Bayesian Inference of Stochastic Dynamical Models

by

Peter Lu

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2013

Author .....
Department of Mechanical Engineering
January 18, 2013

Certified by .....
Pierre F.J. Lermusiaux
Associate Professor of Mechanical Engineering
Thesis Supervisor

Accepted by ....
David E. Hardt
Chairman, Department Committee on Graduate Theses

Bayesian Inference of Stochastic Dynamical Models

by

Peter Lu

Submitted to the Department of Mechanical Engineering
on January 18, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

A new methodology for Bayesian inference of stochastic dynamical models is developed. The methodology leverages the dynamically orthogonal (DO) evolution equations for reduced-dimension uncertainty evolution and the Gaussian mixture model DO filtering algorithm for nonlinear reduced-dimension state variable inference to perform parallelized computation of marginal likelihoods for multiple candidate models, enabling efficient Bayesian update of model distributions. The methodology also employs reduced-dimension state augmentation to accommodate models featuring uncertain parameters. The methodology is applied successfully to two high-dimensional, nonlinear simulated fluid and ocean systems. Successful joint inference of an uncertain spatial geometry, one uncertain model parameter, and $\mathcal{O}(10^5)$ uncertain state variables is achieved for the first. Successful joint inference of an uncertain stochastic dynamical equation and $\mathcal{O}(10^5)$ uncertain state variables is achieved for the second. Extensions to adaptive modeling and adaptive sampling are discussed.

Thesis Supervisor: Pierre F.J. Lermusiaux
Title: Associate Professor of Mechanical Engineering

Acknowledgments

Thank you to my advisor and colleagues.

Thank you to my friends and family.

Thank you to MIT, for five marvelous years.

Contents

Chapter 1

Introduction

1.1 Stochastic Dynamical Systems

A stochastic dynamical system is any system that is both time-varying—dynamical—and affected by uncertainty—stochastic [59]. Examples of such systems are everywhere. Physical systems such as oceans and ecological networks, engineering systems such as power grids and communications networks, and anthropological systems such as financial markets and social networks can all be classified as stochastic dynamical systems. The mathematical tools that have been developed for investigating stochastic dynamical systems are thus highly versatile and have been applied in a wide range of fields [4, 5, 12, 25, 36, 51, 62, 66, 87, 116, 117, 118].

The general procedure for modeling a stochastic dynamical system is best illustrated with a simple example: a ball thrown through the air. This system is certainly dynamical; the ball's position and velocity, the state variables of the system, are changing with time. The system is also stochastic; the ball's trajectory is dependent on its initial conditions, which can be uncertain, as well as on turbulent multiscale aerodynamic forcings, which are often not deterministically predictable. The evolution of the ball's position and velocity can be mathematically described by a stochastic differential equation (SDE) that couples Newton's laws of motion with statistical representations of the turbulent aerodynamic forcings. This SDE represents the stochastic dynamical model for the system. Given a probability distribution for

the ball's initial conditions, the stochastic dynamical model can be used to predict probability distributions for the ball's position and velocity as functions of time, a process known as predicting the uncertainty evolution. If observations of the ball are made during the course of its flight, these probability distributions can be updated to account for the newly-acquired information, a process known as inference or data assimilation. Inference always reduces uncertainty in a system's state variables.

The general procedure for modeling a stochastic dynamical system can be summarized as follows:

1. *Model formulation.* Formulate a stochastic dynamical model that governs the evolution of the system's state variables and their associated probability distribution.

2. *Uncertainty initialization.* Quantify initial uncertainty in the system's state variables by specifying an initial probability distribution.

3. *Uncertainty evolution.* Evolve the initial state variable probability distribution through time using the stochastic dynamical model.

4. *Inference.* Update the state variable probability distribution using information from observations, reducing state variable uncertainty.

Though listed sequentially, uncertainty evolution and inference typically occur in an alternating fashion, with periods of uncertainty evolution interspersed with instances of inference. We also note that uncertainty evolution and inference can also be completed backward in time.

1.2 Model Formulation Uncertainty

For the example featuring the ballistic ball, it was implicitly assumed that the stochastic dynamical model formulated for the system was an accurate mathematical description of its governing physics. Uncertainty in the system's state variables was modeled

as originating solely from uncertainty in the system's initial conditions and the turbulent aerodynamic forcings, whose statistical properties were assumed known. This assumption of absolute validity of the model formulation however is not always defensible. For the ball, Newton's laws of motion were included in the system's stochastic dynamical model, but what if these laws were uncertain? That is, what if the ball were being thrown during a pre-Newtonian era when it was not known whether force was proportional to velocity, acceleration, or a non-linear function of both? This source of uncertainty—model formulation uncertainty—would certainly amplify the overall uncertainty in the ball's trajectory. In contemporary contexts, similar model uncertainty could arise when dealing with complex systems whose governing equations have not yet been derived from known first principles. For these systems, the assumption of absolute validity for any one particular model formulation would surely be inappropriate. In general, uncertainty in model formulation can originate from the choice of state variables themselves, from the functional forms of the model equations, the boundary conditions, and initial conditions, and from the definition of the (spatial) domain of integration. Both the deterministic and stochastic components of the model formulation can be uncertain. In what follows, when possible, we will we refer to model formulation uncertainty as simply model uncertainty.

Model uncertainty in stochastic dynamical systems can be difficult to properly quantify and is thus often ignored. This simplifying assumption is not severely damaging when model uncertainty is insignificant. When throwing a ball on Earth in the 21^{st} century for example, one can have high confidence in the validity of Newton's laws of motion. In other cases however, it can lead to significant underestimation of state variable uncertainty. [34], [55], and [72] review poignant examples from the statistics literature in which ignorance of model uncertainty resulted in overconfidence in state variable estimates, which subsequently led to tragically flawed conclusions.

Perhaps even more unfortunate, ignoring model uncertainty is antithetic to the scientific method, which entails the comparison of competing hypotheses by means of observations. If multiple models are considered, the same observations of a stochastic dynamical system that are used to perform inference of its state variables can also

be used to infer the relative validity of each of the models. This process of model inference can reveal valuable insights regarding the fundamental mechanisms that govern the system under investigation. If model uncertainty is ignored however and only one model—one hypothesis—is assumed, this opportunity for scientific discovery in the classic sense is forfeited.

1.3 Literature Review

Several methods have been developed to handle the coupled issues of model uncertainty and model inference in stochastic dynamical systems. We review a number of these methods here.

Directed search methods are a general class of computational methods for model inference. These methods typically proceed by first performing state variable inference for a large set of candidate models, then scoring the inferred state variables relative to system observations using metrics derived from frequentist statistics. Computational schemes are employed to search through expansive sets of plausible candidate models, with the search process directed by results from successive rounds of candidate evaluations. [123] is a premier example of this strategy. In this work, the authors employed a heuristic optimization scheme known as symbolic regression [75] to search through a space of algebraic expressions with the goal of finding the fundamental physical laws that govern several simple dynamical systems, such as single and double pendula. The authors were able to identify conservation laws for energy and momentum without any prior information regarding the laws' functional forms. Their approach was highly versatile but exceptionally demanding in terms of computational cost, even for the low-dimensional systems considered in their work. Model inference for the double pendulum system, a non-linear system with two state variables, required over 30 hours of computational time in a 32-core parallelized implementation. Extensions of their approach, and other directed search model inference methods (e.g. [19, 22, 74, 110, 144]), to high-dimensional systems will likely prove to be computationally challenging.

Hierarchical Bayesian modeling is a general approach to handling model uncertainty whereby full stochastic dynamical models are represented as hierarchies of simpler, analytically tractable sub-models [145, 147]. If these sub-models are properly formulated, inference can be performed separately for each by exploiting their conditional independences, with the sub-models aggregated afterwards to achieve global model inference. An oceanographic application of this approach is demonstrated in [146], where the authors used a hierarchical Bayesian model to formulate a stochastic dynamical model of the surface wind streamfunction over a region of the Labrador Sea using satellite surface wind velocity data. The aggregate wind model was decomposed into sub-models for observational data, boundary conditions, and the numerical streamfunction, which enabled the quantification of boundary condition uncertainties on the posterior distribution of streamfunction values. [60] features an ecological application of this approach, where the authors used a hierarchical Bayesian model to predict the spatial distribution of ground flora based on sparse data. Sub-models were formulated that enabled the incorporation of geographic covariates, a source of model uncertainty that, when accounted for, significantly enhanced flora distribution predictions. [16], [61], [103], [114], and [131] present further applications of hierarchical Bayesian modeling to problems of spatiotemporal statistics. Multiresolution Bayesian modeling, a variation of the approach for application to signal and image processing, is reviewed in [27], [28], and [148]. [65] reviews an extension of the hierarchical formulation to graphical models that allow for more complex interdependencies between model components at the cost of more computationally intensive inference algorithms.

Reduced-order modeling is a set of methods that have been employed to deal with high-dimensional systems featuring model uncertainty [40]. Though many model inference techniques, such as the directed search methods reviewed above, can be effective when system dimensions are small and candidate model spaces are readily explored, computational difficulties often arise when the same techniques are applied to high-dimensional systems, such as those frequently encountered in oceanography [89]. Reduced-order modeling techniques are designed to find low-dimensional repre-

sentations of high-dimensional models, for which model uncertainty quantification and inference are more readily performed. A wide selection of methods fall under this classification, including proper orthogonal decomposition [7, 20, 53], centroidal Voronoi tessellation [23], neural networks [44, 142], Volterra series [99], kriging [49], certified reduced basis methods based on numerical error bounds [64], empirical emulators [97, 134], error subspace statistical estimation (ESSE) [83, 90],and the dynamically orthogonal (DO) evolution equations [120, 121]. Though these techniques all take advantage of information redundancies in full-order stochastic dynamical models to achieve order reduction, some (e.g. the DO evolution equations) preserve decidedly greater physical meaning in their low-dimensional representations than others (e.g. neural networks and empirical emulators).

The methodology developed in the present work adopts ideas from both hierarchical Bayesian modeling and reduced-order modeling to achieve efficient model uncertainty quantification and inference for stochastic dynamical systems of large dimension. Other works directly relevant to the present methodology are reviewed later in Chapters 1 and 2.

1.4 Problem Statement

1.4.1 System

In this work, we consider a general stochastic dynamical system with state vector $\boldsymbol{X} \in \mathbb{R}^{N_{\boldsymbol{X}}}$ governed by an uncertain stochastic dynamical model $\mathcal{M}$ with uncertain parameter vector $\boldsymbol{\Theta} \in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$, where $N_{\boldsymbol{X}} \in \mathbb{N}$ and $N_{\boldsymbol{\Theta}} \in \mathbb{N}$ are the dimensions of the state and parameter vectors respectively. For realizations $\boldsymbol{x}$, $\boldsymbol{\theta}$, and $\mathcal{M}_n$ of $\boldsymbol{X}$, $\boldsymbol{\Theta}$, and $\mathcal{M}$ respectively, we have

$$\frac{\mathrm{d}\boldsymbol{x}(t;\omega)}{\mathrm{d}t} = \mathcal{M}_n\left[\boldsymbol{x}(t;\omega), \boldsymbol{\theta}(\omega), t; \omega\right] \tag{1.1}$$

$$\mathcal{M}_n \equiv \left(\mathcal{D}_n, \mathrm{SG}_n, \mathrm{BC}_n, \mathrm{IC}_n\right) , \tag{1.2}$$

where t denotes time, ω an index over stochastic realizations (a random event), $\mathcal{D}_n$ the set of stochastic dynamical equations for the model, SG_n the spatial geometry (spatial domain), BC_n the boundary conditions, and IC_n the initial conditions. All of the model components represented in (1.2) are allowed to be uncertain. The joint probability distribution over $\boldsymbol{X}$, Θ, and $\mathcal{M}$ is defined to be $p_{\boldsymbol{X},\Theta,\mathcal{M}}(\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n)$.

1.4.2 Observations

Stochastic observations $\boldsymbol{Y} \in \mathbb{R}^{N_{\boldsymbol{Y}}}$ of the system's state variables are assumed to be available at arbitrarily times, with the probability distribution of observations conditionally independent of both the stochastic dynamical model and the parameters given state variables

$$p_{\boldsymbol{Y}|\boldsymbol{X},\Theta,\mathcal{M}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n) = p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \equiv \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}} \ , \tag{1.3}$$

where $N_{\boldsymbol{Y}} \in \mathbb{N}$ is the dimension of the observation vector, $\boldsymbol{y}$ represents a realization of the observation vector, and $\mathcal{L}(\boldsymbol{y}|\bullet)$ represents the observation likelihood function or observation model for the system.

One particular observation likelihood function we will use is a Gaussian distribution, for which we have

$$\mathcal{L}(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{x}\,,\,\mathbf{R}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}} \ , \tag{1.4}$$

where $\mathcal{N}\big(\bullet\,;\,\boldsymbol{\mu}\,,\,\boldsymbol{\Sigma}\big)$ represents a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, $\mathbf{H} \in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{X}}}$ represents the linear observation matrix that transforms state variables into observation means, and $\mathbf{R} \in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{Y}}}$ represents the matrix of observation covariances. This particular observation model is equivalent to the following linear relation between state and observation vectors:

$$\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} + \boldsymbol{V} \ ,$$

where $\boldsymbol{V} \in \mathbb{R}^{N_{\boldsymbol{Y}}}$ represents zero-mean Gaussian noise with covariance $\mathbf{R}$. The obser-

vation model is assumed to be time-invariant (i.e. the parameters $\mathbf{H}$ and $\mathbf{R}$ in the particular case of (1.4) are assumed to be time-invariant).

1.4.3 Goal

The goal of the present work is two-part:

1. Efficiently evolve $p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}}(\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n)$ in time, accounting for all forms of uncertainty encapsulated in (1.1) and (1.2).

2. When observations are available, perform the update

$$p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}}(\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n) \rightarrow p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}|\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n \,|\, \boldsymbol{y})$$

   using the observation model represented by (1.3).

1.5 Candidate Models

When faced with a stochastic dynamical system whose model is uncertain, a common approach is to formulate a finite set of candidates for the true model governing the system under investigation. These candidate models or beliefs may be derived from first principles, may be inspired by previous observations, or may be based on a combination of theoretical and empirical prior knowledge. A discrete probability distribution $p_{\mathcal{M}}(\bullet)$ can be defined over the set of candidate models to represent the probabilities that each of the candidates are the true model. We note that in general the candidates can be correlated. In a Monte-Carlo approach, each candidate model can be used to predict the evolution of the uncertainties of the system's state variables independent of the other models, producing state variable probability distributions that are conditional on the candidates being the true model.

If the candidate models are assumed to be independent, a state variable probability distribution that accounts for the uncertainty in the formulation of the system's stochastic dynamical model can be estimated at any time as simply the weighted

average of these conditional distributions

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}} \,, \tag{1.5}$$

where $p_{\boldsymbol{X}}(\bullet)$ represents the state variable distribution, $N_{\mathcal{M}} \in \mathbb{N}$ represents the total number of candidate models, $\mathcal{M}_n$ represents the nth candidate model, and $p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$ represents the state variable distribution conditioned on the nth candidate model (the nth model-conditional state variable distribution). This general approach to accounting for model uncertainty has been used in many fields and is known by many names, including Bayesian model averaging [55, 112, 126], multimodel estimation [11, 101], multimodel fusion [98], and (multimodel or super-) ensemble modeling [50, 76, 115]. We note that if the candidate models $\mathcal{M}_n$ are correlated or if the space of model formulation/structures is continuous (instead of discrete as in (1.5)), the distribution (1.5) becomes a correlated weighting or an integral over the continuous model formulation/structures. Our formalism can be extended to these cases and this will be reported elsewhere.

In the present case, the linear nature of (1.5) w.r.t. candidate models leads to several useful properties. Marginal distributions for subsets of state variables can be found as weighted averages of the corresponding model-conditional marginal distributions. Letting $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be mutually exclusive complementary subsets of state variables and $\boldsymbol{x} = \left[\, \boldsymbol{x}_1 \; \boldsymbol{x}_2 \,\right]^{\mathrm{T}}$,

$$\begin{aligned}
p_{\boldsymbol{X}_1}(\boldsymbol{x}_1) &= \int p_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}_2 \\
&= \int p_{\boldsymbol{X}}\Big(\left[\, \boldsymbol{x}_1 \; \boldsymbol{x}_2 \,\right]^{\mathrm{T}}\Big)\, \mathrm{d}\boldsymbol{x}_2 \\
&= \int \left(\sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\mathcal{M}}\Big(\left[\, \boldsymbol{x}_1 \; \boldsymbol{x}_2 \,\right]^{\mathrm{T}}|\,\mathcal{M}_n\Big)\, p_{\mathcal{M}}(\mathcal{M}_n)\right) \mathrm{d}\boldsymbol{x}_2 \\
&= \sum_{n=1}^{N_{\mathcal{M}}} \left(\int p_{\boldsymbol{X}|\mathcal{M}}\Big(\left[\, \boldsymbol{x}_1 \; \boldsymbol{x}_2 \,\right]^{\mathrm{T}}|\,\mathcal{M}_n\Big)\, \mathrm{d}\boldsymbol{x}_2\right) p_{\mathcal{M}}(\mathcal{M}_n) \\
&= \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}_1|\mathcal{M}}(\boldsymbol{x}_1\,|\,\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n) \quad \forall \boldsymbol{x}_1 \in \mathbb{R}^{N_{\boldsymbol{X}_1}} \,,
\end{aligned} \tag{1.6}$$

where $N_{\boldsymbol{X}_1} < N_{\boldsymbol{X}}$. Similarly, the state variable mean can be found as a weighted average of the model-conditional state variable means

$$\begin{aligned}
\mathrm{E}[\boldsymbol{X}] &= \int \boldsymbol{x} p_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\
&= \int \boldsymbol{x} \left(\sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}|\mathcal{M}_n) \, p_{\mathcal{M}}(\mathcal{M}_n) \right) \mathrm{d}\boldsymbol{x} \\
&= \sum_{n=1}^{N_{\mathcal{M}}} \left(\int \boldsymbol{x} p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}|\mathcal{M}_n) \, \mathrm{d}\boldsymbol{x} \right) p_{\mathcal{M}}(\mathcal{M}_n) \\
&= \sum_{n=1}^{N_{\mathcal{M}}} \mathrm{E}[\boldsymbol{X}|\mathcal{M}_n] \, p_{\mathcal{M}}(\mathcal{M}_n) \ . \qquad (1.7)
\end{aligned}$$

1.6 Bayesian Model Inference

When observations of a stochastic dynamical system's state variables are made, both the model-conditional state variable distributions and model distribution within the summation of (1.5) can theoretically be updated using Bayes' theorem [13]

$$p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x}|\boldsymbol{y},\mathcal{M}_n) = \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y}|\boldsymbol{x},\mathcal{M}_n)}{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)} \, p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}|\mathcal{M}_n) \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}, \forall n \in \{1, \ldots, N_{\mathcal{M}}\} \ , \qquad (1.8)$$

$$p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n|\boldsymbol{y}) = \frac{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \, p_{\mathcal{M}}(\mathcal{M}_n) \quad \forall n \in \{1, \ldots, N_{\mathcal{M}}\} \ . \qquad (1.9)$$

In (1.8), the model $\mathcal{M}_n$ plays the role of a 'given parameter'. For this $\mathcal{M}_n$, the distributions $p_{\boldsymbol{X}|\mathcal{M}}(\bullet|\mathcal{M}_n)$ and $p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\bullet|\boldsymbol{y},\mathcal{M}_n)$ are referred to as the prior and posterior conditional state variable distributions for the nth candidate model respectively, while $p_{\mathcal{M}}(\bullet)$ and $p_{\mathcal{M}|\boldsymbol{Y}}(\bullet|\boldsymbol{y})$ are referred to as the prior and posterior model distributions.

