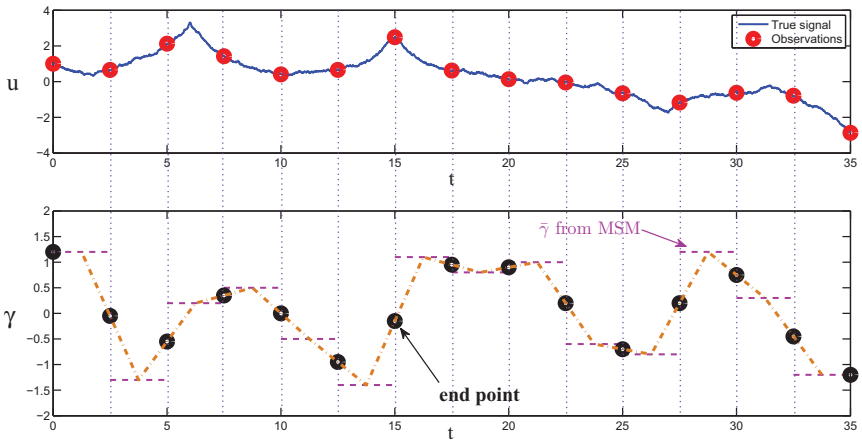

Figure 1. *Illustration of the pre-estimation of the endpoints of γ using an MSM.*

What remains is to switch the estimated endpoints of γ^{mis} to those of α^{mis}. To do this, we regard $\mathbf{\Gamma} = (\Gamma_1, \ldots, \Gamma_n)$ as "observations" of the $\gamma(t)$ process and estimate a background diffusion coefficient $\hat{\sigma}_\gamma^b$ using the quadratic variation

$$\hat{\sigma}_\gamma^b = \sqrt{\frac{1}{T}\sum_{i=1}^{N}(\Gamma_i - \Gamma_{i-1})^2}.$$

Having the estimated diffusion coefficient $\hat{\sigma}_\gamma^b$, the endpoints of γ^{mis} are switched to those of α^{mis} via

$$\boldsymbol{A} = (A_1, \ldots, A_n) = \mathbf{\Gamma}/\hat{\sigma}_\gamma^b. \tag{3.5}$$

With these in mind, we are ready to introduce our algorithm via preconditioning.

3.2. Proposed algorithm. A key feature of the new algorithm is to use the precondition values $\boldsymbol{A}$ in (3.5) to construct an informative prior distribution for the unknown quantities. To that end we split the path of the process α^{mis} as

$$\alpha^{mis} = \alpha^{pre} \cup \alpha_-^{mis},$$

where

$$\begin{aligned}\alpha^{pre} &= [\alpha^{mis}(t_0)\ , \alpha^{mis}(t_1), \ldots, \alpha^{mis}(t_n)]\ ,\\ \alpha_-^{mis} &= \alpha_{(t_0,t_1)}^{mis} \cup \cdots \cup \alpha_{(t_{n-1},t_n)}^{mis}\ .\end{aligned}$$

Moreover, for each subinterval $t \in (t_{i-1}, t_i)$ we consider a conditional SPEKF model

$$\begin{aligned} du(t) &= -\sigma_\gamma \alpha(t) u(t)\, dt + \sigma_u\, dW_u(t),\\ d\alpha(t) &= -\frac{d_\gamma}{\sigma_\gamma}\left(\sigma_\gamma \alpha(t) - \hat{\gamma}\right)\, dt + dW_\gamma(t)\end{aligned} \tag{3.6a}$$

with

$$u(t_{i-1}) = U_{i-1}, \quad u(t_i) = U_i, \quad \alpha(t_{i-1}) = \alpha_{i-1}^{pre}, \quad \alpha(t_i) = \alpha_i^{pre}. \tag{3.6b}$$

Correspondingly, we now aim to explore the posterior distribution

$$p(\boldsymbol{\theta}, \alpha^{pre}, \alpha_-^{mis} \,|\, \boldsymbol{U}) \propto p(\boldsymbol{\theta}) p(\alpha^{pre}) p(\alpha_-^{mis} \,|\, \boldsymbol{\theta}, \alpha^{pre}) p(\boldsymbol{U} \,|\, \alpha^{mis}, \boldsymbol{\theta}) \tag{3.7}$$

with the following prior specifications:

- $p(\boldsymbol{\theta})$ is an uninformative Gaussian prior distribution.
- $p(\alpha^{pre})$ is an *informative* Gaussian distribution, centered at $\boldsymbol{A}$ and a diagonal covariance matrix, with small variances.
- $p(\alpha_-^{mis} \,|\, \alpha^{pre}, \boldsymbol{\theta})$ is a product of OU bridges which are determined by $\boldsymbol{\theta}$ and the endpoints given by α^{pre}. The details on how to construct and simulate such OU bridges are given in [2] and in the appendix.

We note that this prior specification is different from that induced by the OU process (1.2b) in that the marginal distribution of α^{pre} is now forced to be a very informative Gaussian distribution. We now describe the Metropolis–Hastings updates.

Our algorithm alternates between updating the parameters $\boldsymbol{\theta}$ and the path $(\alpha^{pre}, \alpha_-^{mis})$. We initialize the algorithm by simulating $\boldsymbol{\theta}^{\{0\}} \sim p(\boldsymbol{\theta})$ and $\alpha^{pre,\{0\}} \sim p(\alpha^{pre})$. Consequently, we generate a path $\alpha_-^{mis,\{0\}}$ by simulating a collection of OU bridges using $\alpha^{pre,\{0\}}$ as endpoints. Given $(\boldsymbol{\theta}^{\{k\}}, \alpha^{pre,\{k\}}, \alpha_-^{mis,\{k\}})$, $k \geq 0$, we simulate $\boldsymbol{\theta}^*$ from a proposal distribution $\boldsymbol{\theta}^* \sim g(\cdot \,|\, \boldsymbol{\theta}^{\{k\}})$. The proposed state $\boldsymbol{\theta}^*$ is accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\theta}^*) p(\alpha_-^{mis,\{k\}}, \alpha^{pre,\{k\}} \,|\, \boldsymbol{\theta}^*) p(\boldsymbol{U} \,|\, \alpha^{mis,\{k\}}, \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{\{k\}}) p(\alpha^{mis,\{k\}}, \alpha^{pre,\{k\}} \,|\, \boldsymbol{\theta}^{\{k\}}) p(\boldsymbol{U} \,|\, \alpha^{mis,\{k\}}, \boldsymbol{\theta}^{\{k\}})} \cdot \frac{g(\boldsymbol{\theta}^{\{k\}} \,|\, \boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^* \,|\, \boldsymbol{\theta}^{\{k\}})}\right\}. \tag{3.8}$$

We note that the prior distribution $p(\alpha^{pre})$ is constructed by first estimating the $\boldsymbol{A}$ values. This is done through the MSM and using the observations. Our motivation for this approach is twofold. First, we want to exploit the conditional Gaussian structure of the SPEKF model and avoid having to construct complex proposal distributions for the missing paths. Second, we inform the prior on the skeleton α^{pre} in order to help with the exploration of the path space for the OU bridges α_-^{mis}.

3.3. Updating the missing path. To improve the mixing time of our algorithm and ensure continuity of the missing path at each iteration, we update $a^{mis,\{k\}}$ in temporally interleaving blocks of duration $2\,\Delta t$. Specifically, we propose a new path $a^{mis,*}$ using the following procedure:

1. If $k+1$ is odd, select the even time indices $I = \{0, 2, \dots, \}$; otherwise, select the odd indices $I = \{1, 3, \dots\}$.
2. For each $i \in I$, simulate a new endpoint
$$\alpha_i^{pre,*} \sim p_i(\alpha^{pre}). \tag{3.9}$$
3. For each $i \in I$, simulate two adjacent OU bridges $\alpha_{[t_{i-1},t_i]}^{mis,*}$ and $\alpha_{[t_i,t_{i+1}]}^{mis,*}$ with endpoints $\alpha_{i-1}^{pre,\{k\}}, \alpha_i^{pre,*}$ and $\alpha_i^{pre,*}, \alpha_i^{pre,\{k\}}$, respectively. The path segment $\alpha_{[t_{i-1},t_i]}^{mis,*}$ is

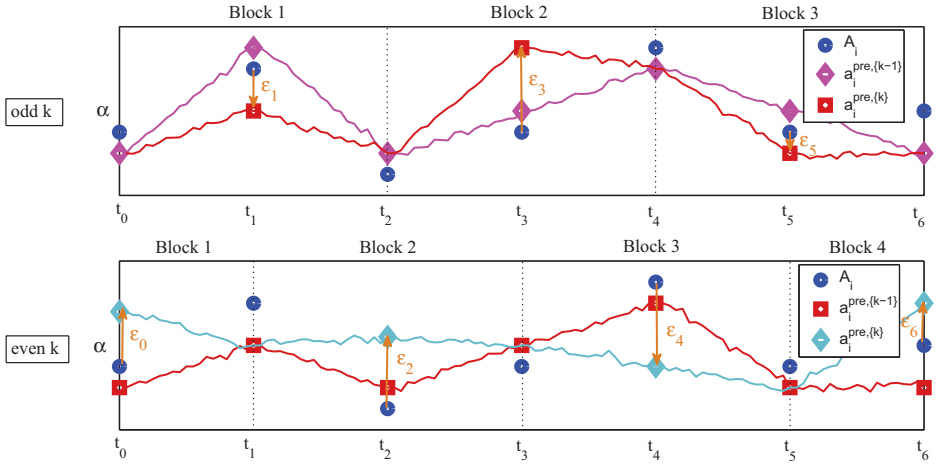

Figure 2. *Illustration of the update procedure for the missing path* α^{mis}.

accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{U} \mid \alpha^{mis,*}_{[t_{i-1},t_i]}, \boldsymbol{\theta}^{\{k\}})}{p(\boldsymbol{U} \mid \alpha^{mis,\{k\}}, \boldsymbol{\theta}^{\{k\}})}\right\}. \tag{3.10}$$

Next, we construct the full proposed missing path $\alpha^{mis,*}$ by concatenating the subpaths, i.e.,

$$\alpha^{mis,*} \equiv \alpha^{mis,\{k\}}_{[t_0,t_1]} \cup \alpha^{mis,\{k\}}_{[t_1,t_2]} \cup \cdots \cup \alpha^{mis,*}_{[t_{i-1},t_i]} \cup \alpha^{mis,*}_{[t_i,t_{i+1}]} \cup \cdots \cup \alpha^{mis,\{k\}}_{[t_{n-1},t_n]}.$$

The $\alpha^{mis,*}$ defined above is then used in the model parameter update in (3.8). In practice, we represent all the paths above using a fine grid for every time interval $[t_{i-1}, t_i]$. All deterministic integrals involved in the algorithm are evaluated using a trapezoidal rule. Moreover, the distribution $p_i(\alpha_i^{pre})$ of the endpoints is a Gaussian with zero mean and variance equal to $0.2^2(\sigma^2_{i-1/2} + \sigma^2_{i+1/2})$, where $\sigma^2_{i-1/2}$ is the variance of $p(\bar{\gamma}_i \mid U_{i-1}, U_i)$ in (3.4).

Observe that by virtue of the conditional model structure in (3.6), the ratio of the prior densities will cancel the ratio of the proposal densities in (3.10), i.e.,

$$\frac{p(\alpha^{mis,*}_{-}, \alpha^{pre,*} \mid \boldsymbol{\theta}^{\{k\}})}{p(\alpha^{mis,\{k\}}, \alpha^{pre,\{k\}} \mid \boldsymbol{\theta}^{\{k\}})} \cdot \frac{g(\alpha^{mis,\{k\}}, \alpha^{pre,\{k\}} \mid \boldsymbol{\theta}^{\{k\}})}{g(\alpha^{mis,*}_{-}, \alpha^{pre,*} \mid \boldsymbol{\theta}^{\{k\}})} = 1, \tag{3.11}$$

since the new path is proposed independently of the old path by simulating from the prior distribution. In particular, (3.11) holds only approximately in the case of the standard (unconditional) SPEKF model in (1.2).

We refer the reader to Figure 2 for an illustration of our block update strategy. As remarked earlier, our updates maintain the continuity of the sample paths corresponding to the $\alpha^{mis}(\cdot)$ process. Our new sampling algorithm is summarized below.

MCMC ALGORITHM WITH PRECONDITIONING.

1. Compute the precondition values A_i for the missing process from (3.5) using the preconditioning algorithm of section 3.1.
2. Set the number of iterations K. Set $k = 0$. Select the initial parameters $\boldsymbol{\theta}^{\{0\}}$, and construct the initial path $\alpha^{mis,\{0\}} = \bigcup_i \alpha^{mis,\{0\}}_{[t_i,t_{i+1}]}$, where $\alpha^{mis,\{0\}}_{[t_i,t_{i+1}]}$ are OU bridges in the interval $[t_i, t_{i+1}]$ with endpoints $[\alpha_i^{pre,\{0\}}, \alpha_{i+1}^{pre,\{0\}}]$ given by (3.5).
3. In step $k+1$, draw the endpoints of $\alpha_i^{pre,*}$ from (3.9).
4. Construct OU bridges $\alpha^{mis,*}_{[i,i+2]}$ with endpoints $[\alpha_i^{pre,\{k\}}, \alpha_{i+2}^{pre,\{k\}}]$ over the intervals $[t_0, t_2], [t_2, t_4], \ldots$ if $k+1$ is odd or $[t_0, t_1], [t_1, t_3], [t_3, t_5], \ldots$ if $k+1$ is even. If the total number of subintervals is odd, then the last block contains only one subinterval.
5. Accept each of the proposed OU bridges with the probability given in (3.10). If $\alpha^{mis,*}_{[i,i+2]}$ is rejected, then replace it by $\alpha^{mis,\{k\}}_{[i,i+2]}$.
6. Propose the parameters $\boldsymbol{\theta}^*$ from some proposal function $g(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{\{k\}})$.
7. Accept $\boldsymbol{\theta}^*$ with probability given by (3.8). If the proposal is rejected, then set $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{\{k\}}$.
8. Set $k = k+1$. Go to step 3, and repeat until $k+1 = K$.

4. Numerical tests. We discuss applications of the algorithm developed in section 3 to three experiments involving intermittent signals, one in a perfect model setting and two in a scenario where the signal is generated by a model which lies outside the simplified SPEKF family (1.2). In addition to parameter estimation, we are interested in assessing the skill of the SPEKF model estimated through MCMC to reproduce the statistics of test data which are not part of the dataset used for parameter estimation.

4.1. Information-theoretic measures of predictive skill and model error. We characterize predictive skill and model error using techniques from information theory [5, 26, 17, 29]. In particular, we consider that the statistics of the observed signal $u(t)$ at the forecast lead time t given initial data $u_0 = u(0)$ are described by a time-dependent PDF $p_t(u \mid u_0)$ with a well-defined equilibrium $p_{eq}(u) = \lim_{t\to\infty} p_t(u \mid u_0)$. The MCMC-estimated SPEKF model produces forecast PDFs $p_t^M(u \mid u_0)$ and $p_{eq}^M(u) = \lim_{t\to\infty} p_t^M(u \mid u_0)$, but these PDFs will in general differ from the PDFs of the true signal. Throughout, we use the term "perfect model" to indicate the model that generates the true signal, and "imperfect model" to represent the simplified SPEKF model with the parameters estimated through the MCMC algorithm.