If the candidate models are assumed independent, as in (1.5), the posterior model-conditional state variable distributions and model distribution can be combined to

form the posterior state variable distribution

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x}\,|\,\boldsymbol{y},\mathcal{M}_n) p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n\,|\,\boldsymbol{y}) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}} \ . \tag{1.10}$$

(1.8) represents Bayesian state variable inference and can be performed for each model-conditional state variable distribution independently, ignoring model uncertainty. Techniques for state variable inference abound, ranging from classic analytical methods such as the Kalman filter [69, 70] to contemporary computational approaches such as particle filters [6], Markov chain Monte Carlo (MCMC) algorithms [3], and forward-backward algorithms [37]. State variable inference and data assimilation have roots in optimal estimation theory e.g. [45, 68, 124] and control theory [95, 82], with now many applications in environmental sciences and engineering (e.g. [15, 71, 79, 100, 111, 117, 145, 149]).

(1.9) represents Bayesian model inference. The comparison of the posteriors for each $\mathcal{M}_n$, $p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n\,|\,\boldsymbol{y})$, is Bayesian hypothesis testing for competing models and each $p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n\,|\,\boldsymbol{y})$ is often referred to as model evidence. Though cosmetically simpler than (1.8), (1.9) is in fact the more challenging of the two Bayesian updates to perform. The chief difficulty lies in the calculation of the marginal likelihood $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)$, which represents the strength of the observational evidence for the nth candidate model, i.e. the likelihood of model $\mathcal{M}_n$, for all states $\boldsymbol{X}$. While $p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y}\,|\,\boldsymbol{x},\mathcal{M}_n)$ is equivalent to the observation likelihood function $\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})$—the function that defines the probability distribution for observations when state variables are known—an explicit expression for the likelihood $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)$ is not available. Instead, $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)$ (the probability distribution for the observation vector realization conditioned on a given candidate model) must be found through oftentimes difficult (large-dimension) integrations [34, 145, 138]

$$\begin{aligned} p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) &= \int p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y}\,|\,\boldsymbol{x},\mathcal{M}_n)\, p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\, \mathrm{d}\boldsymbol{x} \\ &= \int \mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})\, p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\, \mathrm{d}\boldsymbol{x} \quad \forall n \in \{1,\ldots,N_{\mathcal{M}}\} \ . \end{aligned} \tag{1.11}$$

Note that although $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)$ also appears in (1.8), there it serves only as the normalization constant for the posterior conditional state variable distribution and its explicit calculation is usually side-stepped by Bayesian state variable inference schemes. Likewise, $p_{\boldsymbol{Y}}(\boldsymbol{y})$ appearing in (1.9) is not of concern for inference, as it serves only as the normalization constant for the posterior model distribution. Once all marginal likelihoods have been found, $p_{\boldsymbol{Y}}(\boldsymbol{y})$ can be computed as simply their weighted summation

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n) \; . \tag{1.12}$$

It is not uncommon that the difficulty of computing the integral in (1.11) leads to the avoidance of the update of the model distribution (1.9) entirely [76, 112]. In these cases, prior statistical knowledge derived from a fixed set of system observations is typically used to specify an initial model distribution, which is then kept unchanged even when new observations become available. This approach is suboptimal, as update of the model distribution yields two significant benefits: 1) More precise weighting of the conditional state variable distributions in (1.5), leading to improved state variable uncertainty quantification (1.10); and 2) Insight into the true model governing the system under investigation. Techniques for the efficient computation of the marginal likelihood integral in (1.11) are thus of great utility for the study of stochastic dynamical systems featuring uncertainty in model formulation.

1.7 Overview of the Present Work

The present work develops a new methodology for performing Bayesian inference of stochastic dynamical models. The dynamically orthogonal (DO) evolution equations for reduced-dimension uncertainty evolution [121] and the Gaussian mixture model DO filtering algorithm (GMM-DO filter) for nonlinear reduced-dimension inference of state variables [129, 130] are first extended to accommodate stochastic dynamical models featuring uncertain parameters. Another extension then enables the compu-

tationally expedient calculation of the marginal likelihood integral in (1.11). The result is a methodology capable of performing Bayesian model inference for high-dimensional, nonlinear stochastic dynamical systems, including oceanic and atmospheric systems.

Chapter 2 reviews a number of techniques that have been developed to calculate the marginal likelihood integral in (1.11), with particular attention paid to the added difficulties that arise with high-dimensional systems. Chapter 3 reviews the DO evolution equations and the GMM-DO filter, then develops the extensions of the new methodology in detail. Chapter 4 applies the methodology to a simulated stochastic dynamical system featuring a fluid flowing past an obstacle; model uncertainty arises in the shape of the obstacle. Chapter 5 applies the methodology to a second simulated system featuring a marine microorganism convected by a fluid; model uncertainty arises in the reaction equation of the microorganism. Chapter 6 provides a synopsis and a discussion of promising avenues for future investigation.

Chapter 2

Marginal Likelihood Calculation

2.1 Overview

As introduced in Chapter 1, the key computational difficulty associated with Bayesian model inference for a set of model candidates is the calculation of the marginal likelihood integral

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) = \int \mathcal{L}(\boldsymbol{y}|\boldsymbol{x})\, p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}|\mathcal{M}_n)\, \mathrm{d}\boldsymbol{x} \,. \tag{2.1}$$

Three classes of methods for performing this crucial calculation are reviewed in this chapter in order of increasing computational expense. The applicability of each of these method classes is dependent on the functional forms of the observation likelihood function $\mathcal{L}(\boldsymbol{y}|\bullet)$ and model-conditional state variable distribution $p_{\boldsymbol{X}|\mathcal{M}}(\bullet|\mathcal{M}_n)$. Note that (2.1) need not be calculated using the same method for all candidate models. Indeed, the functional form of $p_{\boldsymbol{X}|\mathcal{M}}(\bullet|\mathcal{M}_n)$ for some candidates may enable computational expediencies not applicable to others.

2.2 Analytical Cases

In a handful of special cases, the functional forms of the observation likelihood function $\mathcal{L}(\boldsymbol{y}|\bullet)$ and conditional state variable distribution $p_{\boldsymbol{X}|\mathcal{M}}(\bullet|\mathcal{M}_n)$ allow for the

analytical calculation of the integral in (2.1).

One such case is when both $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)$ and $p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$ are Gaussian distributions. Let

$$\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x}) = \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{x}\,,\,\mathbf{R}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\ ,$$

as first defined in (1.4). Further, let

$$p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n) = \mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n}\,,\,\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\ ,$$

where $\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n} \in \mathbb{R}^{N_{\boldsymbol{X}}}$ represents the vector of state variable means conditional on the nth candidate model and $\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n} \in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}}$ represents the matrix of state variable covariances conditional on the nth candidate model.

Then, (2.1) becomes

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) = \int \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{x}\,,\,\mathbf{R}\big)\,\mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n}\,,\,\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n}\big)\,\mathrm{d}\boldsymbol{x}\ . \tag{2.2}$$

Since the integral in (2.2) is taken over all values of $\boldsymbol{x}$, a linear transformation of the integration variable can be performed without changing the value of the integral. Then, using the linear transformation properties of Gaussian distributions, (2.2) can be rewritten as

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) = \int \mathcal{N}\big(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\,;\,0\,,\,\mathbf{R}\big)\,\mathcal{N}\big(\mathbf{H}\boldsymbol{x}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n}\mathbf{H}^{\mathrm{T}}\big)\,\mathrm{d}\mathbf{H}\boldsymbol{x}\ . \tag{2.3}$$

Observations are assumed to be unbiased. (2.3) represents the convolution of two Gaussian distributions and thus yields another Gaussian distribution whose mean and variance are equal to the sums of the means and variances of the two component distributions respectively [35]

$$\begin{aligned} p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) &= \Big[\mathcal{N}\big(\bullet\,;\,0\,,\,\mathbf{R}\big) * \mathcal{N}\big(\bullet\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n}\mathbf{H}^{\mathrm{T}}\big)\Big](\boldsymbol{y}) \\ &= \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big)\ , \end{aligned} \tag{2.4}$$

where $*$ represents the convolution operator, defined as

$$\big[f(\bullet) * g(\bullet)\big](t) = \int f(t-\tau)g(\tau)\,\mathrm{d}\tau \ .$$

The full derivation of the convolution identity used in (2.4) is provided in Appendix A.

Even if the conditional state variable distribution $p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$ is not Gaussian, (2.4) can still be used to analytically calculate an approximation of the marginal likelihood if the mean and covariance of the conditional distribution is used to form a Gaussian approximation of the distribution. In control theory, state estimation schemes that are based on the classic Kalman filter typically already employ such Gaussian approximations and the extension of these schemes to perform Bayesian model inference using (2.4) has thus been natural [11, 101]. Though these Gaussian approximation methods are computationally expedient, they are limited in the complexity of the systems they can accomodate. Most conspicuously, systems that feature substantially non-Gaussian state variable distributions are handled poorly. A remedy for this—Gaussian mixture model (GMM) approximations—will be explored in Chapter 3. High-dimensional systems also pose a challenge as the size of the conditional state variable covariance matrix appearing in (2.4) grows as ${N_{\boldsymbol{X}}}^2$, the square of the number of state variables. More versatile techniques are reviewed in the following sections.

2.3 Asymptotic Approximations

When the functional forms of $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)$ and $p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$ do not allow for the analytical calculation of the integral in (2.1), a popular alternative is to use closed-form approximations of the integral that are exact in the limit of infinite observations [72]. The majority of these asymptotic approximations are variations of the Laplace approximation. The general Laplace approximation for an integral of the form $\int \mathrm{e}^{f(\boldsymbol{u})}\,\mathrm{d}\boldsymbol{u}$ is

$$\int \mathrm{e}^{f(\boldsymbol{u})}\,\mathrm{d}\boldsymbol{u} \approx (2\pi)^{d/2}\,|\mathbf{A}^*|^{1/2}\mathrm{e}^{f(\boldsymbol{u}^*)} \ , \tag{2.5}$$

where $\boldsymbol{u}$ is a vector, d is the dimension of $\boldsymbol{u}$, $\boldsymbol{u}^*$ is the maximizing argument of $f(\bullet)$, and $\mathbf{A}^*$ is the negative of the inverse Hessian of $f(\bullet)$ evaluated at $\boldsymbol{u}^*$ [9]. Substituting $\log\left[\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})\,p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\right]$ for $f(\boldsymbol{u})$ yields the Laplace approximation for (2.1)

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) \approx (2\pi)^{N_{\boldsymbol{X}}/2}\,|\mathbf{A}^*|^{1/2}\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x}^*)\,p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}^*\,|\,\mathcal{M}_n)\ , \tag{2.6}$$

where $\boldsymbol{x}^*$ is the maximizing argument of $\log\left[\mathcal{L}(\boldsymbol{y}\,|\,\bullet)\,p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)\right]$, which is equivalently the maximizing argument of $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)\,p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$, and $\mathbf{A}^*$ is the negative of the inverse Hessian of $\log\left[\mathcal{L}(\boldsymbol{y}\,|\,\bullet)\,p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)\right]$ evaluated at $\boldsymbol{x}^*$. Depending on the functional forms of $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)$ and $p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$, it may not be possible to find $\boldsymbol{x}^*$ and $\mathbf{A}^*$ analytically, in which case sample approximations for $\boldsymbol{x}^*$ and $\mathbf{A}^*$ must be found before (2.6) can be evaluated, e.g. [92].

The accuracy of the Laplace approximation and its variants increases as the density of $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)\,p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$ increases near its maximum (i.e. the more the density peaks, the better). These asymptotic approximations are thus best-suited for systems featuring unimodal state variable distributions and large numbers of observations. Unfortunately, in high-dimensional, nonlinear systems such as those encountered in oceanography and meteorology, multimodality and sparse observations are the norm (e.g. [8, 31, 88, 87, 30, 100]). Furthermore, calculating and inverting the Hessian of $\log\left[\mathcal{L}(\boldsymbol{y}\,|\,\bullet)\,p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)\right]$, which is necessary for (2.6), is computationally infeasible for systems of high-dimension and hence also needs to be approximated.

2.4 Computational Approximations

For cases where analytical solutions to (2.1) are not available and asymptotic approximations are inappropriate, computational approximations are the only recourse. A general class of computational techniques known as importance sampling [33, 132] makes use of samples drawn from a chosen sampling distribution $g(\bullet)$ to approximate

(2.1) as the weighted average of numerous likelihood function evaluations

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) \approx \frac{\sum_{k=1}^{K} w_k\, \mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x}_k)}{\sum_{k=1}^{K} w_k} \;, \tag{2.7}$$

where $\boldsymbol{x}_k$ represents the kth of $K \in \mathbb{N}$ samples and the weights are defined by

$$w_k = \frac{p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}_k\,|\,\mathcal{M}_n)}{g\,(\boldsymbol{x}_k)} \quad \forall k \in \{1, \ldots, K\} \;. \tag{2.8}$$

A sample weight is large when the density of the prior conditional state variable distribution at the sample value is high relative to the density of the sampling distribution $g(\bullet)$. The sample weights thus balance the frequency with which sample values are drawn from the sampling distribution and their 'importance' relative to the prior conditional state variable distribution.

The simplest choice for $g(\bullet)$ is the prior conditional state variable distribution itself, which reduces all sample weights to one and (2.7) to

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) \approx \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}\big(\boldsymbol{y}\,|\,\boldsymbol{x}_k^{-}\big) \;, \tag{2.9}$$

where $\boldsymbol{x}_k^{-}$ represents the kth sample from the prior distribution. This is known as the arithmetic mean estimator (AME) [81]. Though the simplicity of its implementation is attractive, the AME can exhibit slow rates of convergence if $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)$ is large for only a small subset of its domain.

If $g(\bullet)$ is instead chosen to be the posterior conditional state variable distribution,

(2.7) becomes

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) \approx \frac{\sum_{k=1}^{K} \frac{p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}_k^+|\mathcal{M}_n)}{p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x}_k^+|\boldsymbol{y},\mathcal{M}_n)} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)}{\sum_{k=1}^{K} \frac{p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}_k^+|\mathcal{M}_n)}{p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x}_k^+|\boldsymbol{y},\mathcal{M}_n)}}$$

$$= \frac{\sum_{k=1}^{K} \frac{p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}_k^+|\mathcal{M}_n)}{p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x}_k^+|\boldsymbol{y},\mathcal{M}_n)\, p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)}{\sum_{k=1}^{K} \frac{p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}_k^+|\mathcal{M}_n)}{p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x}_k^+|\boldsymbol{y},\mathcal{M}_n)\, p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)}},$$

where $\boldsymbol{x}_k^+$ represents the kth sample from the posterior distribution. Using (1.8),

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) \approx \frac{\sum_{k=1}^{K} \frac{1}{p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y}|\boldsymbol{x}_k^+,\mathcal{M}_n)} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)}{\sum_{k=1}^{K} \frac{1}{p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y}|\boldsymbol{x}_k^+,\mathcal{M}_n)}}$$

$$= \frac{\sum_{k=1}^{K} \frac{1}{\mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)}{\sum_{k=1}^{K} \frac{1}{\mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)}}$$

$$= \left(\frac{1}{K} \sum_{k=1}^{K} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)^{-1} \right)^{-1}. \tag{2.10}$$

This is known as the harmonic mean estimator (HME) [108]. The convergence rate of the HME is typically faster than that of the AME due to the fact that the posterior conditional state variable distribution tends to have greater density than the prior distribution in areas where $\mathcal{L}(\boldsymbol{y}|\bullet)$ is large. This is a direct result of Bayesian state variable inference; the posterior distribution is generated by shifting the prior

distribution towards state variable values that are more likely to have produced the observations—i.e. values for which $\mathcal{L}(\boldsymbol{y} \mid \bullet)$ is large.

More advanced sampling techniques generally employ recursive schemes that either bias or constrain the sample distribution $g(\bullet)$ to regions where $\mathcal{L}(\boldsymbol{y} \mid \bullet)$ is large. These techniques include bridge sampling [102], path sampling [46], annealed importance sampling [107], and nested sampling [29, 125]. Though the use of recursive adjustments to $g(\bullet)$ can greatly improve both convergence rate and stability, these advanced sampling techniques are still limited in the dimension of the systems they can handle. Even the best contemporary sampling techniques are limited to problems of dimension $\mathcal{O}(10^2)$ [46], which falls far short of the $\mathcal{O}(10^6$–$10^9)$ systems frequently encountered in the geophysical sciences [89]. At a fundamental level, sampling techniques are stymied by the fact that state space volumes grow as exponential functions of the number of the state variables. The number of samples needed to thoroughly canvass a high-dimensional state space in order to accurately approximate the marginal likelihood integral in (2.1) is thus prohibitively large.

2.5 Summary

High system dimension is the common stumbling point of the marginal likelihood calculation methods reviewed in this chapter. In order to perform Bayesian model inference for high-dimensional stochastic dynamical systems, the major challenge is to derive inference schemes that can reduce the dimension of the systems down to a manageable order while capturing and exploiting dominant nonlinear dynamics and non-Gaussian statistics as they arise. Chapter 3 will show that the DO evolution equations are an effective means of achieving this dimension reduction. Furthermore, Chapter 3 will also show that the GMM approximations employed by the GMM-DO filter allow for the analytical calculation of the marginal likelihood integral in (2.1), while still providing the structural flexibility necessary to accommodate systems featuring nonlinear, multimodal state variable distributions. Together, the DO evolution equations and the GMM-DO filter enable Bayesian model inference for

high-dimensional, nonlinear stochastic dynamical systems.