Here, we employ three metrics for model assessment which measure (1) predictability of $u(t)$ in the perfect model; (2) internal predictive skill of $u(t)$ in the MCMC-estimated SPEKF model; and (3) model error of the estimated model relative to the perfect model. These metrics are defined as follows.

Intrinsic predictability, $\mathcal{D}_t$. This metric quantifies the information provided by the initial conditions about the future state of the system beyond the prior knowledge available through equilibrium statistics. This information gain is computed through the relative entropy between the time-dependent and equilibrium PDFs, i.e.,

$$\mathcal{D}_t = \mathcal{P}(p_t, p_{eq}), \tag{4.1}$$

where the functional $\mathcal{P}(p,q) = \int p \log(p/q)$ is the relative entropy between two probability measures.

Internal predictive skill, $\mathcal{D}_t^M$. In direct analogy with (4.1), we introduce the relative-entropy metric

$$\mathcal{D}_t^M = \mathcal{P}(p_t^M, p_{eq}^M), \tag{4.2}$$

which measures information gain in the time-dependent forecast PDF of the imperfect model beyond the model's equilibrium. $\mathcal{D}_t^M$ will convey false predictability if either of p_t^M or p_{eq}^M are biased away from the truth. Nevertheless, it is an important metric for model assessment since it measures potential initial-value predictive skill in the model relative to a trivial forecast drawn from the model's equilibrium.

Model error, $\mathcal{E}_t$. This metric measures the lack of information in the imperfect model density compared to a perfect statistical forecast. It is defined as

$$\mathcal{E}_t = \mathcal{P}(p_t, p_t^M). \tag{4.3}$$

In the following numerical tests, the initial value u_0 is set to the mean value at equilibrium. In separate calculations, we have confirmed that different u_0 values lead to qualitatively similar results.

4.2. Parameter estimation in a perfect-model setting. The first numerical test deals with a perfect-model environment where both the perfect model that generates the signal and the imperfect model for parameter estimation and prediction are in the simplified SPEKF model class. Thus, model error in this example is solely due to poor parameter estimation.

To reflect the wave-like behavior of many signals, a deterministic phase is introduced in the dynamics of the resolved variable u, leading to the simplified SPEKF model,

$$du(t) = (-\gamma(t) + i\omega)u(t)\,dt + \sigma_u\,dW_u(t), \tag{4.4a}$$

$$d\gamma(t) = -d_\gamma\big(\gamma(t) - \hat{\gamma}\big)\,dt + \sigma_\gamma\,dW_\gamma(t). \tag{4.4b}$$

Note that there are five unknown parameters in this system. The algorithm for parameter estimation remains the same as that without phase, except that the variable $u(t)$ and its associated stochastic forcing $dW_u(t)$ become complex, and both ω and σ_u need to be incorporated in the MSM (3.1) for preconditioning.

To generate a signal with intermittency, we set the perfect-model parameter values $\hat{\gamma}^* = 0.8$, $\sigma_u^* = 0.15$, $\sigma_\gamma^* = 0.7$, $d_\gamma^* = 0.5$, and $\omega = 2$. The observation time step is $\Delta t = 0.5$, which is shorter than the averaged decorrelation time of u, $\tau_{corr}^u = 1/d_\gamma^* = 2$. The training time series, shown in Figure 3, has length $T = 250$ and thus contains 501 observations. The large burst of u around $t = 75$, 200, and 225, which corresponds to a transient phase of negative γ, reflects the intermittent instability. The intermittency occurrence is not observed with a high frequency, but the amplitude of the intermittency is large. Thus, this type of intermittent instability produces "black swan"-like events.

The variance of the prior distribution (3.4) of $\bar{\gamma}$ for preconditioning is chosen to be $c_v^b = 4$ so that $\bar{\gamma}$ has access to negative values. The scaling coefficient σ_b^γ to convert the endpoints of

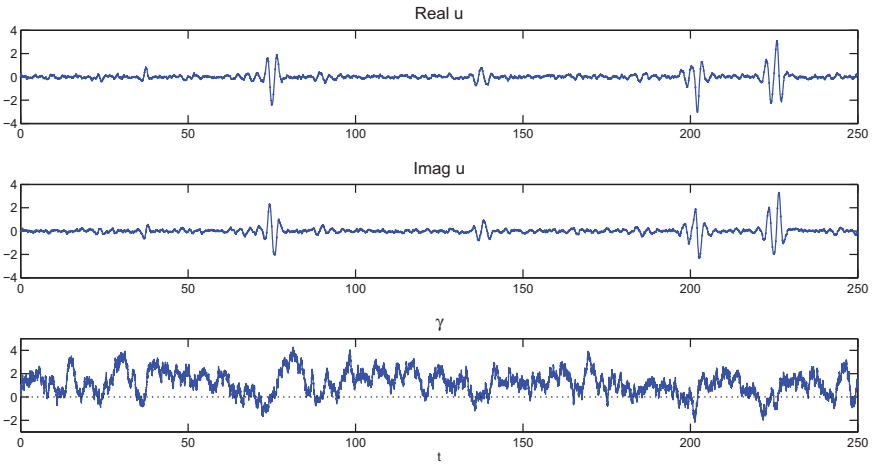

Figure 3. *Training time series for the experiment with no model error. The top and middle panels show the real and imaginary parts of the resolved variable $u(t)$. The bottom panel shows the unresolved stochastic damping $\gamma(t)$. Only observations of $u(t)$ (indicated by point markers) are used for parameter estimation via MCMC. The $\gamma(t)$ process is hidden from the algorithm.*

γ^{mis} to α^{mis} via (3.5) is estimated as $\sigma_\gamma^b = 0.64$. The prior distributions of the five parameters are

$$\sigma_u \sim \Gamma(2, 1/2), \quad \sigma_\gamma \sim \Gamma(2,1), \quad \omega \sim \Gamma(2,1), \quad \hat{\gamma} \sim \mathcal{N}(2,2), \quad d_\gamma \sim \mathcal{N}(2,1), \tag{4.5}$$

where Γ and $\mathcal{N}$ are the PDFs of the Gamma and Gaussian distributions, respectively,

$$\Gamma_x(k,\theta) = \frac{1}{\theta^k}\frac{1}{\Gamma(k)}x^{k-1}e^{-\frac{x}{\theta}}, \quad \mathcal{N}_x(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{4.6}$$

Note that apart from prescribing the sign of the diffusions and phase, the prior distribution has almost no influence on the parameter estimation with our new sampling algorithm. The proposal function $g(\theta)$ in the MCMC algorithm is set to a Gaussian with zero mean and standard deviation $1/4$ and is evaluated for each component separately. The initial MCMC iterates $\boldsymbol{\theta}^{\{0\}}$ are all set to be twice their true values. The missing paths are generated using an Euler numerical scheme with time step $\Delta t = 0.01$.

Figure 4 displays MCMC trace plots of the five parameters together with the estimated equilibrium variance of γ. The acceptance rates of $\hat{\gamma}$ and d_γ are around 70%, and that of σ_u, σ_γ, and ω are around 40%. Moreover, the acceptance rate of the missing path α^{mis} is around 75%, which is a significant improvement compared to the $< 0.1\%$ acceptance rate of the direct algorithm (see section 2). As shown in Figure 5, the parameters decorrelate after approximately 50 iterations. Thus, the new algorithm provides a well-mixed Markov chain.

Figure 6 displays the true value, the given prior distribution, and the posterior PDFs of the five SPEKF parameters, as well as the equilibrium PDF of γ corresponding to the maximum a posteriori estimates of σ_γ and d_γ. All parameters are estimated accurately and have small uncertainty. In particular, the equilibrium variance of γ in the imperfect model,

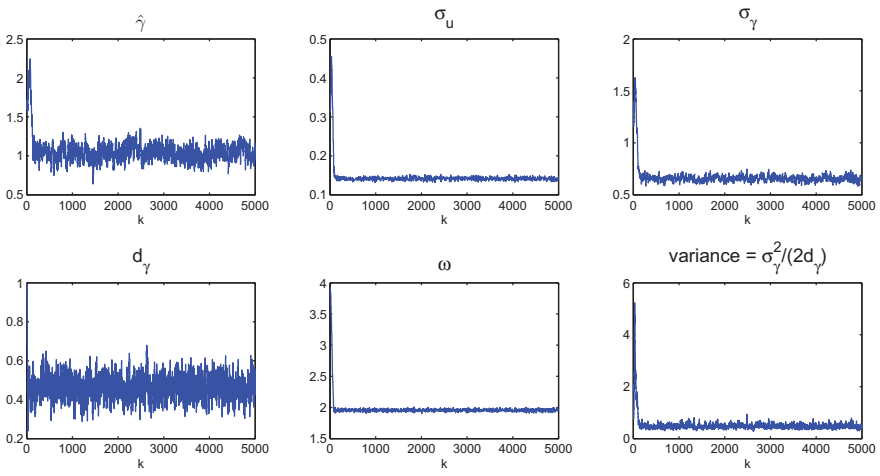

Figure 4. *Trace plots of the parameters of the simplified SPEKF system (4.4) estimated via the new MCMC algorithm of section 3, where the true signal is shown in Figure 3. Here, k denotes the MCMC iteration.*

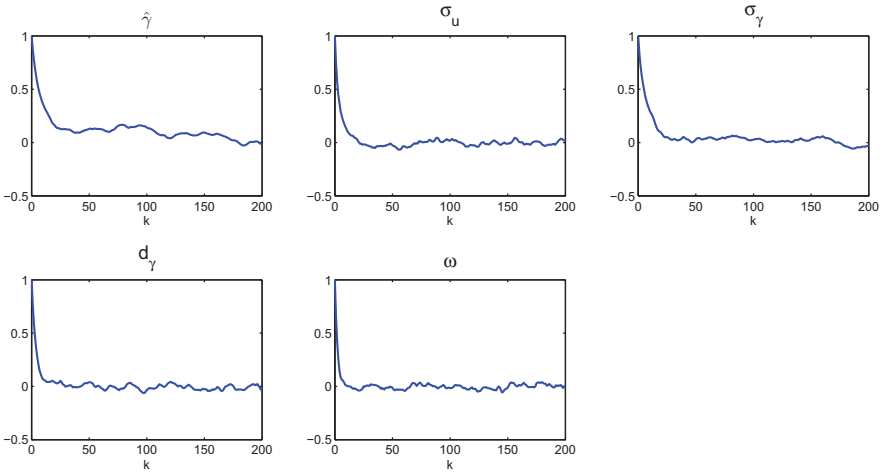

Figure 5. *Autocorrelation functions of the parameter traces of Figure 4 computed after a burn-in period of* 1000 *iterations.*

i.e., $\mathrm{var}(\gamma) = \sigma_\gamma^2/(2d_\gamma)$, agrees well with the true value, $\sigma_\gamma^*/(2d_\gamma^*)$. As a result, the equilibrium distribution of γ has access to negative values, enabling the imperfect model to produce intermittency.

The imperfect model with the parameters from maximum a posteriori estimates via the MCMC algorithm is able to reproduce both the equilibrium and off-equilibrium statistics of the perfect model with high skill. As shown in Figure 7, the autocorrelation functions $\rho_u(t)$ and $\rho_u^M(t)$ for u in the perfect and imperfect models, respectively, oscillate at the same frequency, with ρ_u^M decaying only slightly slower than than ρ_u. Moreover, the internal predictive skill metric $\mathcal{D}_t^M$ from (4.2) matches well with the intrinsic predictability from (4.1). Both $\mathcal{D}_t$ and

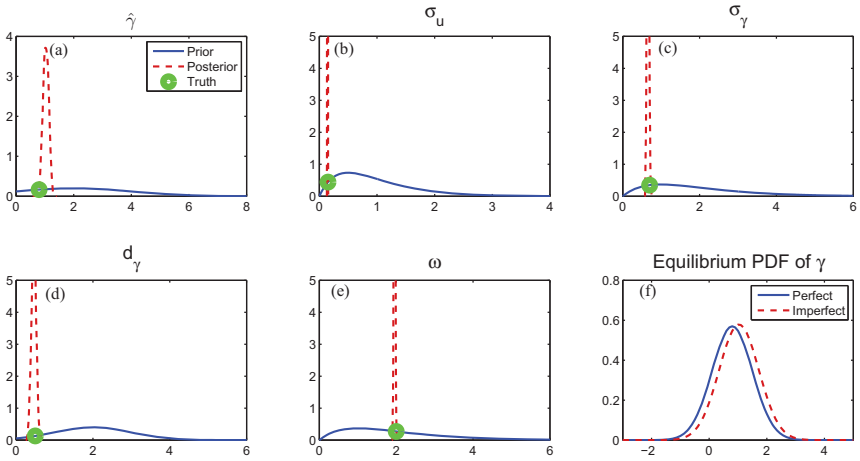

Figure 6. *Parameter estimation for the simplified SPEKF system (4.4) performed via the new MCMC algorithm. Panels* (a)–(e) *show the prior distribution (solid lines), posterior distribution from the MCMC algorithm (dashed lines), and the true value of the parameters (circles) of* $\hat{\gamma}$, σ_u, σ_γ, d_γ, *and* ω, *respectively. Panel* (f) *compares the equilibrium PDF of the unresolved variable* $\gamma(t)$ *of the perfect model and the model equipped with the maximum a posteriori estimates of the parameters.*

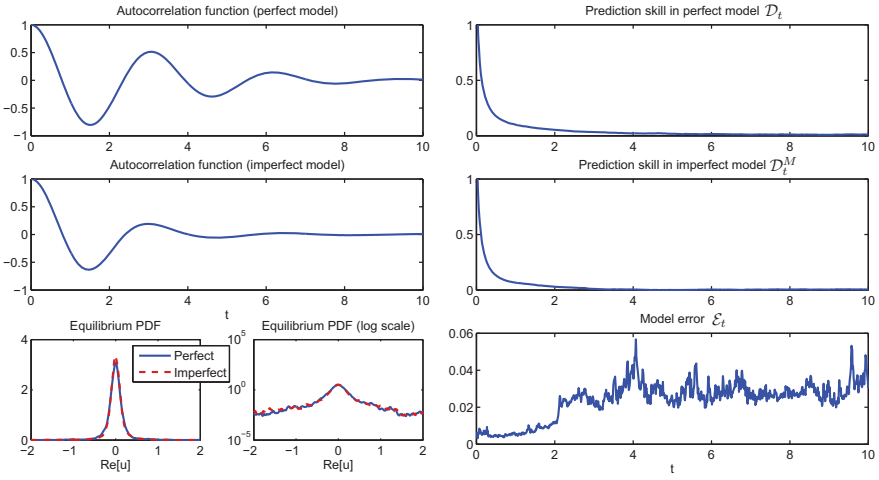

Figure 7. *Perfect-model predictability (left) and predictive skill of the imperfect model equipped with the maximum a posteriori estimated parameter values from Figure 6. The top and middle panels show the autocorrelation functions* $\rho_u(t)$ *and* $\rho_u^M(t)$, *the perfect-model predictability score* $\mathcal{D}_t$ *from (4.1), and internal prediction skill* $\mathcal{D}_t^M$ *from (4.2), respectively. The bottom right panel shows the evolution of the model error* $\mathcal{E}_t$ *from (4.3). The bottom left panels display the equilibrium PDF of the perfect model and the model (solid) with the estimated parameters (dashed) in both linear and logarithmic scales.*