Chapter 3

Methodology

3.1 Overview

To achieve the efficient dimension-reduction necessary for performing the marginal likelihood calculations (1.11) for high-dimensional stochastic dynamical systems, the present methodology employs the DO evolution equations to evolve the dominant model-conditional state variable distributions in (1.5). The GMM-DO filter is used to perform the Bayesian updates of the model-conditional distributions represented by (1.8) in the evolving DO subspace. Reduced-dimension state augmentation is used to extend the DO evolution equations and the GMM-DO filter to accommodate stochastic dynamical models featuring uncertain parameters. The marginal likelihood calculations (1.11) are then performed analytically using Gaussian mixture models (GMM) in the DO subspaces for each candidate model, extending the Gaussian approximation reviewed in Section 2.2. Finally, these marginal likelihoods are used to perform the Bayesian update (1.9) of the model distribution, thus accomplishing Bayesian model inference. Each of these components of the present methodology is developed in detail in the following sections, with an integrated account of the methodology provided at the end of the chapter. A compendium of the notation used in this chapter is provided in Table 3.1 for ease of reference.

Table 3.1: Notation compendium.

General

$N_{\boldsymbol{X}}$	$\in \mathbb{N}$	dimension of state vector
$\boldsymbol{X}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	state vector
$\boldsymbol{x}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	state vector realization
$N_{\boldsymbol{\Theta}}$	$\in \mathbb{N}$	dimension of parameter vector
$\boldsymbol{\Theta}$	$\in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$	parameter vector
$\boldsymbol{\theta}$	$\in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$	parameter vector realization
$\mathcal{M}$		stochastic dynamical model
$N_{\mathcal{M}}$	$\in \mathbb{N}$	number of candidate models
n	$\in \{1, \ldots, N_{\mathcal{M}}\}$	candidate model index
$\mathcal{M}_n$		nth candidate model
$\mathcal{D}_n$		stochastic dynamical equations of nth candidate model
SG_n		spatial geometry of nth candidate model
BC_n		boundary conditions of nth candidate model
IC_n		initial conditions of nth candidate model
$N_{\boldsymbol{Y}}$	$\in \mathbb{N}$	dimension of observation vector
$\boldsymbol{Y}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}}$	observation vector
$\boldsymbol{y}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}}$	observation vector realization

DO Evolution Equations

$\overline{\boldsymbol{x}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	state vector mean
N_{DO}	$\in \mathbb{N}$	dimension of stochastic subspace
i	$\in \{1, \ldots, N_{\mathrm{DO}}\}$	mode index
$\boldsymbol{\mathcal{X}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\mathrm{DO}}}$	matrix of modes
$\tilde{\boldsymbol{x}}_i$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	ith mode
$\boldsymbol{\Phi}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	mode coefficient vector
$\boldsymbol{\phi}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	mode coefficient vector realization
ϕ_i	$\in \mathbb{R}$	ith mode coefficient
N_{MC}	$\in \mathbb{N}$	number of Monte Carlo samples
k	$\in \{1, \ldots, N_{\mathrm{MC}}\}$	Monte Carlo sample index
$\boldsymbol{x}_k$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	kth state vector sample
$\boldsymbol{\theta}_k$	$\in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$	kth mode coefficient vector sample

GMM-DO Filter

$\mathbf{H}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{X}}}$	linear observation matrix
$\mathbf{R}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{Y}}}$	observation covariance matrix
N_{GMM}	$\in \mathbb{N}$	number of GMM components
j	$\in \{1, \ldots, N_{\text{GMM}}\}$	GMM component index
$\pi_{\boldsymbol{X},j}$	$\in \mathbb{R}$	jth component weight of prior state GMM
$\boldsymbol{\mu}_{\boldsymbol{X},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	jth component mean vector of prior state GMM
$\boldsymbol{\Sigma}_{\boldsymbol{X},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}}$	jth component covariance matrix of prior state GMM
$\pi_{\boldsymbol{\Phi},j}$	$\in \mathbb{R}$	jth component weight of prior mode coefficient GMM
$\boldsymbol{\mu}_{\boldsymbol{\Phi},j}$	$\in \mathbb{R}^{N_{\text{DO}}}$	jth component mean vector of prior mode coefficient GMM
$\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}$	$\in \mathbb{R}^{N_{\text{DO}} \times N_{\text{DO}}}$	jth component covariance matrix of prior mode coefficient GMM
$\pi_{\boldsymbol{X}\|\boldsymbol{Y},j}$	$\in \mathbb{R}$	jth component weight of posterior state GMM
$\boldsymbol{\mu}_{\boldsymbol{X}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	jth component mean vector of posterior state GMM
$\boldsymbol{\Sigma}_{\boldsymbol{X}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}}$	jth component covariance matrix of posterior state GMM
$\mathbf{K}_j$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{Y}}}$	jth gain matrix
$\pi_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}$	jth component weight of posterior mode coefficient GMM
$\boldsymbol{\mu}_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\text{DO}}}$	jth component mean vector of posterior mode coefficient GMM
$\boldsymbol{\Sigma}_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\text{DO}} \times N_{\text{DO}}}$	jth component covariance matrix of posterior mode coefficient GMM
$\tilde{\boldsymbol{y}}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}}$	transformed observation vector realization
$\tilde{\mathbf{H}}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\text{DO}}}$	transformed observation matrix
$\tilde{\mathbf{K}}_j$	$\in \mathbb{R}^{N_{\text{DO}} \times N_{\boldsymbol{Y}}}$	jth transformed gain matrix
$\boldsymbol{\mu}'_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\text{DO}}}$	jth intermediate component mean vector

Reduced-Dimension State Augmentation

$\boldsymbol{X}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	augmented state vector
$\boldsymbol{x}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	augmented state vector realization
$\mathcal{M}_{\boldsymbol{\Theta}}$		augmented stochastic dynamical model
$\overline{\boldsymbol{x}}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	augmented state vector mean
$\tilde{\boldsymbol{x}}_{\boldsymbol{\Theta},i}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	ith augmented mode
$\mathbf{H}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}\times(N_{\boldsymbol{X}}+N_{\mathrm{DO}})}$	augmented linear observation matrix

Operators, Functions, and Indicators

$\mathrm{E}[\bullet]$	expectation operator
$*$	convolution operator
$\mathcal{L}(\boldsymbol{y}\,\vert\,\bullet)$	observation likelihood function
$\mathcal{N}(\bullet\,;\,\boldsymbol{\mu}\,,\,\boldsymbol{\Sigma})$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
$(\cdot)^{-}$	prior
$(\cdot)^{+}$	posterior

3.2 DO Evolution Equations

The DO evolution equations are a closed set of reduced-dimension stochastic differential equations that effectively approximate general, high-dimensional nonlinear stochastic differential equations. These equations are premised on the fact that any vector of stochastic dynamical state variables can be approximated to arbitrary accuracy using a DO expansion (a generalized, time-dependent Karhunen-Loeve decomposition) of the state vector

$$\boldsymbol{x}(t;\omega) \approx \overline{\boldsymbol{x}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\,\tilde{\boldsymbol{x}}_i(t)\ , \tag{3.1}$$

where $\overline{\boldsymbol{x}}(t) \in \mathbb{R}^{N_{\boldsymbol{X}}}$ denotes the state vector mean, $\tilde{\boldsymbol{x}}_i(t) \in \mathbb{R}^{N_{\boldsymbol{X}}}$ the ith of N_{DO} orthonormal basis vectors, and $\phi_i(t;\omega) \in \mathbb{R}$ the ith of N_{DO} zero-mean stochastic processes [96]. The basis vectors and stochastic processes are referred to as the modes and mode coefficients of the expansion respectively. The expansion is exact in the

case where N_{DO} equals the dimension of the state vector $N_{\boldsymbol{X}}$.

If the modes in (3.1) are properly selected, the total number of modes needed to achieve high approximation accuracy can be orders of magnitude less than the state vector dimension. Specifically, if the modes are chosen to be oriented in the directions of 'largest' state uncertainty, then a small number of modes can capture a large majority of the total uncertainty in the state vector. These modes then define a stochastic subspace $\mathbf{V}_{\mathrm{DO}} = \mathrm{span}\{\tilde{\boldsymbol{x}}_i(t)\}_{i=1}^{N_{\mathrm{DO}}}$ embedded in $\mathbb{R}^{N_{\boldsymbol{X}}}$ within which the majority of the state uncertainty resides. At any given time, a reduced-dimension probability distribution for the N_{DO} mode coefficients then efficiently represents the full probability distribution for the N_{X} state variables, as the expansion (3.1) relates the two sets of variables through an affine transformation.

To evolve the probability density of the state vector, equations for the terms in the expansion (3.1) are obtained from the original stochastic dynamical model equations governing the evolution of the state vector

$$\frac{\mathrm{d}\boldsymbol{x}(t;\omega)}{\mathrm{d}t} = \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] \ . \tag{3.2}$$

We assume for now that the true model for the system is known and hence use $\mathcal{M}$ in (3.2) as opposed to the $\mathcal{M}_n$ used in (1.1). Specifically, evolution equations for $\overline{\boldsymbol{x}}(t)$, $\tilde{\boldsymbol{x}}_i(t)$, and $\phi_i(t;\omega)$ [121] are obtained by insertion of (3.1) into (3.2), noting that while only the dynamical evolution of the modes can capture state uncertainty evolution orthogonal to the stochastic subspace $\mathbf{V}_{\mathrm{DO}}$ at time t, both dynamical evolution of the modes and of the mode coefficients can capture uncertainty evolution within $\mathbf{V}_{\mathrm{DO}}$. This redundancy is eliminated by constraining the evolution of the modes to directions orthogonal to $\mathbf{V}_{\mathrm{DO}}$

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \perp \mathbf{V}_{\mathrm{DO}} \Leftrightarrow \left\langle \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \, , \, \tilde{\boldsymbol{x}}_j(t) \right\rangle = 0 \quad \forall i,j \in \{1,\ldots,N_{\mathrm{DO}}\} \ , \tag{3.3}$$

where the operator $\langle \mathbf{a}, \mathbf{b} \rangle$ represents the vector inner product of $\mathbf{a}$ and $\mathbf{b}$. (3.3) is known as the DO condition for mode evolution [121]. The DO condition, critically,

preserves the orthonormality of the modes

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \tilde{\boldsymbol{x}}_i(t)\,,\,\tilde{\boldsymbol{x}}_j(t)\right\rangle = \left\langle \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\,,\,\tilde{\boldsymbol{x}}_j(t)\right\rangle + \left\langle \tilde{\boldsymbol{x}}_i(t)\,,\,\frac{\mathrm{d}\tilde{\boldsymbol{x}}_j(t)}{\mathrm{d}t}\right\rangle = 0$$
$$\forall i,j \in \{1,\ldots,N_{\mathrm{DO}}\}\ .$$

Using (3.3) in conjunction with the expansion (3.1) and stochastic dynamical model (3.2), a unique set of evolution equations can be derived for the state vector mean, modes, and mode coefficients. These are the DO evolution equations

$$\frac{\mathrm{d}\overline{\boldsymbol{x}}(t)}{\mathrm{d}t} = \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\ , \tag{3.4}$$

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N_{\mathrm{DO}}} \mathbf{C}^{-1}_{(i,j)}\,\mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right]$$
$$\forall i \in \{1,\ldots,N_{\mathrm{DO}}\}\ , \tag{3.5}$$

$$\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t} = \left\langle \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] - \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\,,\,\tilde{\boldsymbol{x}}_i(t)\right\rangle$$
$$\forall i \in \{1,\ldots,N_{\mathrm{DO}}\}\ , \tag{3.6}$$

where $\mathrm{E}\left[\bullet\right]$ represents the expectation operator, $\mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathbf{a}\right]$ represents the projection of the vector $\mathbf{a}$ onto the space orthogonal to $\mathbf{V}_{\mathrm{DO}}$

$$\mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathbf{a}\right] = \mathbf{a} - \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}}\left[\mathbf{a}\right] = \mathbf{a} - \sum_{k=1}^{N_{\mathrm{DO}}}\left\langle \mathbf{a}\,,\,\tilde{\boldsymbol{x}}_k(t)\right\rangle\tilde{\boldsymbol{x}}_k(t)$$

and $\mathbf{C}^{-1}_{(i,j)}$ represents the (i,j)th entry of the inverse of the mode coefficient covariance matrix

$$\mathbf{C}_{(i,j)} = \mathrm{E}\left[\phi_i(t;\omega)\phi_j(t;\omega)\right]\ .$$

The complete derivation of (3.4)–(3.6) from (3.1)–(3.3) is provided in Appendix B. The imposition of additional constraints on (3.4)–(3.6) can be shown to result in either the proper orthogonal decomposition (POD) evolution equations [109] or the

polynomial chaos (PC) evolution equations [47], indicating that these two more conventional methods for reduced-dimension uncertainty evolution are subclasses of the DO equations [121].

For the present state vector formulation (3.2), (3.4) and (3.5) represent $(N_{\mathrm{DO}}+1)$ ordinary differential equations (ODEs) of dimension $N_{\boldsymbol{X}}$, while (3.6) represents N_{DO} SDEs of unit dimension or, equivalently, a single SDE of dimension N_{DO}. These equations are coupled. Numerical implementations of the DO evolution equations can employ classic solvers for the deterministic evolution of the state vector mean and modes (of course specific original system equations can lead to powerful DO solvers, see [141]). Meanwhile, stochastic evolution of the coefficients can be carried out using Monte Carlo (MC) sampling methods, whereby $N_{\mathrm{MC}} \gg N_{\mathrm{DO}}$ samples are drawn from the initial mode coefficient distribution and evolved by solving (3.6) as an ODE of dimension N_{DO} for each sample [141]. These evolved samples then constitute a sample approximation for the mode coefficient distribution at any point in time, keeping a rich description of this distribution since $N_{\mathrm{MC}} \gg N_{\mathrm{DO}}$. However, an equivalent sampling approach applied to the original SDE in (3.2) would require the solution of N_{MC} ODEs of dimension $N_{\boldsymbol{X}} \gg N_{\mathrm{DO}}$, an endeavor of substantially greater computational expense. The DO evolution equations thus enable computationally expedient reduced-dimension uncertainty evolution for general, nonlinear stochastic dynamical systems.

In [121], the DO equations were derived for infinite-dimensional stochastic dynamical state fields $\boldsymbol{x}(\boldsymbol{r}, t; \omega)$, where $\boldsymbol{r}$ represents a coordinate vector within a continuous domain $\mathrm{SG} \in \mathbb{R}^n$. Such state fields are commonly encountered in the physical sciences, where SG typically represents a spatial domain of dimension 1, 2, or 3. When dealing with state fields, the inner products appearing in (3.3), (3.5), and (3.6) become spatial inner products

$$\langle \mathbf{a}, \mathbf{b} \rangle = \int_{\mathrm{SG}} \mathbf{a}(\boldsymbol{r})^{\mathrm{T}} \mathbf{b}(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r}$$

and (3.2) represents a stochastic partial differential equation (SPDE) rather than just a SDE. The ODEs represented by (3.4) and (3.5) subsequently become partial

differential equations (PDEs), which in general can be solved using numerical schemes that discretize the continuous domain SG [141]. If the stochastic boundary conditions for the original SPDE represented by (3.2) are defined as

$$\mathcal{B}\left[\boldsymbol{x}(\boldsymbol{r},t;\omega)\right] = b(\boldsymbol{r},t;\omega),\ \boldsymbol{r} \in \partial\mathrm{SG}\ , \tag{3.7}$$

where $\mathcal{B}$ represents a linear differential operator, then the boundary conditions for the PDEs governing the evolution of the state vector mean and modes are given by

$$\mathcal{B}\left[\overline{\boldsymbol{x}}(\boldsymbol{r},t;\omega)\right] = \mathrm{E}\left[\,b(\boldsymbol{r},t;\omega)\,\right],\ \boldsymbol{r} \in \partial\mathrm{SG}\ , \tag{3.8}$$

$$\mathcal{B}\left[\tilde{\boldsymbol{x}}_i(\boldsymbol{r},t;\omega)\right] = \sum_{j=1}^{N_{\mathrm{DO}}} \mathbf{C}^{-1}_{(i,j)}\,\mathrm{E}\left[\,\phi_j(t;\omega)\,b(\boldsymbol{r},t;\omega)\,\right],\ \boldsymbol{r} \in \partial\mathrm{SG}$$

$$\forall i \in \{1,\ldots,N_{\mathrm{DO}}\}\ . \tag{3.9}$$

When numerically implemented with a discretized domain, all infinite-dimensional state fields $\boldsymbol{x}(\boldsymbol{r},t;\omega)$ reduce to finite-dimensional state vectors $\boldsymbol{x}(t;\omega)$. Mathematical development of the state and model inference schemes in the following sections will thus be premised on the finite-dimensional representation (3.1), with the implicit assumption that numerically discretized representations are used to accommodate systems featuring infinite-dimensional state fields.

3.3 GMM-DO Filter

As mentioned in Section 3.2, the expansion (3.1) represents an affine transformation between mode coefficients and state variables, a relation that is more salient when

(3.1) is written in matrix form

$$\begin{aligned}
\boldsymbol{x}(t;\omega) &\approx \overline{\boldsymbol{x}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\, \tilde{\boldsymbol{x}}_i(t) \\
&= \overline{\boldsymbol{x}}(t) + \left[\begin{array}{ccc} \tilde{\boldsymbol{x}}_1(t) & \cdots & \tilde{\boldsymbol{x}}_{N_{\mathrm{DO}}}(t) \end{array} \right] \left[\begin{array}{c} \phi_1(t;\omega) \\ \vdots \\ \phi_{N_{\mathrm{DO}}}(t;\omega) \end{array} \right] \\
&= \overline{\boldsymbol{x}}(t) + \boldsymbol{\mathcal{X}}(t)\boldsymbol{\phi}(t;\omega) \ . \qquad (3.10)
\end{aligned}$$

where $\boldsymbol{\mathcal{X}}(t) \in \mathbb{R}^{N_{\boldsymbol{x}} \times N_{\mathrm{DO}}}$ represents the matrix of modes and $\boldsymbol{\phi}(t;\omega) \in \mathbb{R}^{N_{\mathrm{DO}}}$ represents a realization of the vector of mode coefficients $\boldsymbol{\Phi}$. Since this relation dictates that any probability distribution for state variables can be equivalently represented by a reduced-dimension probability distribution for mode coefficients, a Bayesian update of the state variable distribution can theoretically be achieved through an equivalent update of the mode coefficient distribution. For specific prior and observation model distributions, this affine transformation to a subspace allows an explicit (analytical) update of the prior distribution parameters in the subspace. The GMM-DO filter is a scheme that takes advantage of this fact to achieve efficient reduced-dimension Bayesian state variable inference [129, 130].