$\mathcal{D}_t^M$ are significantly larger than the model error $\mathcal{E}_t$ for short-range forecasts with $t \lesssim 0.5$. The medium-range ($0.5 \lesssim t \lesssim 2$) prediction skill of the imperfect model is slightly worse than the prefect model, but still larger than $\mathcal{E}_t$. Both the perfect and imperfect model equilibrate around $t = 3.0$, at which point the model error in the equilibrium distribution $p_{eq}^M(u)$ becomes

negligible. In particular, the imperfect model succeeds in capturing the fat tails of the true distribution $p_{eq}(u)$, which are the outcomes of extreme events. We remark that the residual errors in the estimated parameters are possibly due to the shortness of the training time series. More accurate estimation should be achievable using longer training time series.

4.3. Parameter estimation with model error. The more realistic situation for parameter estimation and prediction is in the presence of model error, i.e., for $u(t)$ signals generated by models which are not in the simplified SPEKF class. Here, we discuss two such experiments where intermittent instability in $u(t)$ is the outcome of a two-state Markov jump process (section 4.3.1) and a nonlinear stochastic process with correlated additive and multiplicative (CAM) noise (section 4.3.2).

4.3.1. Intermittent instability from a Markov jump process. In this application, the unresolved variable $\gamma(t)$ of the perfect model that generates the signal is assumed to be driven by a two-state Markov jump process, i.e.,

$$\begin{aligned} du(t) &= (-\gamma(t) + i\omega)u(t)\,dt + \sigma_u\,dW_u(t), \\ \gamma(t) &\text{ satisfies a two-state Markov jump process.} \end{aligned} \tag{4.7}$$

The system (4.7) features regime switching and can be utilized to mimic unresolved baroclinic instabilities in the atmosphere or ocean [28]. Here, the damping $\gamma(t)$ switches between the stable phase $\gamma^+ = 2.27$ and the unstable phase $\gamma^- = -0.04$. The switching rate from the stable to unstable phase is $\nu = 0.1$ and that from the unstable to stable phase is $\mu = 0.2$. Therefore, the time-averaged damping is given by $\bar{\gamma} = (\nu\gamma^- + \mu\gamma^+)/(\mu + \nu) = 1.5$. The phase and diffusion coefficients are $\omega = 1.782$ and $\sigma_u = 0.1095$, respectively.

The observed $u(t)$ signal for parameter estimation and the underlying $\gamma(t)$ process are shown in Figure 8, where negative $\gamma(t)$ values corresponds to intermittent instability. The observation time step is $\Delta t = 0.5$, which is shorter than the decorrelation time of $u(t)$. The length of the observed time series is $T = 250$, corresponding to 501 observations. The variance of $\bar{\gamma}$ in the preconditioning algorithm of section 3 is set to $c_v^b = 6$ to ensure access to negative values. The scaling coefficient in (3.5) is around $\sigma_\gamma^b = 0.8$. The prior distributions of the parameters are the same as in (4.5).

Figure 9 shows a proposal of the missing path γ^{mis} constructed by OU bridges in step 6 of the new algorithm (after rescaling $\alpha^{mis,\{k\}}$ by σ_γ^b). Clearly, the proposed missing path has a similar pattern as the true signal, and therefore it succeeds in recovering the intermittency. Figures 10–12 show trace plots and autocorrelation functions of the MCMC iterations, and the PDFs for parameter estimation obtained via the new algorithm, respectively. Similar to the example in section 4.2, the new algorithm has a high acceptance rate and is able to recover the intermittent instability.

Figure 13 displays the prediction skill resulting from maximum a posteriori estimates of the parameters obtained via the new algorithm. The autocorrelation function of the imperfect model has the same oscillation frequency as that of the perfect model, but it decays slightly faster than its perfect model counterpart. Moreover, the model error $\mathcal{E}_t$ from (4.3) in the equilibrium PDF $p_{eq}^M(u)$ is small. In particular, the imperfect model has a similar fat-tailed density as the equilibrium PDF $p_{eq}(u)$ in the perfect model, which implies that the imperfect model is able to capture the extreme events. The imperfect model has good short-range

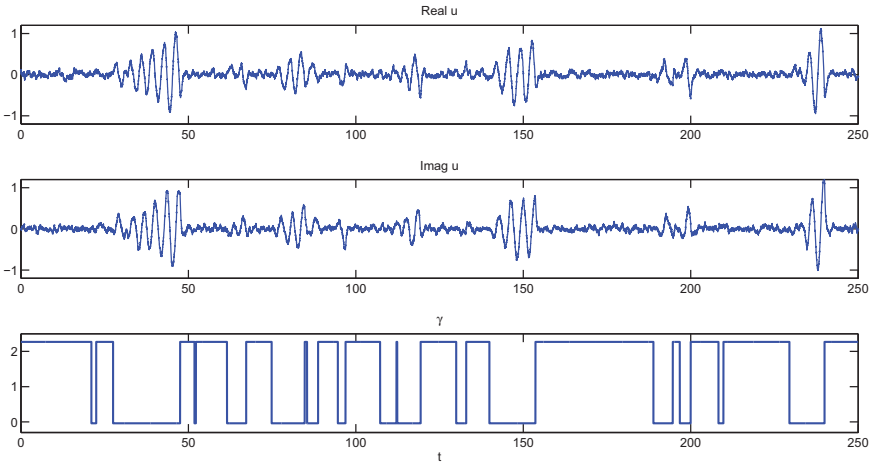

Figure 8. *True signal generated by system (4.7) featuring a two-state Markov jump process. The top and middle panels show the true signal of the real and imaginary parts of the resolved variable u and the bottom panel shows that of the unresolved variable γ, which is a two state Markov jump process with stable phase $\gamma^+ = 2.27$ and unstable phase $\gamma^- = -0.04$. Only observations of u (indicated by point markers) are used for parameter estimation via MCMC. The γ process is hidden from the algorithm.*

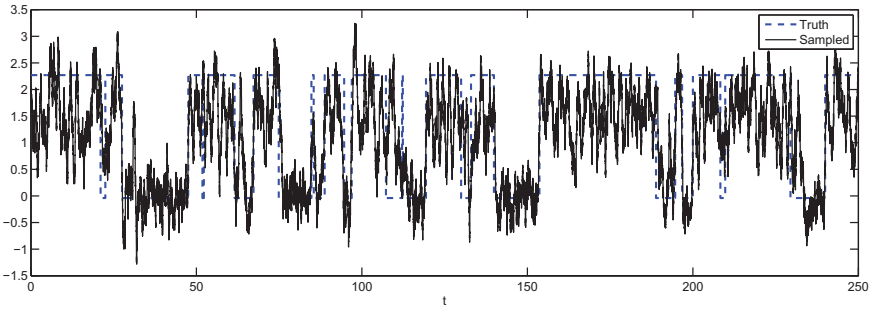

Figure 9. *A sample of the missing path γ^{mis} using the new algorithm (solid) compared with the true two-state Markov jump process signal (dashed) from (4.7).*

internal prediction $\mathcal{D}_t^M$ from (4.2) skill for $t < 0.70$. However, its medium-range forecasts from that model are not accurate in the sense that $\mathcal{D}_t^M$, which decays slightly faster than the perfect model predictability score $\mathcal{D}_t$, is exceeded by the model error. Nevertheless, the model error remains small at later times, giving a good estimation of the equilibrium distribution.

4.3.2. Intermittent instability with correlated additive-multiplicative noise. We have also studied an example in which the unresolved variable is driven by a nonlinear and non-Gaussian dynamics with cubic nonlinearity and CAM noise [27],

$$du(t) = (-\gamma(t) + i\omega_u)u(t)\,dt + \sigma_u\,dW_u(t), \tag{4.8a}$$

$$d\gamma(t) = (-a\gamma(t) + b\gamma^2(t) - c\gamma^3(t) + f_\gamma)\,dt + (A - B\gamma(t))\,dW_c(t) + \sigma\,dW_\gamma(t), \tag{4.8b}$$

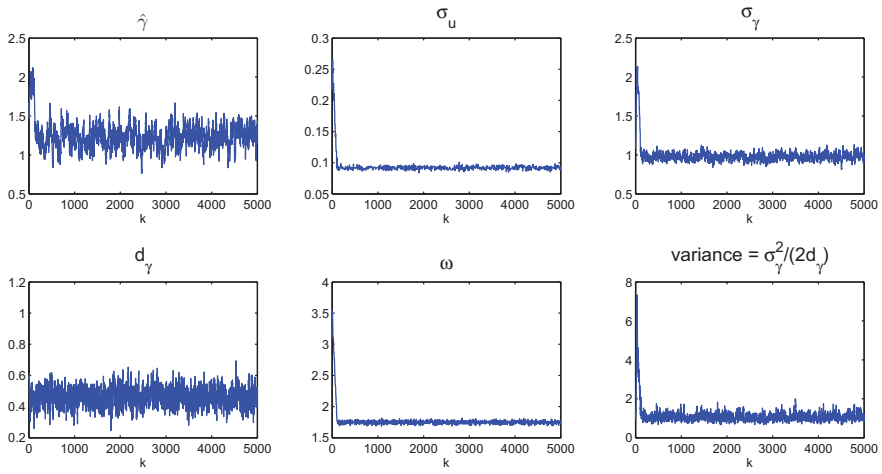

Figure 10. *Trace plots of the parameters of the simplified SPEKF system (4.4) estimated via the new MCMC algorithm of section 3, where the true signal is generated by system (4.7) and is shown in Figure 8. Here, k denotes the MCMC iteration.*

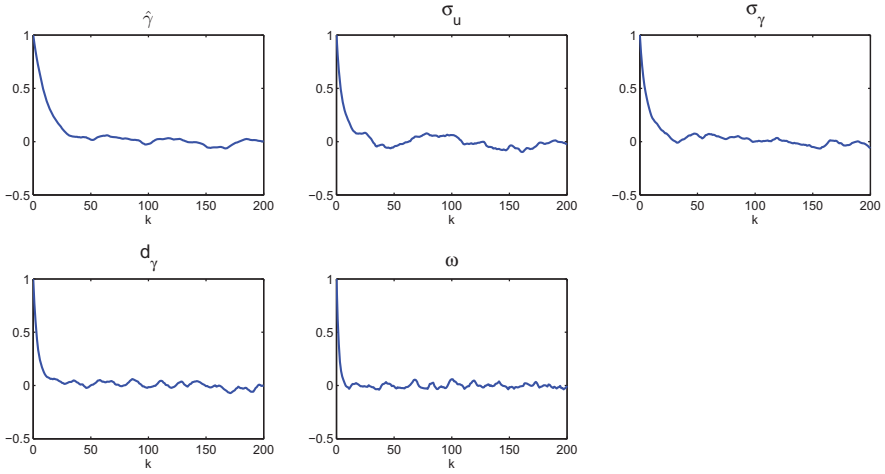

Figure 11. *Autocorrelation functions of the parameter traces of Figure 10 after a burn-in period of* 1000 *samples.*

We selected a dynamical regime in which $\gamma(t)$ is bimodal [see Figure 15(f)]. The training time series is shown in Figure 14. There, $\gamma(t)$ is qualitatively similar to two-state Markov jump process of section 4.3.1 but has a continuous path. The observation time is $\Delta t = 0.5$, which is again shorter than the averaged decorrelation time of $u(t)$. The theoretic optimized parameters σ_γ, d_γ, and $\hat{\gamma}$ can be computed by matching the mean, variance, and decorrelation time of the cubic model (4.8b) with those of SPEKF model (4.4b) and are illustrated by the green dot in Figure 15.

Parameter estimation results from the new algorithm are shown in Figure 15. The parameters $\hat{\gamma}$, σ_u, and ω are estimated with high accuracy. Both damping d_γ and diffusion σ_γ are

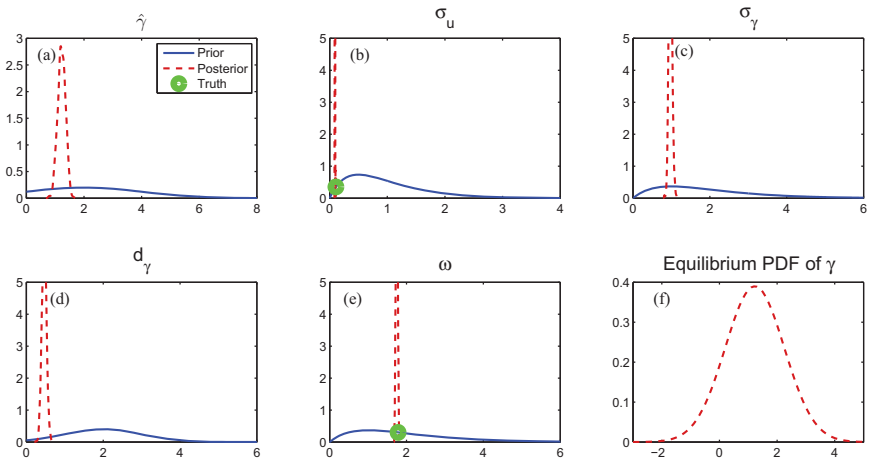

Figure 12. *Estimation of SPEKF parameters via the new algorithm for an observed signal generated by the Markov jump process system (4.7). Panels* (a)–(e) *show the prior distribution (solid line), posterior distribution from MCMC algorithm (dashed line), and the true parameter values if available (circle) of* $\hat{\gamma}$, σ_u, σ_γ, d_γ, *and* ω, *respectively. Panel* (e) *shows the equilibrium PDF of the unresolved variable* γ *of the SPEKF model equipped with maximum a posteriori estimates of the parameters.*