For any set of stochastic dynamical state variables evolved using the DO evolution equations, the GMM-DO filter operates as follows. Anytime observations correlated with state variables are made according to the observation model (1.3), the prior probability distribution for mode coefficients in the DO subspace is approximated using a GMM

$$p_{\boldsymbol{\Phi}}(\boldsymbol{\phi}) \approx \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi},j} \times \mathcal{N}\big(\, \boldsymbol{\phi}\,;\, \boldsymbol{\mu}_{\boldsymbol{\Phi},j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{\Phi},j} \big) \quad \forall \boldsymbol{\phi} \in \mathbb{R}^{N_{\mathrm{DO}}} \ , \qquad (3.11)$$

where N_{GMM} is the to-be-determined number of GMM components, $\pi_{\boldsymbol{\Phi},j}$ the jth component weight, $\boldsymbol{\mu}_{\boldsymbol{\Phi},j}$ the jth component mean vector, and $\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}$ the jth component covariance matrix. This approximation is found by performing a semiparametric fit to

the Monte Carlo samples used to numerically evolve the stochastic mode coefficients. Specifically, the expectation-maximization (EM) algorithm for GMMs [17] is used to find maximum likelihood estimates for the parameters $\pi_{\Phi,j}$, $\boldsymbol{\mu}_{\Phi,j}$, and $\boldsymbol{\Sigma}_{\Phi,j}$, while the selection of the number of GMM components N_{GMM} is directed by the Bayesian information criterion (BIC) [133].

Due to the affine transformation (3.10) relating mode coefficients to state variables, the GMM approximation of the prior mode coefficient distribution (3.11) equivalently represents a GMM approximation of the prior state variable distribution, with variances restricted to the dominant directions of state uncertainty

$$p_{\boldsymbol{X}}(\boldsymbol{x}) \approx \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{X},j} \times \mathcal{N}\big(\boldsymbol{x}\,;\, \boldsymbol{\mu}_{\boldsymbol{X},j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{X},j}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N\boldsymbol{x}}\ , \tag{3.12}$$

where

$$\pi_{\boldsymbol{X},j} = \pi_{\Phi,j} \tag{3.13}$$

$$\boldsymbol{\mu}_{\boldsymbol{X},j} = \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\Phi,j} \tag{3.14}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X},j} = \boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\Phi,j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \tag{3.15}$$

are the jth component weight of the prior state variable GMM approximation, the jth component mean vector, and the jth component covariance matrix, respectively. Further, if the Gaussian observation likelihood function

$$\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x}) = \mathcal{N}\big(\boldsymbol{y}\,;\, \mathbf{H}\boldsymbol{x}\,,\, \mathbf{R}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N\boldsymbol{x}} \tag{3.16}$$

as first defined in (1.4) is used, the Bayesian update of the GMM prior (3.12) is another GMM by conjugacy [129]; the posterior state variable distribution is thus

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) = \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{X}|\boldsymbol{Y},j} \times \mathcal{N}\big(\boldsymbol{X}\,;\, \boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N\boldsymbol{x}}\ , \tag{3.17}$$

with parameters

$$\pi_{\boldsymbol{X}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{X},j} \times \mathcal{N}\big(\boldsymbol{y}\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},j}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big)}{\sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X},k} \times \mathcal{N}\big(\boldsymbol{y}\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},k}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},k}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big)}\,,$$

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j} = \boldsymbol{\mu}_{\boldsymbol{X},j} + \mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},j}\right)\,,$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j} = \left(\mathbf{I} - \mathbf{K}_j\mathbf{H}\right)\boldsymbol{\Sigma}_{\boldsymbol{X},j}$$

$$\forall j \in \{1, \ldots, N_{\mathrm{GMM}}\}$$

and gain matrices defined by

$$\mathbf{K}_j = \boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}}\big(\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big)^{-1} \quad \forall j \in \{1, \ldots, N_{\mathrm{GMM}}\}\,.$$

Though analytically accessible, the posterior GMM state variable distribution (3.17) cannot be directly computed for systems (3.2) of high dimension. Specifically, the storage and manipulation of the prior and posterior state variable GMM component covariance matrices, which are of size $N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}$, is computationally prohibitive. A key advantage of the GMM-DO filter is that the update of the prior GMM state distribution (3.12) is equivalently obtained from the update of the prior GMM coefficient distribution (3.11) into the following posterior GMM coefficient distribution:

$$p_{\boldsymbol{\Phi}|\boldsymbol{Y}}(\boldsymbol{\phi}\,|\,\boldsymbol{y}) = \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} \times \mathcal{N}\big(\boldsymbol{\phi}\,;\, \boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\big) \quad \forall \boldsymbol{\phi} \in \mathbb{R}^{N_{\mathrm{DO}}}\,, \tag{3.18}$$

where

$$\pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{\Phi},j} \times \mathcal{N}\big(\tilde{\boldsymbol{y}}\,;\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\,,\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\big)}{\sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi},k} \times \mathcal{N}\big(\tilde{\boldsymbol{y}}\,;\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},k}\,,\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},k}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\big)}\,,$$

$$\boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j} - \sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \times \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},k}\,,$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \left(\mathbf{I} - \tilde{\mathbf{K}}_j\tilde{\mathbf{H}}\right)\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}$$

$$\forall j \in \{1, \ldots, N_{\mathrm{GMM}}\}$$

and the transformed observation vector realization, transformed observation matrix, transformed gain matrices, and intermediate component mean vectors are defined by

$$\begin{aligned}
\tilde{\boldsymbol{y}} &= \boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}} \ , \\
\tilde{\mathbf{H}} &= \mathbf{H}\boldsymbol{\mathcal{X}} \ , \\
\tilde{\mathbf{K}}_j &= \boldsymbol{\mathcal{X}}^{\mathrm{T}}\mathbf{K}_j \quad \forall j \in \{1, \ldots, N_{\mathrm{GMM}}\} \ , \\
\boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j} &= \boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \tilde{\mathbf{K}}_j \left(\tilde{\boldsymbol{y}} - \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j} \right) \quad \forall j \in \{1, \ldots, N_{\mathrm{GMM}}\} \ . \qquad (3.19)
\end{aligned}$$

This posterior GMM coefficient distribution (3.18) can be shown to be equivalent to the posterior GMM state variable distribution (3.17) through the affine transformation (3.10) if the state vector mean is also updated according to

$$\overline{\boldsymbol{x}}(t^+) = \overline{\boldsymbol{x}}(t^-) + \boldsymbol{\mathcal{X}} \sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \times \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \ .$$

Since all uncertainty resides in the DO subspace, Bayesian updates can only change the mean of the DO coefficients. This state vector update is thus responsible for bringing all mean updates back in to the state space. The full demonstration of this equivalency is presented in Appendix C.

We first note that all computations in this update are defined by analytical equations. Then, whereas the explicit calculation of (3.17) is infeasible, the calculation of (3.18) is untroublesome. Critically, no matrices of size larger than $N_{\boldsymbol{X}} \times N_{\mathrm{DO}} \ll N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}$ are manipulated in the update (3.18), rendering the Bayesian update of the mode coefficient distribution computationally tractable for high-dimensional systems.

Finally, new Monte Carlo samples are drawn from the posterior GMM mode coefficient distribution (3.18), which are dynamically evolved with the DO evolution equations until new observations are made and the GMM-DO filter is applied again. By using GMM approximations, the GMM-DO filter is able to accommodate systems featuring multimodal state variable distributions while avoiding the excessive granularity of kernel approximations [129]. The GMM-DO filter, in conjunction with the DO evolution equations, is thus a computationally expedient method for performing

Bayesian state variable inference for high-dimensional, nonlinear stochastic dynamical systems.

3.4 Reduced-Dimension State Augmentation

Parameter uncertainty in stochastic dynamical models can be directly coupled to state uncertainty using a well-known technique from estimation theory known as state augmentation [45]. This technique enables joint Bayesian inference of state variables and parameters from observations correlated to the state variables. The conceptual premise of state augmentation is simple: treat all uncertain parameters as time-invariant uncertain state variables. Concatenating vectors of state variables and parameters and using the notation specified in Table 3.1 then leads to

$$\boldsymbol{X}_{\boldsymbol{\Theta}}(t) = \begin{bmatrix} \boldsymbol{X}(t) \\ \boldsymbol{\Theta} \end{bmatrix} \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}} , \tag{3.20}$$

which is referred to as augmented state vector and for which realizations are given by

$$\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega) = \begin{bmatrix} \boldsymbol{x}(t;\omega) \\ \boldsymbol{\theta}(\omega) \end{bmatrix} . \tag{3.21}$$

The general, nonlinear differential operator that governs the evolution of the augmented state vector—the augmented stochastic dynamical model—is given by

$$\mathcal{M}_{\boldsymbol{\Theta}}\left[\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega);\omega\right] = \frac{\mathrm{d}\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{x}(t;\omega) \\ \boldsymbol{\theta}(\omega) \end{bmatrix} = \begin{bmatrix} \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] \\ 0 \end{bmatrix} . \tag{3.22}$$

Reduced-dimension state augmentation is achieved using an expansion similar to

(3.1) for the augmented state vector

$$\begin{bmatrix} \boldsymbol{x}(t;\omega) \\ \boldsymbol{\theta}(\omega) \end{bmatrix} \approx \begin{bmatrix} \overline{\boldsymbol{x}}(t) \\ \overline{\boldsymbol{\theta}} \end{bmatrix} + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega) \begin{bmatrix} \tilde{\boldsymbol{x}}_i(t) \\ \tilde{\boldsymbol{\theta}}_i(t) \end{bmatrix}$$

$$\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega) \approx \overline{\boldsymbol{x}}_{\boldsymbol{\Theta}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\, \tilde{\boldsymbol{x}}_{\boldsymbol{\Theta},i}(t) \,, \tag{3.23}$$

where $\overline{\boldsymbol{x}}_{\boldsymbol{\Theta}}(t) \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}}$ denotes the mean of the augmented state vector, $\tilde{\boldsymbol{x}}_{\boldsymbol{\Theta},i}(t) \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}}$ the ith mode of the augmented state vector, $\overline{\boldsymbol{\theta}} \in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$ the mean of the parameter vector, and $\tilde{\boldsymbol{\theta}}_i(t) \in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$ the parameter elements of the ith augmented state vector mode. Given the augmented stochastic dynamical model (3.22), DO evolution equations as those developed in Section 3.2 (i.e. (3.4)–(3.6)) can be used to evolve the augmented state vector mean, augmented state vector modes, and mode coefficients defined in (3.23).

It can be recognized that there is no mathematical distinction between state variables and parameters in the augmented state vector expansion (3.23). Mode coefficients are shared between the two groups of uncertain variables and the DO modes of the augmented state vector represent directions of joint state and parameter uncertainty. Statistical correlations between state variables and parameters manifest as concurrently nonzero elements in the augmented modes. If state variables and parameters are largely uncorrelated however, the modes will feature only either nonzero state variable elements or nonzero parameter elements, indicating orthogonality between state uncertainty and parameter uncertainty. Note that even though parameters are specified as time-invariant in the augmented stochastic dynamical model (3.22) and hence the mean of the parameter vector in (3.23) is time-invariant (in between observation updates), the parameter elements of the augmented state vector modes are still subject to evolution. This reflects the fact that the statistical relations between state uncertainty and parameter uncertainty may change as the stochastic dynamical system under investigation evolves. Since the augmented state vector modes collectively

define the stochastic subspace within which the majority of the joint state and parameter uncertainty resides, all elements of the modes are subject to change in order to reflect all possible reorientations of this subspace. Of course, when an observation is made, the Bayesian update of the parameter distribution also updates the mean of the parameter values; these updated parameter means are then subsequently utilized in (3.22) and (3.23) until the next observation.

State augmentation also enables joint Bayesian inference of state variables and parameters. If parameters were to be inferred separate from state variables, Bayesian update of the probability distribution for parameters would entail the calculation

$$p_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}\,|\,\boldsymbol{y}) = \frac{p_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}\,|\,\boldsymbol{\theta})}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\, p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$$

$$= \frac{\int p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Theta}}(\boldsymbol{y}\,|\,\boldsymbol{x},\boldsymbol{\theta})\, p_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\, p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$$

$$= \frac{\int \mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})\, p_{\boldsymbol{X}|\boldsymbol{\Theta}}(\boldsymbol{x}\,|\,\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\, p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{N_{\boldsymbol{\Theta}}} \,. \tag{3.24}$$

(3.24) is analogous to (1.9) and the last integral in (3.24) is analogous to (1.11), the marginal likelihood integral required for model inference, and the same difficulties discussed in Chapter 2 for the calculation of (1.11) apply to the integral here. In the present 'parameter evidence' computation, the role of $\mathcal{M}_n$ is played by $\boldsymbol{\theta}$, the parameter mean value (and its pdf) is estimated as model evidence was in (1.9).

State augmentation allows us to solve these difficulties for parameter inference the same way we solved them for model inference, i.e. using the GMM-DO filter to estimate the state pdf and so allow the computation of the integral in (3.24). Of course, as we will see, once the parameter pdf has been updated at a given time, the posterior parameter pdf is used at later times.

Specifically, the observation likelihood $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)$ can be replaced by an equivalent

augmented likelihood function

$$\mathcal{L}_{\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{x}_{\boldsymbol{\Theta}}) = \mathcal{L}(\boldsymbol{y}|\left[\,\mathbf{I}^{N_{\boldsymbol{X}}}\;\mathbf{0}\,\right]\boldsymbol{x}_{\boldsymbol{\Theta}}) \quad \forall \boldsymbol{x}_{\boldsymbol{\Theta}} \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}} \tag{3.25}$$

that simply ignores the parameter elements of the augmented state vector when calculating observation likelihoods. If the original state likelihood function is a Gaussian function, as assumed by the GMM-DO filter in (3.16), the augmented likelihood function becomes

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\Theta}}(\boldsymbol{y}|\boldsymbol{x}_{\boldsymbol{\Theta}}) &= \mathcal{N}\big(\boldsymbol{y};\,\mathbf{H}\left[\,\mathbf{I}^{N_{\boldsymbol{X}}}\;\mathbf{0}\,\right]\boldsymbol{x}_{\boldsymbol{\Theta}},\,\mathbf{R}\big)\\ &= \mathcal{N}\big(\boldsymbol{y};\,[\,\mathbf{H}\;\mathbf{0}\,]\,\boldsymbol{x}_{\boldsymbol{\Theta}},\,\mathbf{R}\big)\\ &= \mathcal{N}\big(\boldsymbol{y};\,\mathbf{H}_{\boldsymbol{\Theta}}\boldsymbol{x}_{\boldsymbol{\Theta}},\,\mathbf{R}\big) \quad \forall \boldsymbol{x}_{\boldsymbol{\Theta}} \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}}\end{aligned} \tag{3.26}$$

and the GMM-DO filter can be applied as usual using the augmented observation matrix $\mathbf{H}_{\boldsymbol{\Theta}} = [\,\mathbf{H}\;\mathbf{0}\,] \in \mathbb{R}^{N_{\boldsymbol{Y}}\times(N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}})}$ in place of the original observation matrix $\mathbf{H}$. As long as some of the modes in the expansion (3.23) feature both nonzero state variable and parameter elements—i.e. as long as there is some finite correlation between state variables and parameters—the probability distribution for parameters will be jointly updated with the probability distribution for state variables when the GMM-DO filter updates the reduced-dimension distribution for the shared mode coefficients. After this GMM-DO update, the integration of (3.22) and (3.23) will utilize the updated parameter mean and evolve the joint state-parameter probability distribution in accord with the dynamics until the next observation.

In summary, reduced-dimension state augmentation extends both the DO evolution equations and the GMM-DO filter to enable joint uncertainty evolution and inference of state variables and parameters. For the remainder of this chapter, the augmented state vector $\boldsymbol{X}_{\boldsymbol{\Theta}}$ (and realization $\boldsymbol{x}_{\boldsymbol{\Theta}}$) will be denoted as simply $\boldsymbol{X}$ (and $\boldsymbol{x}$) for notational expediency, with the implicit assumption that $\boldsymbol{X}$ encapsulates both uncertain state variables and uncertain parameters coupled through state augmentation.

3.5 GMM-DO Marginal Likelihood Calculation

3.5.1 Analytical GMM Approach

As reviewed in Section 2.2, the marginal likelihood integral in (1.11) that is essential for Bayesian model inference can be calculated analytically for the special case of Gaussian model-conditional state variable distributions coupled with a Gaussian observation likelihood function. In this section, we will show that for the GMM state variable distributions and Gaussian observation models and likelihood function used by the GMM-DO filter, a similar analytical solution for the marginal likelihood integral is available.

Specifically, using (3.16) and (3.12) for $\mathcal{L}(\boldsymbol{y}\,|\,\bullet)$ and $p_{\boldsymbol{X}|\mathcal{M}}(\bullet\,|\,\mathcal{M}_n)$ respectively, we obtain

$$\begin{aligned} p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) &= \int \mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})\; p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\; \mathrm{d}\boldsymbol{x} \\ &= \int \mathcal{N}\big(\boldsymbol{y};\, \mathbf{H}\boldsymbol{x},\, \mathbf{R}\big) \\ &\qquad \left(\sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\mathcal{M}_n,j} \times \mathcal{N}\big(\boldsymbol{x};\, \boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n,j},\, \boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n,j}\big) \right) \mathrm{d}\boldsymbol{x}\,. \end{aligned} \tag{3.27}$$

Here, since we are once again considering model uncertainty, we use $\mathcal{M}_n$ in place of $\mathcal{M}$ to represent a realization of the uncertain model. Note that state augmentation is also featured implicitly, with both state variables and parameters encapsulated in the vector $\boldsymbol{x}$. Analogous to the derivation presented in Section 2.2, since the integral in (3.27) is taken over all values of $\boldsymbol{x}$, a linear transformation of the integration variable can be performed without changing the value of the integral. Then, using the linear

transformation properties of Gaussian distributions, (3.27) can be rewritten as

$$
\begin{aligned}
p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) &= \int \mathcal{N}\big(\boldsymbol{y}-\mathbf{H}\boldsymbol{x}\,;\,0\,,\,\mathbf{R}\big) \\
&\qquad \left(\sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\mathcal{M}_n,j} \times \mathcal{N}\big(\mathbf{H}\boldsymbol{x}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n,j}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\big)\right) \mathrm{d}\mathbf{H}\boldsymbol{x} \\
&= \int \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\mathcal{M}_n,j} \times \mathcal{N}\big(\boldsymbol{y}-\mathbf{H}\boldsymbol{x}\,;\,0\,,\,\mathbf{R}\big) \\
&\qquad \mathcal{N}\big(\mathbf{H}\boldsymbol{x}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n,j}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\big)\,\mathrm{d}\mathbf{H}\boldsymbol{x}\;. \qquad (3.28)
\end{aligned}
$$

Interchanging integration and summation, then factoring,

$$
\begin{aligned}
p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) &= \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\mathcal{M}_n,j} \times \int \mathcal{N}\big(\boldsymbol{y}-\mathbf{H}\boldsymbol{x}\,;\,0\,,\,\mathbf{R}\big) \\
&\qquad \mathcal{N}\big(\mathbf{H}\boldsymbol{x}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n,j}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\big)\,\mathrm{d}\mathbf{H}\boldsymbol{x} \\
&= \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\mathcal{M}_n,j} \times \big[\mathcal{N}\big(\bullet\,;\,0\,,\,\mathbf{R}\big) \\
&\qquad * \mathcal{N}\big(\bullet\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n,j}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\big)\big](\boldsymbol{y})\;, \qquad (3.29)
\end{aligned}
$$

where $*$ again represents the convolution operator. Using once more the convolution identity introduced in Section 2.2 and derived in Appendix A,

$$
p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) = \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\mathcal{M}_n,j} \times \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}|\mathcal{M}_n,j}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}|\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\big)\;. \qquad (3.30)
$$

Analogous to (3.17), the analytically accessible (3.30) cannot be directly computed for systems of high dimension due to the prohibitive size of the state variable GMM component covariance matrices involved. Fortunately, the utilization of the GMM-DO approach to compute (3.30) is possible, again because of (3.10), the affine transformation that relates state variables and mode coefficients. Substituting the

prior GMM-DO fit in the DO subspace, (3.13)–(3.15), into (3.30) yields

$$
\begin{aligned}
p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) &= \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\mathcal{M}_n,j} \\
&\qquad \times \mathcal{N}\big(\boldsymbol{y}\,;\, \mathbf{H}\left(\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\right)\,,\, \mathbf{H}\left(\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\boldsymbol{\mathcal{X}}^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big) \\
&= \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\mathcal{M}_n,j} \\
&\qquad \times \mathcal{N}\big(\boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}}\,;\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\,,\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\left(\mathbf{H}\boldsymbol{\mathcal{X}}\right)^{\mathrm{T}} + \mathbf{R}\big) \\
&= \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\mathcal{M}_n,j} \times \mathcal{N}\big(\tilde{\boldsymbol{y}}\,;\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\,,\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\big)\,, \qquad (3.31)
\end{aligned}
$$

where the definitions in (3.19) have been used. The explicit evaluation of the Gaussian functions in (3.31) is computationally expedient. None of the matrices involved exceed $N_{\boldsymbol{Y}} \times N_{\boldsymbol{X}}$ in size and $N_{\boldsymbol{Y}}$ is typically orders of magnitude smaller than $N_{\boldsymbol{X}}$, especially for systems featuring infinite-dimensional state fields (or their large-dimensional discretized versions) [48]. Furthermore, all the quantities in (3.31) are found in the process of performing Bayesian state variable inference with the GMM-DO filter, thus necessitating no new computations. (3.31) can be used to calculate marginal likelihoods for all candidate models, which can subsequently be used to perform Bayesian update of the model distribution as represented by (1.9).

3.5.2 Computational Alternatives

If an alternative to the GMM-DO filter is used to perform Bayesian state variable inference in conjunction with the DO evolution equations, reduced-dimension marginal likelihood calculation can still be accomplished even without the analytical expediency provided by GMM state variable distributions. Since our present numerical implementations of the DO evolution equations [141] evolve the mode coefficients in the expansion (3.1) using Monte Carlo sampling methods, a set of samples from the prior mode coefficient distribution is at hand anytime observations are made and inference is to be performed. These samples can be used to find marginal likelihood

AMEs, as reviewed in Section 2.4, by first transforming the samples using (3.10) then calculating (2.9)

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) \approx \frac{1}{N_{\mathrm{MC}}} \sum_{k=1}^{N_{\mathrm{MC}}} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^-)$$
$$= \frac{1}{N_{\mathrm{MC}}} \sum_{k=1}^{N_{\mathrm{MC}}} \mathcal{L}(\boldsymbol{y}|\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\phi}_k^-) \ , \tag{3.32}$$

where $\boldsymbol{\phi}_k^-$ represents the kth of $N_{\mathrm{MC}} \in \mathbb{N}$ samples from the prior mode coefficient distribution and $\boldsymbol{x}_k^-$ represents the equivalent kth sample from the prior state variable distribution.

Any state variable inference scheme operating in conjunction with the DO evolution equations will also necessarily generate a set of Monte Carlo samples from the posterior mode coefficient distribution (3.18) in order to continue the evolution of the mode coefficients following inference. This set of posterior samples can be used to find marginal likelihood HMEs, as reviewed in Section 2.4, by again transforming the samples using (3.10) then calculating (2.10)

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n) \approx \left(\frac{1}{N_{\mathrm{MC}}} \sum_{k=1}^{N_{\mathrm{MC}}} \mathcal{L}(\boldsymbol{y}|\boldsymbol{x}_k^+)^{-1} \right)^{-1}$$
$$= \left(\frac{1}{N_{\mathrm{MC}}} \sum_{k=1}^{N_{\mathrm{MC}}} \mathcal{L}(\boldsymbol{y}|\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\phi}_k^+)^{-1} \right)^{-1} \ , \tag{3.33}$$

where $\boldsymbol{\phi}_k^+$ represents the kth of $N_{\mathrm{MC}} \in \mathbb{N}$ samples from the posterior mode coefficient distribution and $\boldsymbol{x}_k^+$ represents the equivalent kth sample from the posterior state variable distribution. Note that while directly drawing samples from either the prior or posterior state variable distributions may be intractable for high-dimensional systems, drawing samples from the reduced-dimension prior or posterior mode coefficient distributions is both feasible and efficient as the modes in the expansion (3.1) restrict sampling to the time-dependent stochastic subspace within which the majority of the state uncertainty resides.

3.6 Summary

The methodology developed here for Bayesian inference of stochastic dynamical models can be summarized in the following four-step procedure premised on the general four-step procedure for stochastic dynamical modeling introduced in Section 1.1:

1. *Model formulation.* Formulate a set of candidate models for the true stochastic dynamical model governing the system under investigation. For each candidate, specify stochastic dynamical equations, spatial geometry, boundary conditions, and initial conditions, as represented in (1.2).

2. *Uncertainty initialization.*

   (a) *State variable and parameter uncertainty initialization.* For each candidate model, specify an initial joint probability distribution for state variables and parameters. Initialize the augmented state vector mean as represented in the expansion (3.23) by finding the statistical mean of this distribution. Initialize augmented modes and mode coefficients by orienting the modes in the directions of greatest joint state variable and parameter uncertainty and specifying corresponding probability distributions for the mode coefficients. A general computational procedure for accomplishing this initialization of modes and mode coefficients is to perform Karhunen-Loève transforms of the initial joint state variable and parameter covariance matrices [84, 86], a procedure that leads to the optimal selection of modes and mode coefficients when the initial probability distributions are Gaussian [121].

   (b) *Model uncertainty initialization.* Define an initial discrete probability distribution over the set of candidate models. The distribution value corresponding to any candidate model represents the probability that that candidate model is the true model.

   (c) *Initial Bayesian uncertainty quantification.* As needed, compute the initial probability distribution for state variables and parameters that accounts

for model uncertainty by combining the model-conditional initial distributions using (1.5).

3. *Uncertainty evolution.*

   (a) *Model-conditional uncertainty evolution.* Use the DO evolution equations (3.4)–(3.6) to evolve the augmented state vector means, augmented modes, and mode coefficients for each candidate model.

   (b) *Prior Bayesian uncertainty quantification.* As needed, compute the prior probability distribution for state variables and parameters that accounts for model uncertainty by combining the model-conditional prior distributions using (1.5).

4. *Inference.*

   (a) *State variable and parameter inference.* When an observation is made, use the GMM-DO filter to perform a Bayesian update of the mode coefficient distribution for each candidate model.

   (b) *Marginal likelihood calculation.* Calculate marginal likelihoods for all candidate models using (3.31).

   (c) *Model inference.* Use (1.9) along with the marginal likelihoods to perform a Bayesian update of the model distribution.

   (d) *Posterior Bayesian uncertainty quantification.* As needed, compute the posterior probability distribution for state variables and parameters that accounts for model uncertainty by combining the model-conditional posterior distributions using (1.5).

A graphical representation of this four-step procedure is provided in Figure 3-1. N_{DO} and N_{GMM} are in general both time- and model-dependent but for notational simplicity, these dependencies are not explicitly represented in the figure. Similarly, the number of candidate models $N_{\mathcal{M}}$ could in general be a function of time, but this too is omitted in the graphical representation.

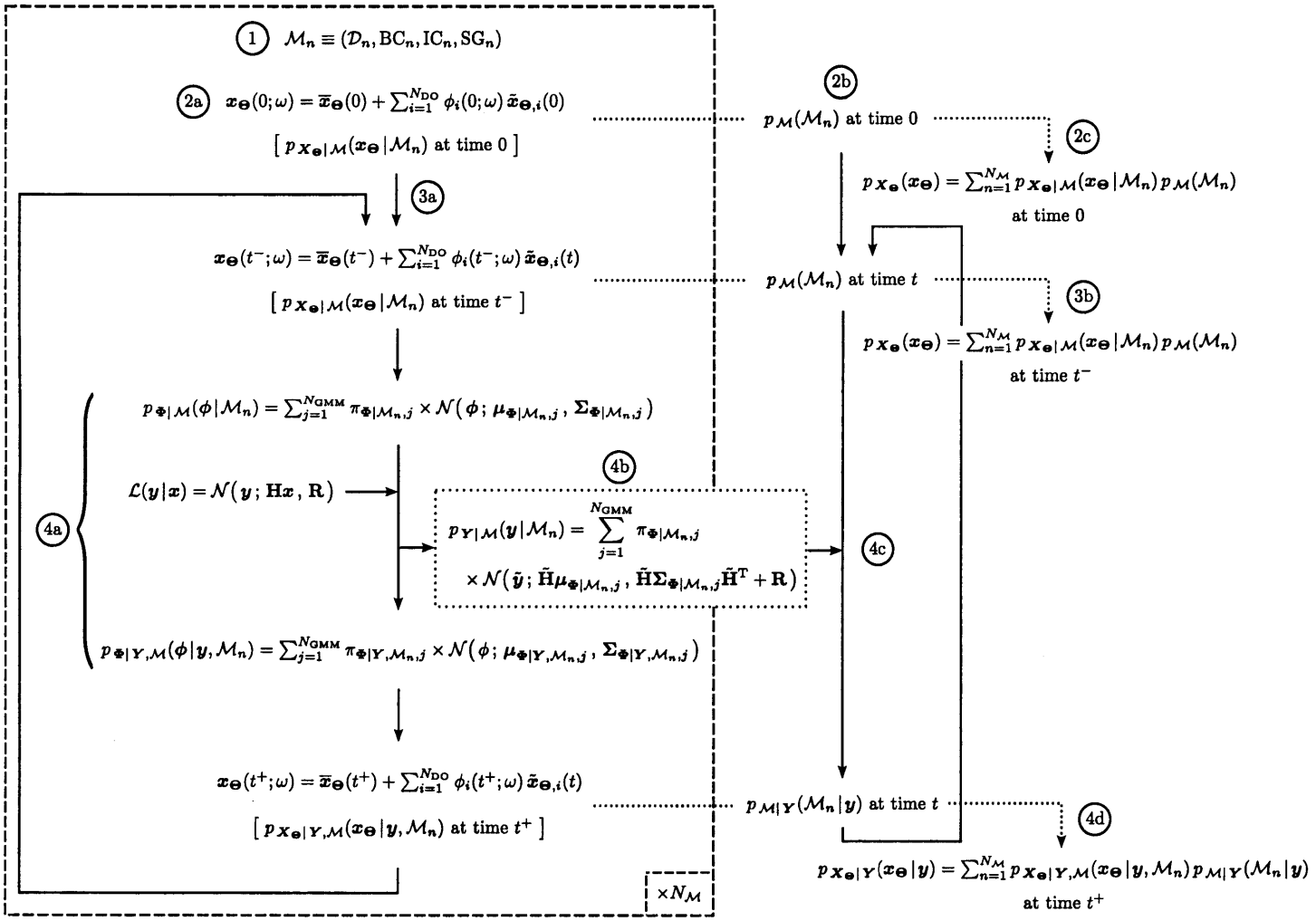

Figure 3-1: Four-step procedure for Bayesian inference of stochastic dynamical models.

3.6.1 Comments

Independence between Candidate Models

Note that all operations involved in Steps 2a, 3a, 4a, and 4b (i.e. the left half of Figure 3-1) are performed independently for each candidate model. All computations relevant to these steps can hence be performed in parallel for each candidate model. Another important consequence of this independence is that idiosyncratic variations of the DO evolution equations and the GMM-DO filter can be applied to any of the candidate models without affecting the others. For example, each candidate model could employ a different stochastic subspace dimension for the expansion (3.23), with some possibly adopting dynamically evolving subspace dimensions as used in [122]. Furthermore, as mentioned in Section 2.1, the marginal likelihood calculations that constitute Step 4b can be performed differently for each candidate model. If an alternative to the GMM-DO filter is used for one particular candidate for example, the reduced-dimension computational approximations (3.32) and (3.33) could be used in place of (3.31).

Independence from Model Distribution

Another important feature of the present methodology is that the operations involved in Steps 2a, 3a, 4a, and 4b are also independent of the model distribution (i.e. no arrows point from the right half of Figure 3-1 to the left). In fact, as long as the marginal likelihoods calculated in Step 4b are stored, Bayesian update of the model distribution as represented by Step 4c need not be performed unless an explicit rendering of the state variable distribution with model uncertainty accounted for is required (i.e. if Steps 2c, 3b, and 4d are necessary). This allows for new candidate models to be added to the set of viable candidates, even after Steps 2a, 3a, 4a, and 4b have been carried out for existing candidates. One simply needs to respecify the initial model distribution to include the new candidates, then perform the independent steps afresh for the new candidates only. The marginal likelihoods generated for the new candidates can be combined with those already stored to perform Bayesian update of the

new model distribution.

Note that one can always choose to perform model inference either immediately after each observation or only after all observations have been made. The decision between these two alternatives should be made considering available computational resources and the utility of real-time state variable and parameter distributions that account for model uncertainty. If real-time distributions are crucial, model inference should naturally be performed at all observation times in order to carry out Steps 3b and 4d. If computational resources are relatively limited however, model inference may, out of necessity, be left for post-processing. The present methodology allows for both courses of action.

Recursive Model Inference

The model inference procedure described above could also be extended to a form of 'recursive model inference' that is similar in fashion to parameter inference. In this case, when the model distribution is updated at each observation time, the model itself could be treated as a parameter, in that we could sample model uncertainty as we sample parameter uncertainty. Consequently, the set of candidate models could be updated with time, perhaps directly with an algorithm such as the GMM-DO filter, and these updates would then influence model predictions at future times. This would represent a new extension to the current methodology since in this case model inference and state inference would not be independent: updates to the set of candidate models would influence the dynamic evolution of the state variables, just as updates to parameter distributions influence state variables in the current formulation. Further investigation of this possibility is left for future work.

In sum, the present methodology constitutes a versatile approach to Bayesian inference of stochastic dynamical models. Applications to two high-dimensional, nonlinear systems are presented in the following chapters.

Chapter 4

Flow Past an Obstacle

4.1 Flow Past a Cylinder or Island

The flow past a cylinder or idealized island is a classic dynamical system in fluid mechanics that features a two-dimensional fluid flowing around an impermeable circular obstacle. The fluid is assumed to be incompressible and Newtonian while the obstacle is assumed to be fixed. Furthermore, the fluid velocity far away from the obstacle is assumed to be uniform and steady—i.e. both spatially and temporally invariant—at an infinite distance away from the obstacle. This system has been extensively studied within the fluid mechanics community due to both its exceptionally simple formulation and the startling intricacy of the fluid flow patterns it can evoke [32, 41, 42, 127, 137, 150].

The key parameter that governs the properties of the fluid flow in the flow past a cylinder system is the non-dimensional Reynolds number

$$\mathrm{Re} = \frac{V_\infty L}{\nu} , \tag{4.1}$$

where V_∞ represents the velocity of the fluid an infinite distance away from the obstacle, L represents the projected width of the obstacle, and ν represents the dynamic viscosity of the fluid. The Reynolds number can be interpreted as the ratio between the inertial and viscous forces in the fluid [78]. For Reynolds numbers below ap-

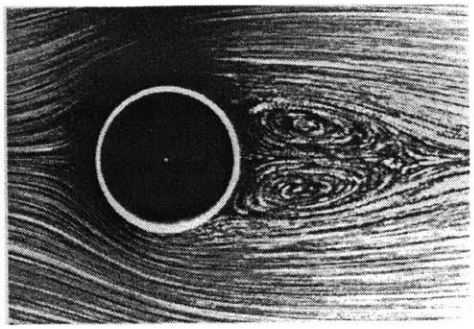

Figure 4-1: Visualization of stable recirculation zones in an experimental investigation of the flow past a cylinder system with a Reynolds number of 26 [143].

proximately 40, the fluid flow downstream of the obstacle is symmetric, with a pair of stable recirculation zones appearing behind the obstacle for Reynolds numbers greater than 5. These recirculation zones can be seen in Figure 4-1 and increase in size with increasing Reynolds number. At Reynolds numbers greater than 40, the recirculation zones become unstable and exhibit periodic, asymmetric vortex shedding. This results in dynamic fluid flow patterns known as von Kármán vortex streets downstream of the obstacle, as illustrated in Figure 4-2. Systems featuring Reynolds numbers greater than 200 exhibit more complex aperiodic patterns while Reynolds numbers greater than 10^5 result in turbulent fluid flow [135].