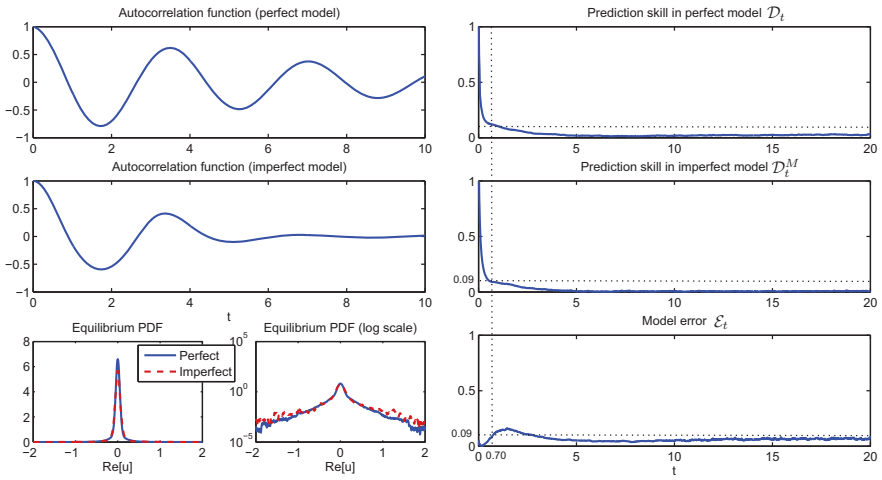

Figure 13. *Perfect-model predictability (left) and predictive skill of the imperfect model equipped with the maximum a posteriori estimated parameter values from Figure 12. The true signal in 8 is generated by system (4.7). The top and middle panels show the autocorrelation functions* $\rho_u(t)$ *and* $\rho_u^M(t)$, *the perfect-model predictability score* $\mathcal{D}_t$ *from (4.1), and internal prediction skill* $\mathcal{D}_t^M$ *from (4.2), respectively. The bottom right panel shows the evolution of the model error* $\mathcal{E}_t$ *from (4.3). The bottom left panels display the equilibrium PDF of the perfect model and the model (solid) with the estimated parameters (dashed) in both linear and logarithmic scales.*

slightly overestimated, but the ratio $\sigma_\gamma^2/(2d_\gamma)$, which is the variance of $\gamma(t)$, is very close to the variance of the Gaussian approximation of the perfect model. The prediction skill results (not shown here) are qualitatively similar to those in Figure 13 for system (4.7), but in this

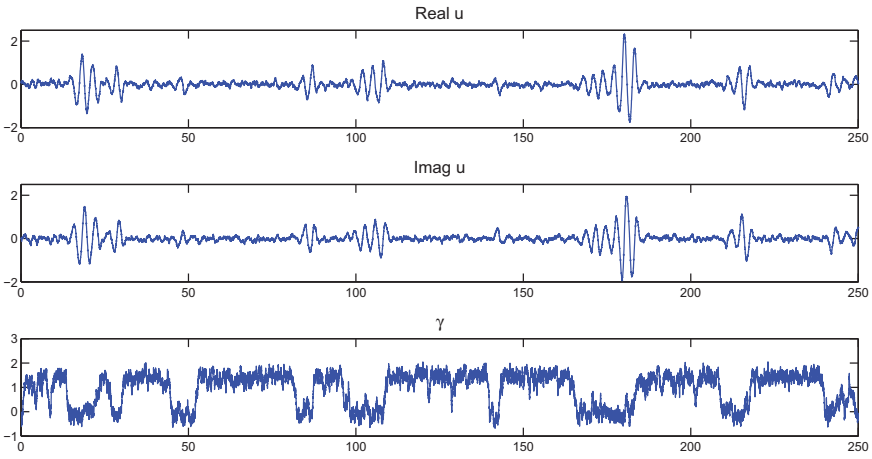

Figure 14. *True signal generated by system (4.8). The top and middle panels show the true signal of the real and imaginary parts of the resolved variable $u(t)$ and the bottom panel shows that of the unresolved variable $\gamma(t)$, the equilibrium distribution of which is bimodal. Only observations of $u(t)$ (indicated by point markers) are used for parameter estimation via MCMC. The $\gamma(t)$ process is hidden from the algorithm.*

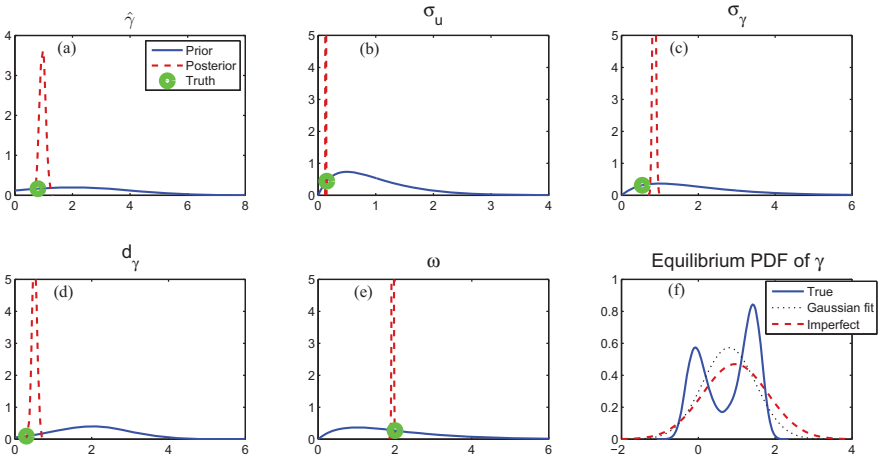

Figure 15. *Estimation of SPEKF parameters via the new algorithm for an observed signal generated by the system (4.8). Panels (a)–(e) show the prior distribution (solid line), posterior distribution from MCMC algorithm (dashed line), and the true parameter values if available (circle) of $\hat{\gamma}$, σ_u, σ_γ, d_γ, and ω, respectively. Panel (e) shows the equilibrium PDF of the unresolved variable $\gamma(t)$ of the SPEKF model equipped with maximum a posteriori estimates of the parameters.*

case the continuity of the $\gamma(t)$ in the perfect model leads to a somewhat smaller model error of the imperfect model.

5. Conclusions. In this paper, the simplified SPEKF model (1.2) is utilized as a low-order process model to approximate signals with hidden intermittent instability. Bayesian inference incorporating data augmentation [39] is applied for parameter estimation in this

class of models via MCMC algorithms. Direct applications of MCMC with data augmentation [34] were found to result in a low acceptance rate of the missing path and poor parameter estimation skill. A new MCMC algorithm was developed, which involves a preconditioning procedure to pre-estimate the unresolved process at the discrete moments that are consistent with the observations of the resolved variable. This new algorithm provides a high acceptance rate of the proposed missing path for data augmentation and produces posterior PDFs for parameter estimation of high accuracy and low uncertainty.

We have performed a suite of numerical tests of the new algorithm in both perfect-model settings and applications where the observed signal is generated by a model which is not of SPEKF type. In all cases, the SPEKF models with parameters estimated via the new algorithm were able to capture the fat-tailed PDFs of the observed signal resulting from extreme events due to hidden intermittent instability. In a challenging application with model error where intermittent instability is generated by a discontinuous two-state Markov jump process, the new algorithm led to SPEKF models of high short-range predictive skill and high fidelity relative to the true-model statistics in both medium- and long-range forecasts.

Future work will involve generalizing the algorithm to perform parameter estimation in the full SPEKF model (1.1), as well as the even more complicated case [6] with a stochastic phase $\omega(t)$ in $u(t)$. The full SPEKF model with stochastic phase has 11 explicit parameters and several implicit parameters in the forcing. Whether some of the parameters are redundant and how to estimate a large number of parameters with only partial observations in $u(t)$ are both important research questions. The application of SPEKF models to parameter estimation and prediction with real-world data with intermittency, such as data for the Madden–Julian oscillation, ENSO, and planetary boundary layers, will be taken into consideration in the future.

Appendix. This section summarizes Lemma 3.1 in [2] for sampling the OU bridge. Consider an OU bridge, which is a solution to the stochastic differential equation

$$dX_t = -\theta X_t dt + \sigma dW_t$$

conditionally on $X_0 = a$ and $X_T = b$ for some $a, b \in \mathbb{R}$.