Countless variations of the classic flow past a cylinder system have been investigated, including systems featuring non-Newtonian fluids [24], rotating obstacles [104], oscillating obstacles [18], noncircular obstacles [67], multiple obstacles [94], and three-dimensional fluids and obstacles [43]. In addition to serving as standardized platforms for conducting theoretical, experimental, and computational studies of fluid phenomena, the flow past a cylinder system and its variants have also been used as practical models of real systems found in both natural and engineered environments [151]. An example of an atmospheric system that is well-modeled as a flow past a cylinder is depicted in Figure 4-3.

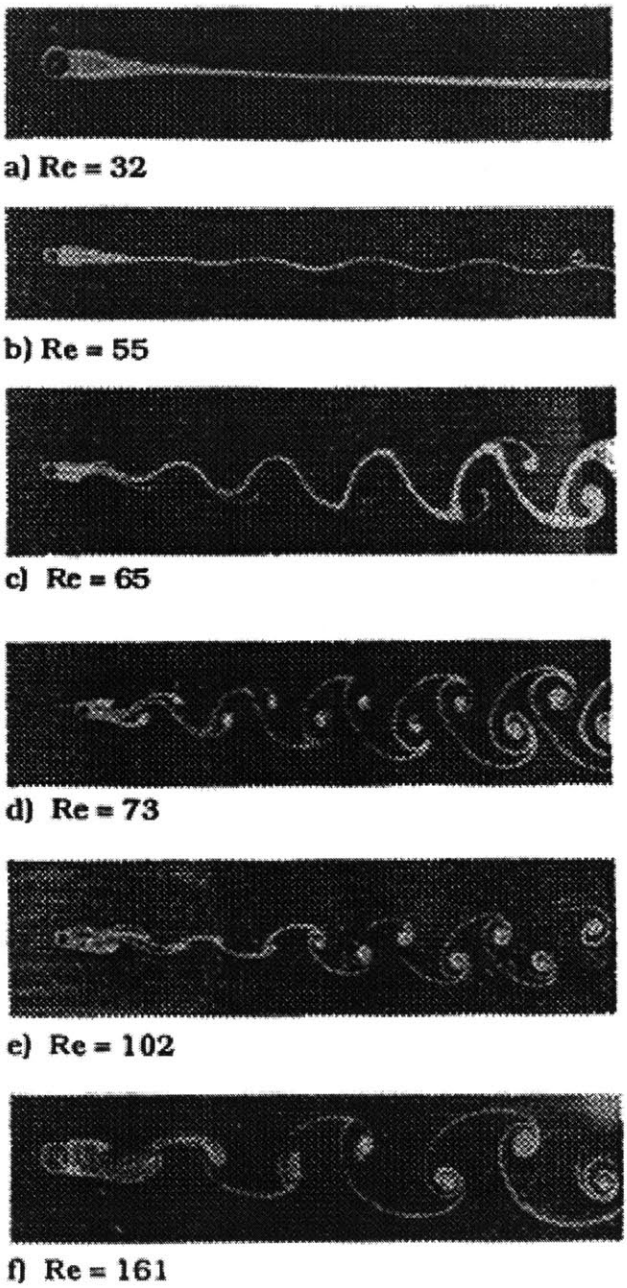

Figure 4-2: Visualizations of von Kármán vortex streets in an experimental investigation of the flow past a cylinder system with several Reynolds numbers [58].

Figure 4-3: Image of clouds off the Chilean coast near the Juan Fernandez Islands taken by the Landsat 7 satellite on September 15, 1999 [106]. The fluid flow pattern behind the island at the bottom-left of the image can be modeled as a variation of the flow past a cylinder system.

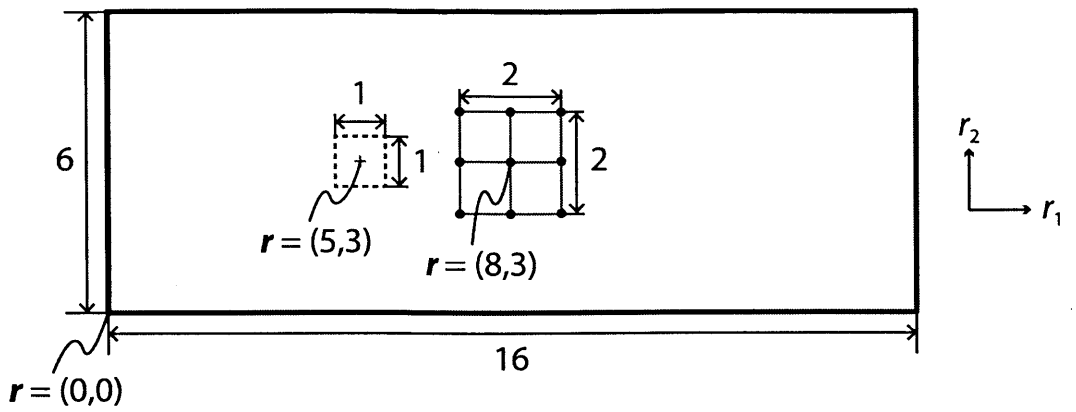

Figure 4-4: Two-dimensional spatial domain of the stochastic flow past an obstacle system. All lengths and coordinates are given in non-dimensionalized length units. The bounding area for the obstacle is indicated by the dashed square. The observation array is indicated by the nine points situated at the center of the domain.

4.2 Stochastic Flow Past an Obstacle System

The first stochastic dynamical system considered for application of the Bayesian model inference methodology developed in Chapter 3 is a simulated stochastic variation of the flow past a cylinder system. The two-dimensional spatial domain of the system is shown in Figure 4-4, with fluid flow proceeding left to right in the positive r_1 direction. As in the classic rendition of the system, the fluid is assumed to be incompressible and Newtonian while the obstacle is assumed to be fixed. Uncertainty in the fluid velocity field—the infinite-dimensional state field of the system—is introduced in three ways: 1) uncertainty in the initial velocity field; 2) uncertainty in the Reynolds number of the fluid flow; and 3) uncertainty in the shape of the obstacle.

As mentioned in Section 4.1, the fluid flow pattern downstream of the obstacle in the flow past a cylinder system is periodic for Reynolds numbers between 40 and 200. Vortex shedding frequency and morphology are predictable for any given Reynolds number within this range. The phase of vortex shedding however is highly unpredictable, as shedding is always triggered chaotically by asymmetric perturbations of the otherwise symmetric fluid domain. Even mild uncertainty in the initial conditions of the system can thus lead to substantial uncertainty in the shedding phase, which,

as exhibited in Figure 4-5, can correspond to substantial uncertainty in the velocity field.

Further uncertainty in the velocity field can arise if the Reynolds number of the fluid flow, which governs both the frequency and morphology of vortex shedding downstream of the obstacle, is also uncertain. As exhibited in Figure 4-6, higher Reynolds numbers correspond to greater shedding frequencies and vortices of higher energy.

Finally, uncertainty in the shape of the obstacle within the generic spatial domain illustrated in Figure 4-4 can have a significant effect on the uncertainty of the velocity field. As exhibited in Figure 4-7, different obstacle shapes can result in substantially different downstream fluid flow patterns. Even when projected widths are identical, different shapes can lead to different vortex shedding frequencies and morphologies for the same Reynolds number. For the system considered here, five shapes are allowed for the obstacle, with spatial dimensions as specified in Figure 4-8. The details of the flow patterns for these five shapes are different, as can be seen in Figure 4-7. Specifically, the circle and square exhibit both similar near-field and far-field vortex shedding patterns, but with slight variations in vortex intensities (most notably in the near-field). The other three shapes also exhibit similar near-field vortex shedding patterns, but the far-field pattern of the downstream-pointing triangle differs markedly from those of the upstream-pointing triangle and diamond. These differences can be exploited to infer the correct obstacle shape for the system when this feature is uncertain.

4.3 Description of the Experiments

4.3.1 Model Formulation

Five possible shapes are considered for the obstacle in the system, leading to five possible configurations for the generic spatial domain illustrated in Figure 4-4. Each of these domains represents a distinct candidate model for the system, with each

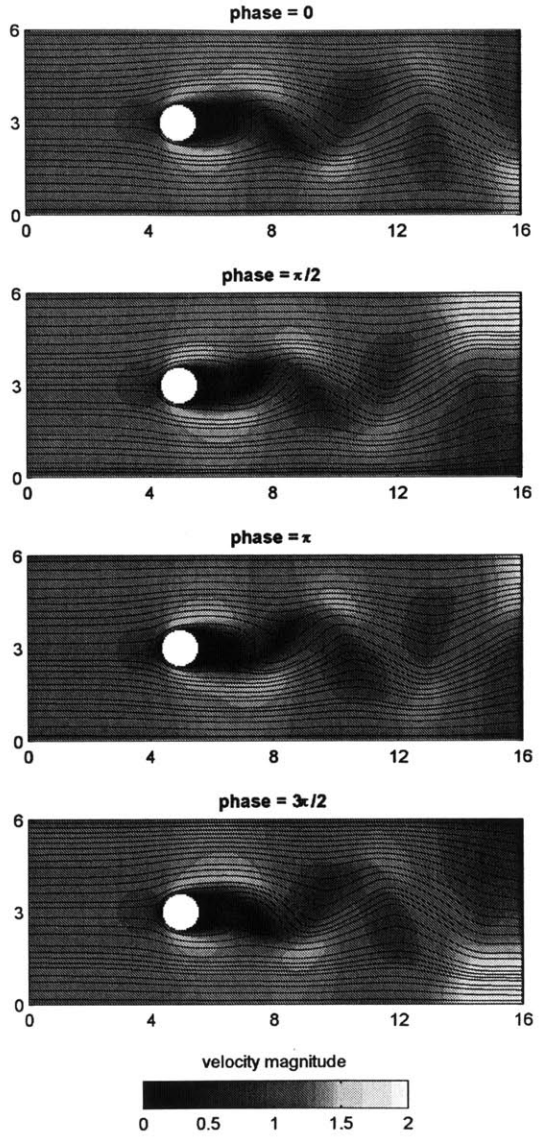

Figure 4-5: Fluid flow patterns for realizations of the stochastic flow past an obstacle system featuring a circular obstacle, a Reynolds number of 75, and phase shifts of 0, $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$. Directions of fluid flow are indicated with streamlines. Note that the flow patterns for phase shifts of 0 and π are a mirrored pair about the $r_2 = 3$ centerline, as are the flow patterns for phase shifts of $\frac{\pi}{2}$ and $\frac{3\pi}{2}$.

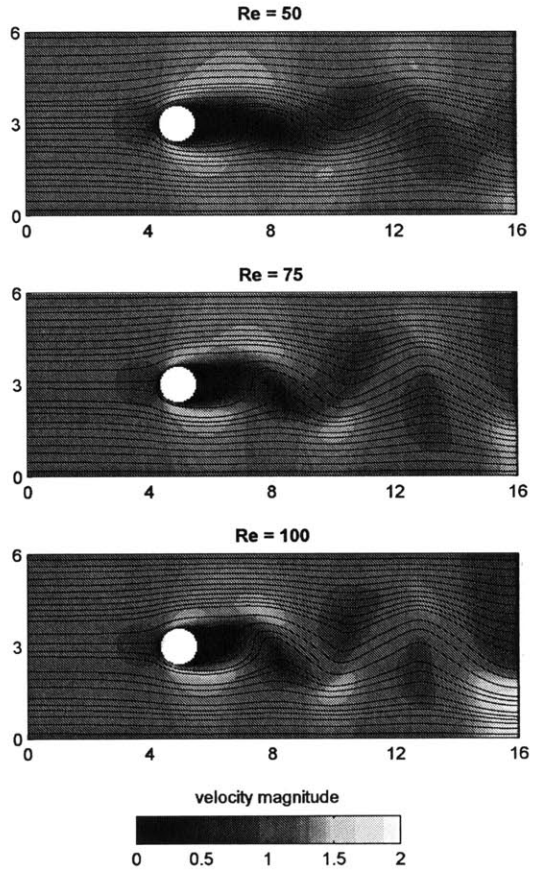

Figure 4-6: Fluid flow patterns for realizations of the stochastic flow past an obstacle system featuring a circular obstacle and Reynolds numbers of 50, 75, and 100 at the same non-dimensional time. Phase shifts indicate different vortex shedding frequencies.

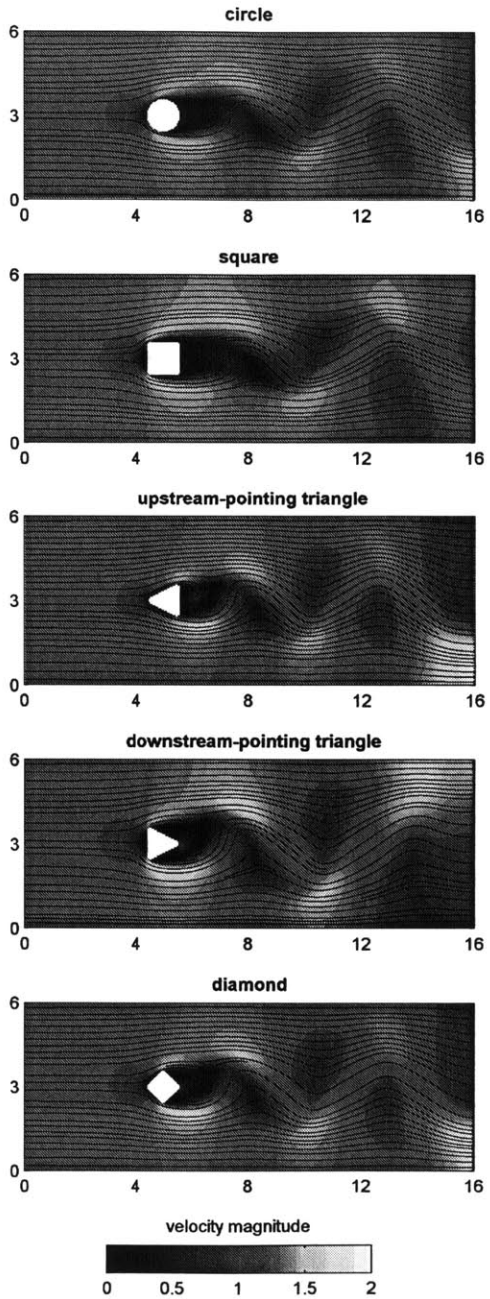

Figure 4-7: Fluid flow patterns for realizations of the stochastic flow past an obstacle system featuring different obstacle shapes at the same non-dimensional time, all with a Reynolds number of 75. The projected width of all the obstacles is identical. Phase shifts indicate different vortex shedding frequencies.

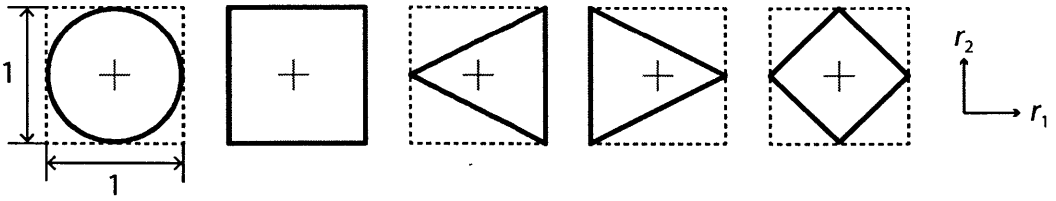

Figure 4-8: Five possible obstacle shapes: a circle, a square, an upstream-pointing triangle, a downstream-pointing triangle, and a diamond. All lengths are given in non-dimensionalized length units. The bounding area indicated for each of the shapes corresponds to the area indicated in Figure 4-4.

candidate assumed to be equally likely a priori (i.e. a uniform initial model distribution is used). We proceed to describe each component of (1.2) for each of these five candidate models.

$\mathcal{D}_n$ — The same set of stochastic dynamical equations is used to model the evolution of the fluid velocity field for each of the candidate models: the two-dimensional Navier-Stokes equations for an incompressible, Newtonian fluid in non-dimensional form

$$\frac{\partial v_1}{\partial t} = -\frac{\partial (v_1^2)}{\partial r_1} - \frac{\partial (v_1 v_2)}{\partial r_2} + \frac{1}{\mathrm{Re}(\omega)} \left(\frac{\partial^2 v_1}{\partial r_1^2} + \frac{\partial^2 v_1}{\partial r_2^2} \right) - \frac{\partial p}{\partial r_1} , \tag{4.2}$$

$$\frac{\partial v_2}{\partial t} = -\frac{\partial (v_1 v_2)}{\partial r_1} - \frac{\partial (v_2^2)}{\partial r_2} + \frac{1}{\mathrm{Re}(\omega)} \left(\frac{\partial^2 v_2}{\partial r_1^2} + \frac{\partial^2 v_2}{\partial r_2^2} \right) - \frac{\partial p}{\partial r_2} , \tag{4.3}$$

$$\frac{\partial v_1}{\partial r_1} + \frac{\partial v_2}{\partial r_2} = 0 , \tag{4.4}$$

where $v_1 = v_1(\boldsymbol{r}, t; \omega)$ and $v_2 = v_2(\boldsymbol{r}, t; \omega)$ denote the r_1 and r_2 components of the non-dimensionalized velocity field $\boldsymbol{v}(\boldsymbol{r}, t; \omega)$ respectively [73]. The Reynolds number for the system $\mathrm{Re}(\omega)$ is left as an uncertain parameter to be inferred. $p = p(\boldsymbol{r}, t; \omega)$ appearing in (4.2) and (4.3) represents the non-dimensionalized pressure field, which, despite lacking an explicit time derivative term, can be solved for implicitly from the velocity field using the continuity equation (4.4).

SG_n — The geometries of the spatial domains for the five candidate models are distinct, with each domain featuring its own obstacle shape from the set depicted in Figure 4-8. The projected widths of all the obstacles are identical and are all 1

non-dimensional length unit. For purposes of numerical implementation, the spatial domain for each of the candidate models is discretized using a regular rectangular grid of 320 and 120 elements along the r_1 and r_2 directions respectively ($\Delta r_1 = \Delta r_2 = \frac{1}{20}$). This spatial discretization results in approximately 320×120 r_1-velocity and 320×120 r_2-velocity state variables for each of the candidate models, with slight variations due to the different surface areas of the five possible obstacle shapes. The augmented state vector for any of the candidate models is then

$$\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega) = \begin{bmatrix} \boldsymbol{v_1}(t;\omega) \\ \boldsymbol{v_2}(t;\omega) \\ \mathrm{Re}(\omega) \end{bmatrix} , \tag{4.5}$$

where $\boldsymbol{v_1}(t;\omega)$ and $\boldsymbol{v_2}(t;\omega)$ represent the vectors of r_1-velocity and r_2-velocity state variables respectively. Again, the Reynolds number of the system appears as an uncertain parameter. $N_{\boldsymbol{X}} + N_{\boldsymbol{\Theta}}$, the dimension of the augmented state vector, is thus of order 10^5.

BC_n — Stochastic Dirichlet boundary conditions are used for the left domain limits for all five candidate models

$$\boldsymbol{v} = (V_\infty(\omega), 0) = \left(\frac{\mathrm{Re}(\omega)}{\mathrm{Re}'}, 0 \right) \text{ for } r_1 = 0 , \tag{4.6}$$

where $V_\infty(\omega)$ denotes the stochastic non-dimensionalized inlet fluid velocity and

$$\mathrm{Re}' = \frac{V'L'}{\nu}$$

is a constant comprised of the non-dimensionalizing velocity and length constants V' and L' as well as the dynamic viscosity of the fluid ν. V', L', and ν are chosen to be 1, 1, and $\frac{1}{80}$ for this system respectively. Deterministic Neumann boundary conditions

are used for the top, bottom, and right domain limits

$$\frac{\partial \boldsymbol{v}}{\partial r_2} = 0 \text{ for } r_2 = 0 \text{ and } r_2 = 6 \ ,$$
$$\frac{\partial \boldsymbol{v}}{\partial r_1} = 0 \text{ for } r_1 = 16 \ .$$

Additional deterministic Dirichlet boundary conditions are used to impose no-slip conditions on the surfaces of the obstacles for all candidate models.

IC_n — For each of the five candidate models, an initial uniform distribution from 50 to 100 is assumed for the Reynolds number of the fluid flow. The mean of this distribution—Re = 75—corresponds to a mean non-dimensionalized inlet fluid velocity of $\frac{75}{80}$, as given by the relation in (4.6). Initial means for the fluid velocity in each of the candidate model spatial domains are found by numerically solving the continuity equation (4.4) with this mean inlet velocity and the other boundary conditions specified above. For each of the candidate models, a covariance matrix is then constructed for the velocity state variables that respects symmetry about the $r_2 = 3$ centerline using the boundary-mollified spatial covariance method suggested in [121]. Discrete Karhunen-Loève transforms are performed on these covariance matrices to initialize eight pairs of augmented modes and mode coefficients for each candidate model. The symmetry properties of this procedure lead to unbiased initialization of uncertainty in vortex shedding phase (i.e. no information regarding the phase of vortex shedding is assumed at the start), as exhibited in Figure 4-10. A ninth augmented mode is introduced for each candidate model that features only a single nonzero entry for the uncertain Reynolds number, for a total DO subspace dimension $N_{\mathrm{DO}} = 9$.

4.3.2 True Solution Generation

We generate two different true solutions for the stochastic flow past an obstacle system, the first—Experiment A1—featuring the circular obstacle depicted in Figure 4-8 and the second—Experiment A2—featuring the downstream-pointing triangular obstacle. For both solutions, Re is chosen to be 80. Deterministic initial velocity fields

satisfying the continuity equation (4.4) are constructed for both, with arbitrary asymmetries introduced to induce vortex shedding. Deterministic evolution of the velocity fields is performed by numerically solving the Navier-Stokes equations (4.2)–(4.4) using a finite-volume approach employing a regular rectangular grid of 320 and 120 elements along the r_1 and r_2 directions respectively and a non-dimensional time-step of $\Delta t = \frac{1}{120}$ up to a final non-dimensional time $T = 50$. These deterministic solutions represent two possible realizations of the stochastic flow past an obstacle system, each with a true obstacle shape, true Reynolds number, and true dynamical velocity field that are to be inferred. Each of these two solutions constitutes an independent experiment conducted with this system and results for both will be described in the following section.

4.3.3 Observations and Inference

Every 1 non-dimensional time unit (i.e. every 120 numerical time-steps), noisy observations of both r_1-velocity and r_2-velocity are made at the nine locations indicated in Figure 4-4, for a total of $N_{\mathbf{Y}} = 18$ observations at every observation time. A single non-dimensional time unit for this system corresponds to a phase shift of approximately $\frac{5\pi}{13}$ when $\mathrm{Re} = 80$. Unbiased Gaussian noise is applied to the observations with a standard deviation equal to approximately 10% of the mean fluid velocity, which are then used by the GMM-DO filter to perform joint model-conditional inference of the velocity state variables and the Reynolds number for each of the candidate models. The linear observation matrix $\mathbf{H}$ appearing in (3.16) is specified so as to identify the velocity state variables corresponding to the nine observation locations while the matrix of observation covariances $\mathbf{R}$ is specified as a diagonal matrix with diagonal elements equal to the variance of the observation noise.

Following the procedure outlined in [129], the Bayesian information criterion (BIC), in conjunction with the expectation-maximization (EM) algorithm for GMMs, is used to select the optimal number of GMM components N_{GMM} at every observation time. Typical BIC-optimized values for N_{GMM} were found to lie between 20 and 60 for experiments conducted with the stochastic flow past an obstacle system. Marginal

likelihoods for the candidate models are calculated using (3.31) following every iteration of the GMM-DO filter and the model distribution is updated according to (1.9).

4.3.4 Numerical Method

Numerical evolution of the augmented state vector mean, augmented modes, and mode coefficients for each of the candidate models is performed using the DO evolution equations (3.4)–(3.6) and the Navier-Stokes equations (4.2)–(4.4) with stochastic boundary conditions (4.6) according to the finite-volume methodology developed in [140]. A non-dimensional time-step of $\Delta t = \frac{1}{120}$ is used. $N_{\text{MC}} = 10^4$ Monte Carlo samples are used for the stochastic evolution of the mode coefficients.

For numerical robustness, the logarithms of the marginal likelihoods calculated using (3.31) are used to perform model inference rather than the marginal likelihoods themselves, which are typically of exceptionally small magnitude ($< 10^{-12}$) for high-dimensional systems such as the stochastic flow past an obstacle system considered here. These logarithms can be used directly to perform Bayesian update of the model

distribution, as can be seen by rearranging the terms in (1.9)

$$p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n|\boldsymbol{y}) = \frac{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\, p_{\mathcal{M}}(\mathcal{M}_n)$$

$$= \frac{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n)}{\sum_{m=1}^{N_{\mathcal{M}}} p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_m)\, p_{\mathcal{M}}(\mathcal{M}_m)}$$

$$= \frac{p_{\mathcal{M}}(\mathcal{M}_n)}{\sum_{m=1}^{N_{\mathcal{M}}} \frac{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_m)}{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)} p_{\mathcal{M}}(\mathcal{M}_m)}$$

$$= \frac{p_{\mathcal{M}}(\mathcal{M}_n)}{\sum_{m=1}^{N_{\mathcal{M}}} \exp\Big[\log p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_m) - \log p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_n)\Big] p_{\mathcal{M}}(\mathcal{M}_m)}$$

$$\forall n \in \{1, \ldots, N_{\mathcal{M}}\}\,. \qquad (4.7)$$

Note that the argument of the exponential in (4.7) is the logarithm of the Bayes factor [72] between the mth and nth candidate models.

A summary of the experimental properties used for the stochastic flow past an obstacle system is provided in Table 4.1.

4.4 Results

Three metrics are used to evaluate the success of the Bayesian model inference methodology as applied to the stochastic flow past an obstacle system. The first metric is the root mean square error (RMSE) between the velocity state variables given by the deterministic simulation from which observations are made—i.e. the true velocity state variables—and the Bayesian mean of the candidate model velocity state variables, as determined by (1.7). A RMSE approaching 0 indicates successful state inference. The second metric is the integral of the Bayesian probability

Table 4.1: Experimental properties for the stochastic flow past an obstacle system.

Property	Value
$N_{\mathcal{M}}$	5
$N_{\boldsymbol{X}}$	$\sim 75,000$
$N_{\boldsymbol{\Theta}}$	1
$N_{\boldsymbol{Y}}$	18
N_{DO}	9
N_{MC}	10^4
Δr_1	1/20
Δr_2	1/20
Δt	1/120
T	50

distribution for the Reynolds number of the system from 77.5 to 82.5, a range that is centered on the Reynolds number of the deterministic simulations—i.e. the true Reynolds number—and encompasses only 10% of the total initial prior range assumed for the Reynolds number. This metric is referred to as the true Reynolds number probability and is calculated by applying the marginalization (1.6) to the augmented state vector (4.5). A true Reynolds number probability approaching 1 indicates successful parameter inference. The third metric is simply the probability of the model of the deterministic simulation from which observations are made—i.e. the true model—within the model distribution. A true model probability approaching 1 indicates successful model inference (since in our example, the true model is among the candidate models). This 'probability of the model' metric easily extends to cases where the truth is not part of the candidate models (e.g. is a linear combination) or is a probabilistic combination of several candidate models.

4.4.1 Experiment A1: Circular Obstacle

The first experiment we consider, Experiment A1, is that of a simulated truth defined by the deterministic simulation featuring the circular obstacle. The goal is to jointly infer the velocity state variables, Reynolds number, and obstacle shape when noisy

observations are made from this simulated truth. The evolution of this inference is illustrated in Figures 4-9 to 4-18.

Figure 4-9 and Figure 4-10 illustrate the system at the initial non-dimensional time ($t = 0$), before any observations are made. The fluid velocity field of the deterministic simulation, which was chosen for the true velocity field of the system, is in a time-periodic state. Regular asymmetric vortex shedding is occurring downstream of the obstacle, with a period of approximately 5.4 non-dimensional time units. Each of the five model-conditional velocity field means (i.e. $\overline{\boldsymbol{x}}$ in (3.1)) for the five candidate models is symmetric about the $r_2 = 3$ centerline, indicating no bias towards any particular vortex shedding phase. The Bayesian mean of these five model-conditional velocity fields (i.e. the mean of $p_{\boldsymbol{X}}(\bullet)$ in (1.5) or equivalently the mean of the five $\overline{\boldsymbol{x}}$ for the five candidate models) is consequently also symmetric. The model-conditional Reynolds number distributions (found using the marginalization represented by (1.6)) are uniform, as is the model distribution (i.e. $p_{\mathcal{M}}(\bullet)$ in (1.5)). The model-conditional velocity field standard deviations are also all symmetric about the $r_2 = 3$ centerline.

Figure 4-11 and Figure 4-12 illustrate the system after 1 non-dimensional time unit and 1 observation episode (i.e. one use of the GMM-DO filter). The true velocity field remains in a time-periodic state. The model-conditional velocity field means for the five candidate models (each mean estimated using the GMM-DO filter) have shifted slightly from symmetry, as have the model-conditional velocity field standard deviations. Standard deviation magnitudes have also decreased for all candidate models as a result of inference. The model-conditional Reynolds number distributions have shifted slightly from uniform, as has the model distribution.

Figure 4-13 and Figure 4-14 illustrate the system after 10 non-dimensional time units and 10 observation episodes. All five of the model-conditional velocity field means are beginning to align in phase with the true velocity field. Consequently, the Bayesian mean of these conditional velocity field means is also beginning to align in phase with the true field, though the energy of the vortices in this Bayesian mean is still lower than that of those in the true field. The Reynolds number distribution is beginning to shift towards the true Reynolds number of 80 and the probability

of one of the incorrect candidate models, the one featuring the downstream-pointing triangular obstacle, has fallen quickly.

Figure 4-15 and Figure 4-16 illustrate the system after 30 non-dimensional time units and 30 observation episodes. Phase alignments between the five conditional velocity field means and the true field all appear to have improved significantly. Since the five obstacle shapes produce different vortex shedding frequencies for the same Reynolds number however (see Figure 4-7), the five model-conditional Reynolds number distributions have begun to concentrate around different values. For the candidate model featuring the circular obstacle, the conditional Reynolds number distribution is simply concentrating around 80, the true Reynolds number. For the other candidates though, the conditional distributions are concentrating around values that allow the candidates to match the shedding frequency of the true field despite featuring incorrect obstacle shapes. This leads to the multimodality exhibited in the Reynolds number distribution. The candidate model featuring the circular obstacle is now also clearly favored by the model distribution.

Figure 4-17 and Figure 4-18 illustrate the system after 50 non-dimensional time units and 50 observation episodes. The Bayesian velocity field mean is now consistent with the true velocity field. The Reynolds number distribution is concentrated around the true Reynolds number of 80, with slight multimodality still present. Note that had an incorrect model been assumed for the system instead of properly accounting for model uncertainty, the inferred Reynolds number distribution at this point would correspond entirely to one of the smaller modes constituting the multimodal Reynolds number distribution, all of which assign infinitesimal probability density to the true Reynolds number of the system. The candidate model featuring the circular obstacle is now also the only plausible model for the system, as indicated by the highly peaked model distribution. The model-conditional velocity field standard deviations for all five candidate models have decreased significantly, with the greatest decreases exhibited by the circular obstacle (the true model) and the square obstacle (the candidate model 'closest' to the circle, as seen in Figure 4-7).

Figure 4-19 illustrates the time progression of the three inference metrics intro-

duced above. The RMSE approaches 0, the true Reynolds number probability approaches 1, and the true model probability approaches 1, all indicating successful inference.

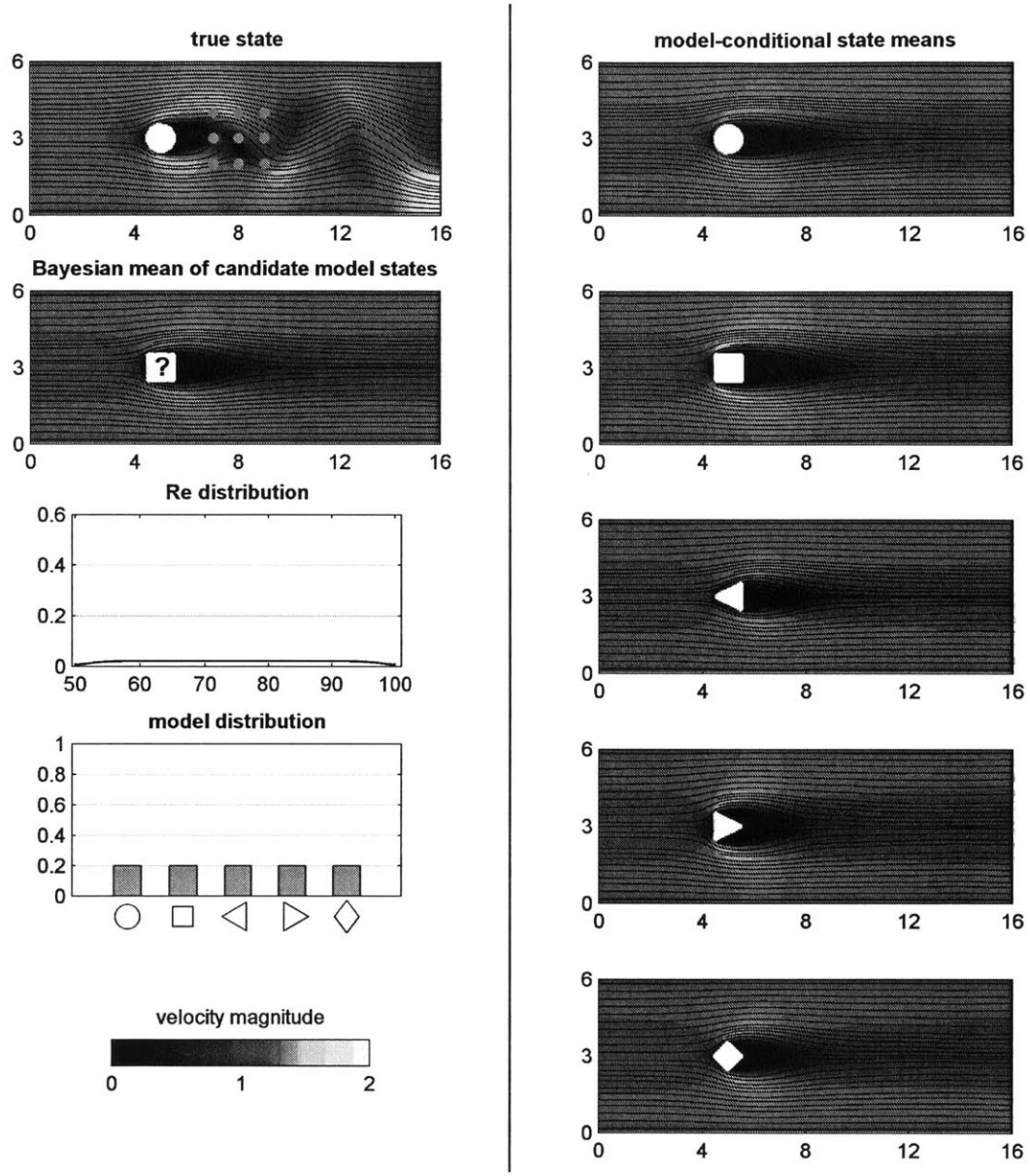

Figure 4-9: The stochastic flow past an obstacle system with observations made from the deterministic simulated truth featuring the circular obstacle (Experiment A1) at non-dimensional time $t = 0$. Observation locations are indicated on the true state field in light blue. The bounding area for the obstacle in the Bayesian mean of the candidate model states is labeled with a question mark since the obstacle shape is uncertain. Model-conditional velocity field means for each of the five candidate models are displayed on the right.