Lemma 3.1 in [2]. Generate $X_{t_0}, X_{t_1}, \ldots, X_{t_n}, X_{t_{n+1}}$, where $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = T$, by $X_0 = a$ and

$$X_{t_i} = e^{-\theta(t_i - t_{i-1})} X_{t_{i-1}} + W_i, \qquad i = 1, \ldots, n+1,$$

where the W_is are independent and

$$W_i \sim \mathcal{N}\left(0, \sigma^2 \left(\frac{1 - e^{-2\theta(t_i - t_{i-1})}}{2\theta}\right)\right).$$

Define

$$Z_{t_i} = X_{t_i} + (b - X_{t_{n+1}}) \frac{e^{\theta t_i} - e^{-\theta t_i}}{e^{\theta t_{n+1}} - e^{-\theta t_{n+1}}}, \qquad i = 0, \ldots, n+1.$$

Then $(Z_{t_0}, Z_{t_1}, \ldots, Z_{t_n}, Z_{t_{n+1}})$ is distributed like an OU bridge with $Z_{t_0} = a$ and $Z_{t_{n+1}} = b$.

REFERENCES

[1] A. Beskos and G. Roberts *Exact simulation of diffusions*, Ann. Appl. Probab., 15 (2005), pp. 2422–2444.

[2] M. Bladt and M. Sorensen, *Simple Simulation of Diffusion Ridges with Application to Likelihood Inference for Diffusions*, Working paper, Centre for Analytical Finance, University of Copenhagen, Copenhagen, 2007.

[3] M. Branicki, B. Gershgorin, and A. J. Majda, *Filtering skill for turbulent signals for a suite of nonlinear and linear extended Kalman filters*, J. Comput. Phys., 231 (2012), pp. 1462–1498.

[4] M. Branicki, N. Chen, and A. J. Majda, *Non-Gaussian test models for prediction and state estimation with model errors*, Chinese Ann. Math. Ser. B, 34 (2013), pp. 29–64.

[5] M. Branicki and A. J. Majda, *Quantifying uncertainty for predictions with model error in non-Guassian systems with intermittency*, Nonlinearity, 25 (2012), pp. 2543–2578.

[6] M. Branicki and A. J. Majda, *Dynamic stochastic superresolution of sparseley observed turbulent systems*, J. Comput. Phys., 241 (2013), pp. 333–363.

[7] K. Cochrane, C. De Young, D. Soto, and T. Bahri, *Climate Change Implications for Fisheries and Aquaculture*, FAO Fisheries and Aquaculture technical report 530, Food and Agriculture Organization of the United Nations, Rome, 2009.

[8] S. Cotter, G. Roberts, A. Stuart, and D. White, *MCMC methods for functions: Modifying old algorithms to make them faster*, Statist. Sci., 28 (2013), pp. 424–446.

[9] S. Chib, M. K. Pitt, and N. Shephard, *Likelihood based inference for diffusion driven models*, Working paper, Nuffield College, University of Oxford, Oxford, 2004.

[10] S. Chib, M. K. Pitt, and N. Shephard, *Likelihood based inference for diffusion driven models*, Working paper, Nuffield College, University of Oxford, Oxford, 2010.

[11] O. Elerian, S. Chib, and N. Shephard, *Likelihood inference for discretely observed non-linear diffusions*, Econometrica, 69 (2001), pp. 959–993.

[12] B. Eraker, *Markov chain Monte Carlo analysis of diffusion models with application to finance*, J. Bus. Econom. Statist., 19 (2001), pp. 177–191.

[13] C. Fuchs, *Inference for Diffusion Processes with Applications in Life Sciences*, Springer, Heidelberg, 2013.

[14] B. Gershgorin, J. Harlim, and A. J. Majda, *Test models for improving filtering with model errors through stochastic parameter estimation*, J. Comput. Phys., 229 (2010), pp. 1–31.

[15] B. Gershgorin, J. Harlim, and A. J. Majda, *Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation*, J. Comput. Phys., 229 (2010), pp. 32–57.

[16] B. Gershgorin and A. J. Majda, *Filtering a statistically exactly solvable test model for turbulent tracers from partial observations*, J. Comput. Phys., 230 (2011), pp. 1602–1638.

[17] D. Giannakis, A. J. Majda, and I. Horenko, *Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems*, Phys. D, 241 (2012), pp. 1735–1752.

[18] A. Golightly and D. J. Wilkinson, *Bayesian inference for stochastic kinetic models using a diffusion approximation*, Biometrics, 61 (2005), pp. 781–788.

[19] A. Golightly and D. J. Wilkinson, *Bayesian sequential inference for nonlinear multivariate diffusions*, Stat. Comput., 16 (2006), pp. 323–338.

[20] A. Golightly and D. J. Wilkinson, *Bayesian sequential inference for stochastic kinetic biochemical network models*, J. Comput. Biol., 13 (2006), pp. 838–851.

[21] A. Golightly and D. J. Wilkinson, *Bayesian inference for nonlinear multivariate diffusion models observed with error*, Comput. Statist. Data Anal., 52 (2008), pp. 1674–1693.

[22] A. Golightly and D. J. Wilkinson, *Bayesian parameter inference for stochastic biochemical network models using particle MCMC*, Interface Focus, 1 (2011), pp. 807–820.

[23] J. Harlim and A. J. Majda, *Filtering turbulent sparsely observed geophysical flows*, Monthly Weather Rev., 138 (2010), pp. 1050–1083.

[24] S. E. Hills and A. F. M. Smith, *Parameterization Issues in Bayesian Inference*, in Bayesian Statistics 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., Oxford University Press, Oxford, 1992.

[25] K. Kalogeropoulos, *Likelihood-based inference for a class of multivariate diffusions with unobserved paths*, J. Statistical Planning and Inference, 137 (2007), pp. 3092–3102.
[26] A. J. Majda and M. Branicki, *Lessons in uncertainty quantification for turbulent dynamical system*, Discrete Contin. Dyn. Syst., 32 (2013), pp. 3133–3221.
[27] A. J. Majda, C. Franzke, and D. Crommelin, *Normal forms for reduced stochastic climate models*, Proc. Natl. Acad. Sci., 16 (2009), pp. 3649–3653.
[28] A. J. Majda and J. Harlim, *Filtering Complex Turbulent Systems*, Cambridge University Press, Cambridge, 2012.
[29] A. J. Majda and B. Gershgorin, *Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error*, Proc. Natl. Acad. Sci., 108 (2011), pp. 12599–12604.
[30] A. J. Majda, J. Harlim, and B. Gershgorin, *Mathematical strategies for filtering turbulent dynamical systems*, Discrete Contin. Dyn. Syst., 27 (2010), pp. 441–486.
[31] G. A. Meehl and C. Tebaldi, *More intense, more frequent, and longer lasting heat waves in the* 21*st century*, Science, 305 (2004), pp. 994–997.
[32] O. Papaspiliopoulos, G. O. Roberts, and O. Stramer, *Data augmentation for diffusions*, J. Comput. Graph. Statist., 22 (2013), pp. 665–688.
[33] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed., Springer, Heidelberg, 2003.
[34] G. Roberts and O. Stramer, *On inference for partial observations from partial observed nonlinear diffusion model using Metropolis–Hastings algorithms*, Biometrika 88 (2001), pp. 603–621.
[35] G. Sermaidis, O. Papaspiliopoulos, G. O. Roberts, A. Beskos, and P. Fearnhead, *Markov chain Monte Carlo for exact inference for diffusions*, Scand. J. Statist., 40 (2013), pp. 294–321.
[36] H. Sorensen, *Parametric inference for diffusion processes observed at discrete points in time: A survey*, Int. Statist. Rev., 72 (2004), pp. 337–354.
[37] O. Stramer and M. Bognar, *Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach*, Bayesian Anal., 6 (2011), pp. 231–258.
[38] N. N. Taleb, *The Fourth Quadrant: A Map of the Limits of Statistics*, Edge, 2008.
[39] M. A. Tanner and W. H. Wong, *The calculation of posterior distributions by data augmentation*, J. Amer. Statist. Assoc., 82 (1987), pp. 528–540.