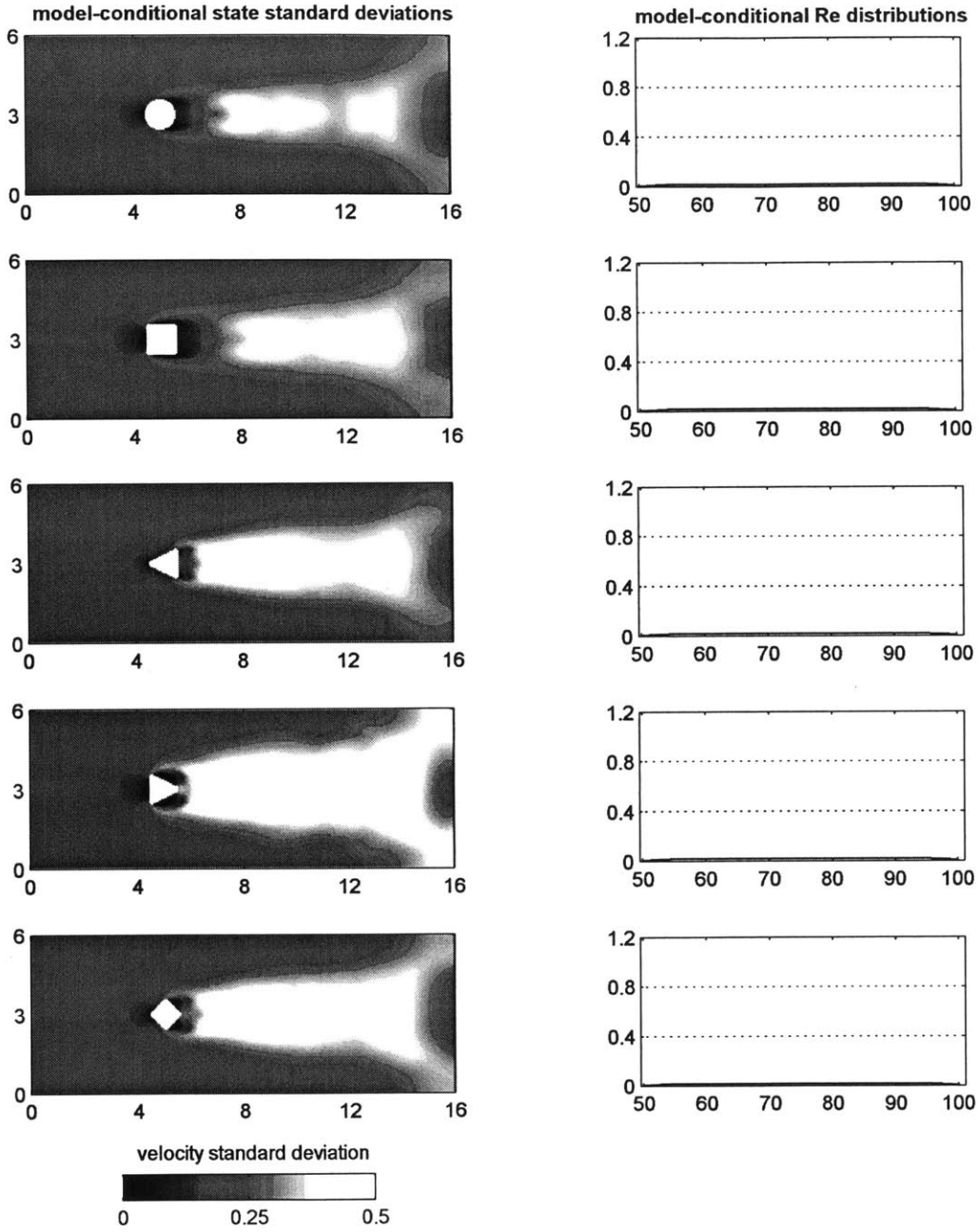

Figure 4-10: As Figure 4-9 but showing model-conditional uncertainties at non-dimensional time $t = 0$. Specifically, model-conditional velocity field standard deviations for each of the five candidate models are displayed on the left. Model-conditional Reynolds number distributions for each of the five candidate models are displayed on the right.

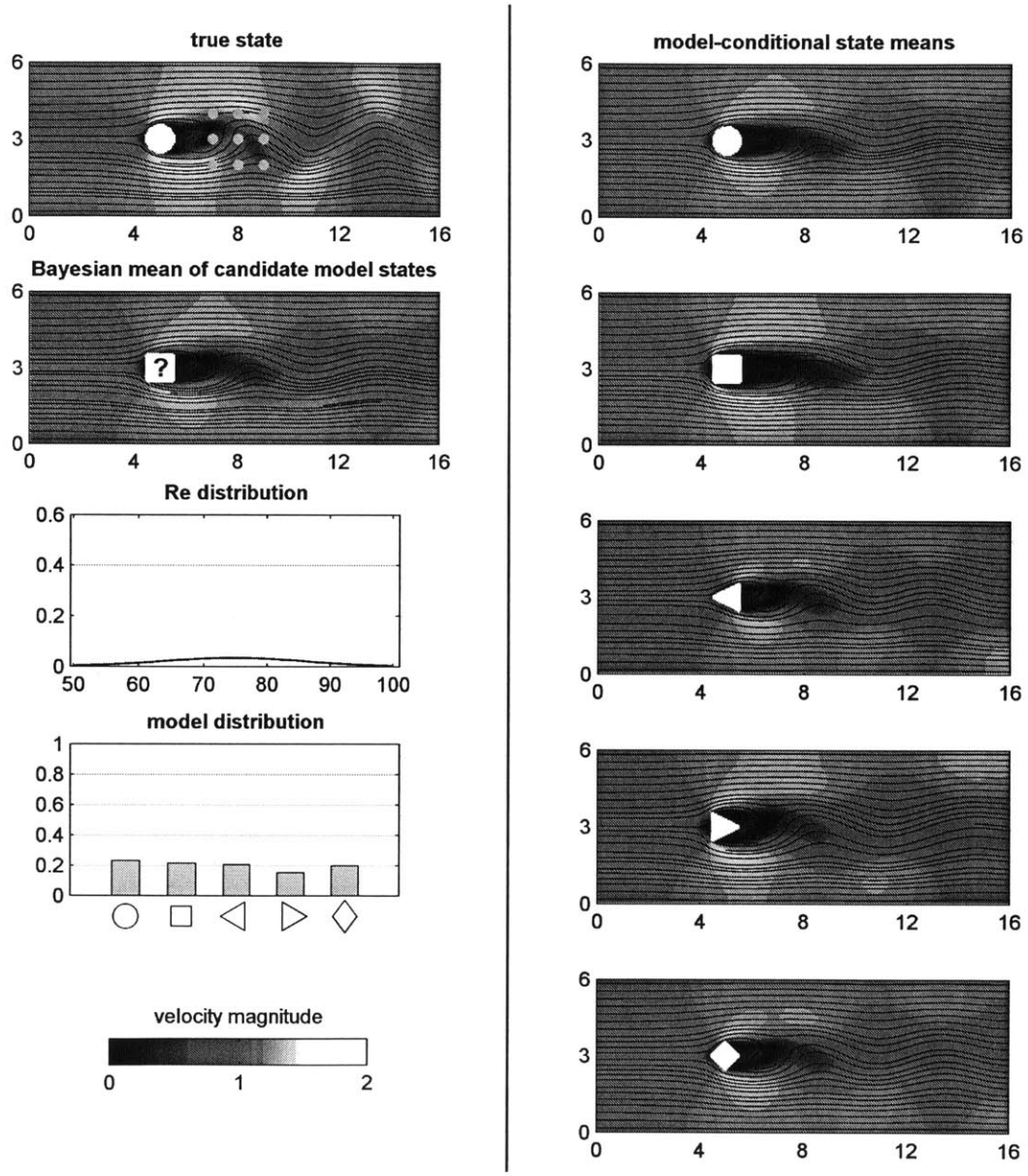

Figure 4-11: As Figure 4-9 but at non-dimensional time $t = 1$ (i.e. after 1 observation episode).

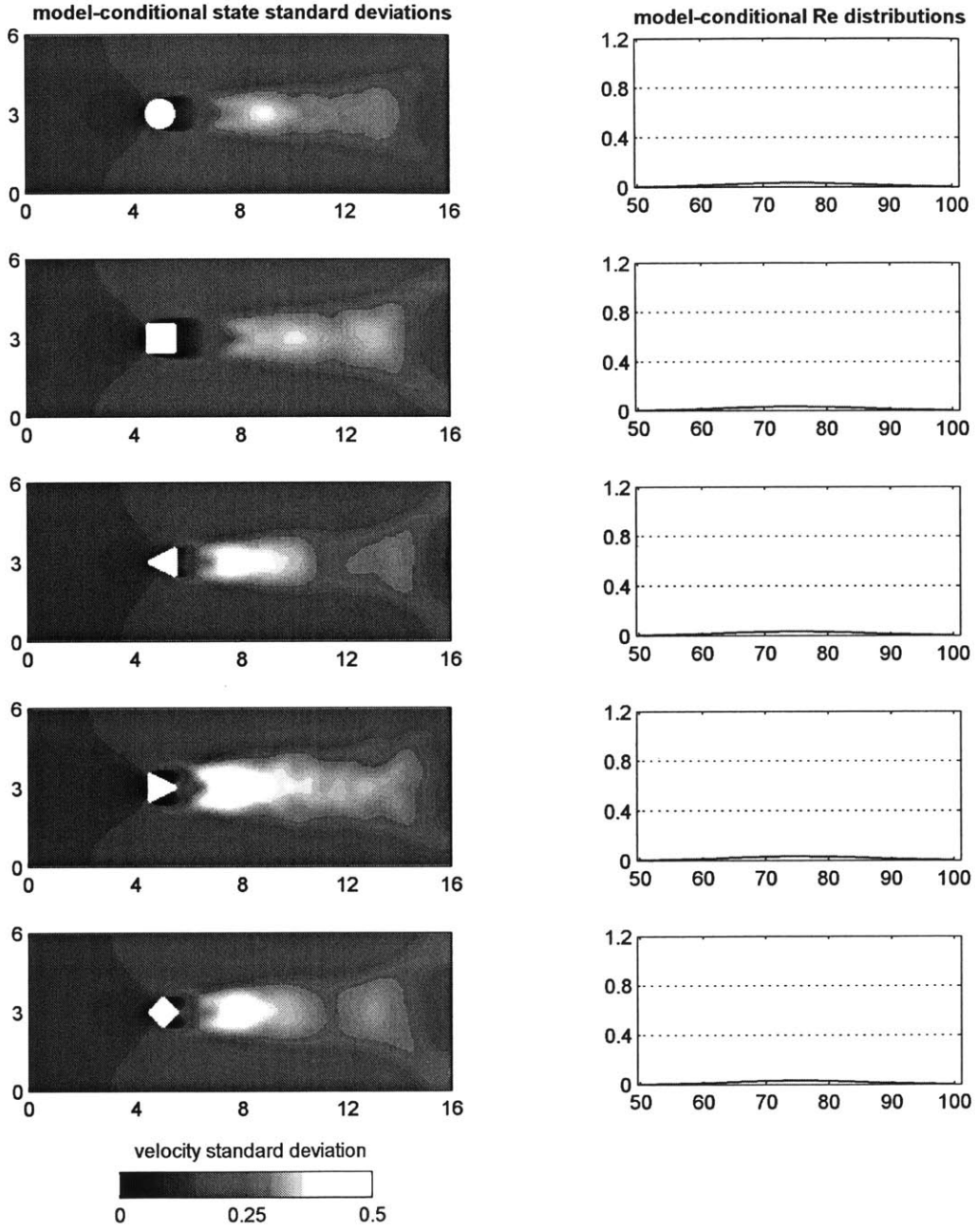

Figure 4-12: As Figure 4-10 but at non-dimensional time $t = 1$ (i.e. after 1 observation episode).

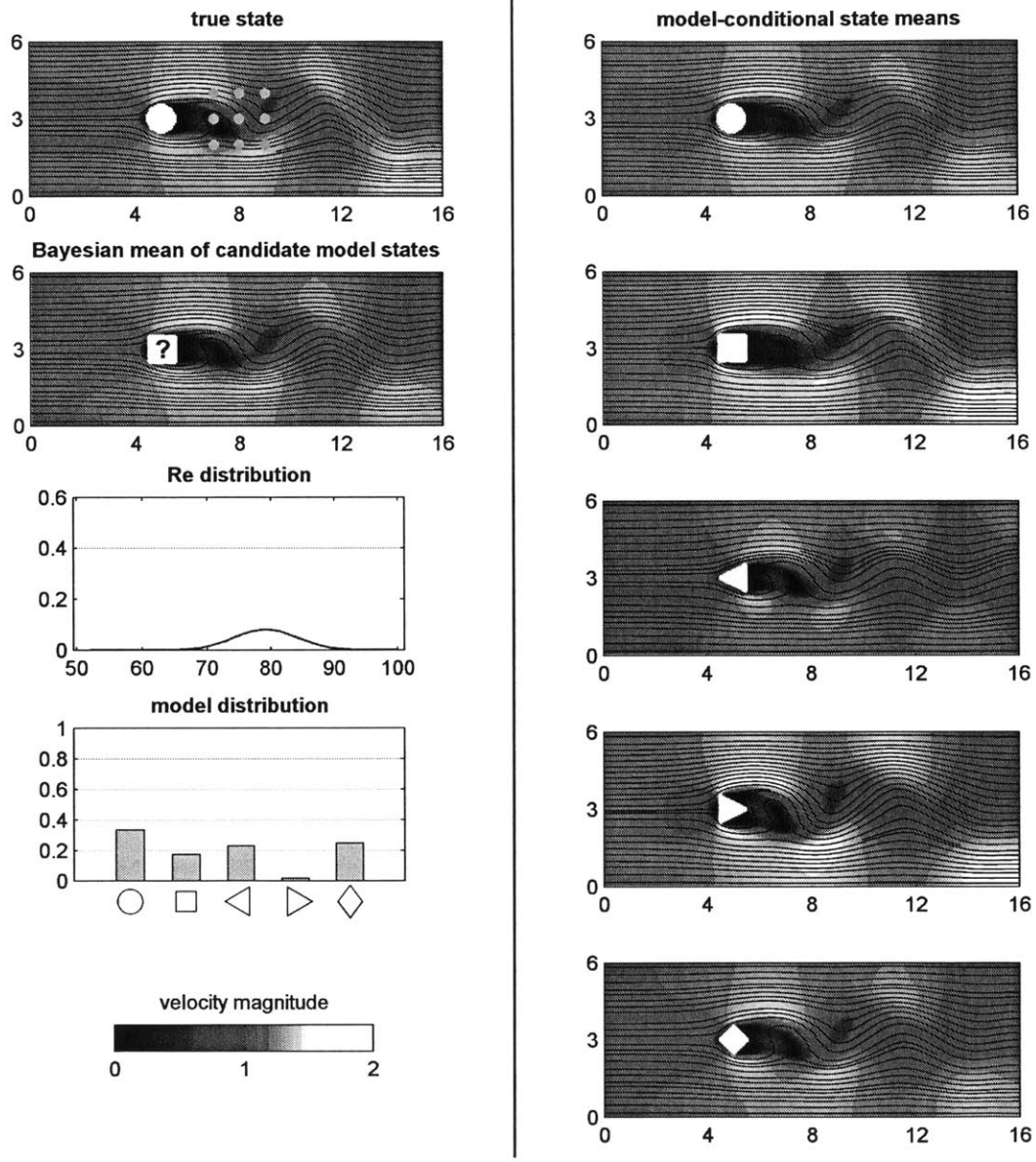

Figure 4-13: As Figure 4-9 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

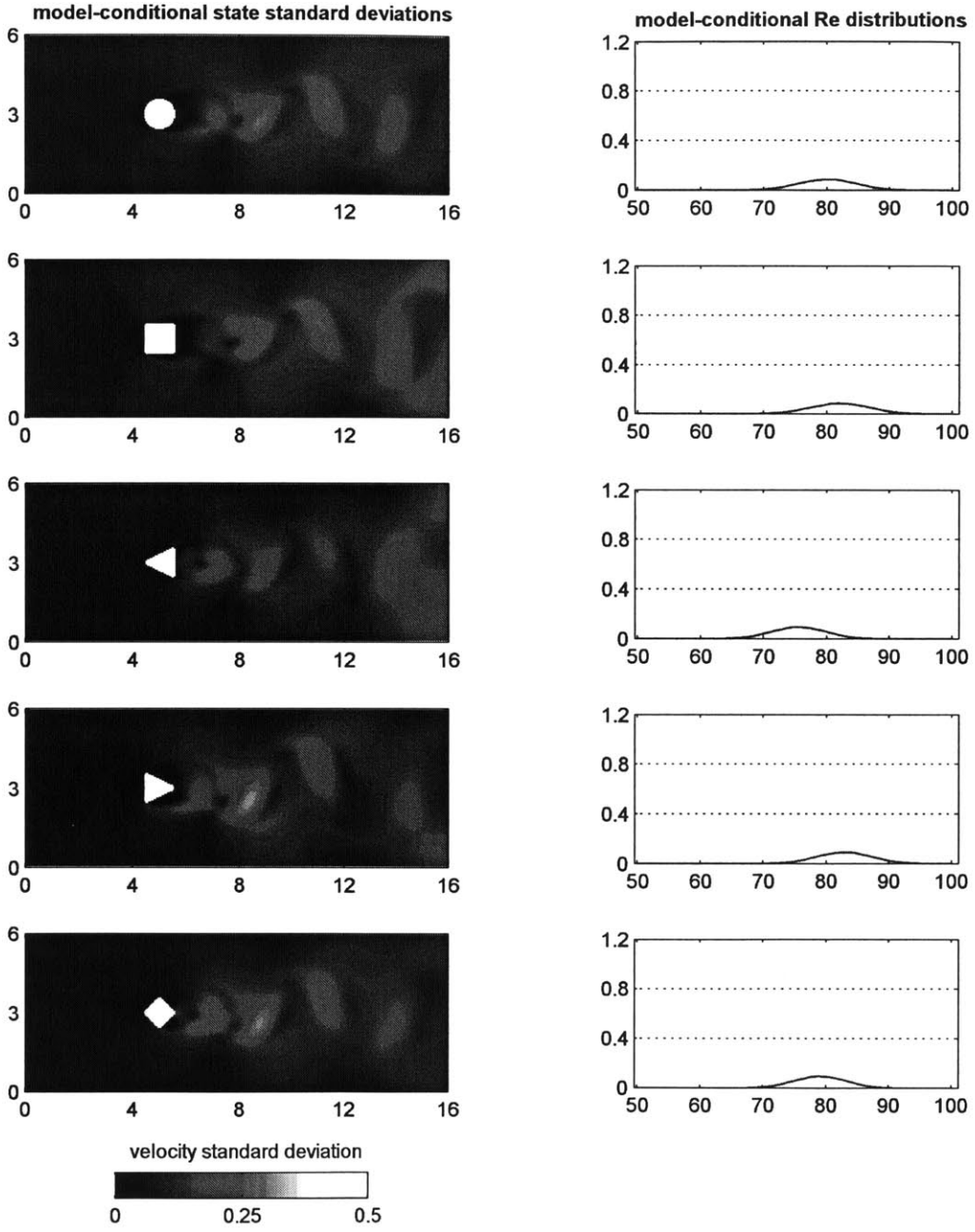

Figure 4-14: As Figure 4-10 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

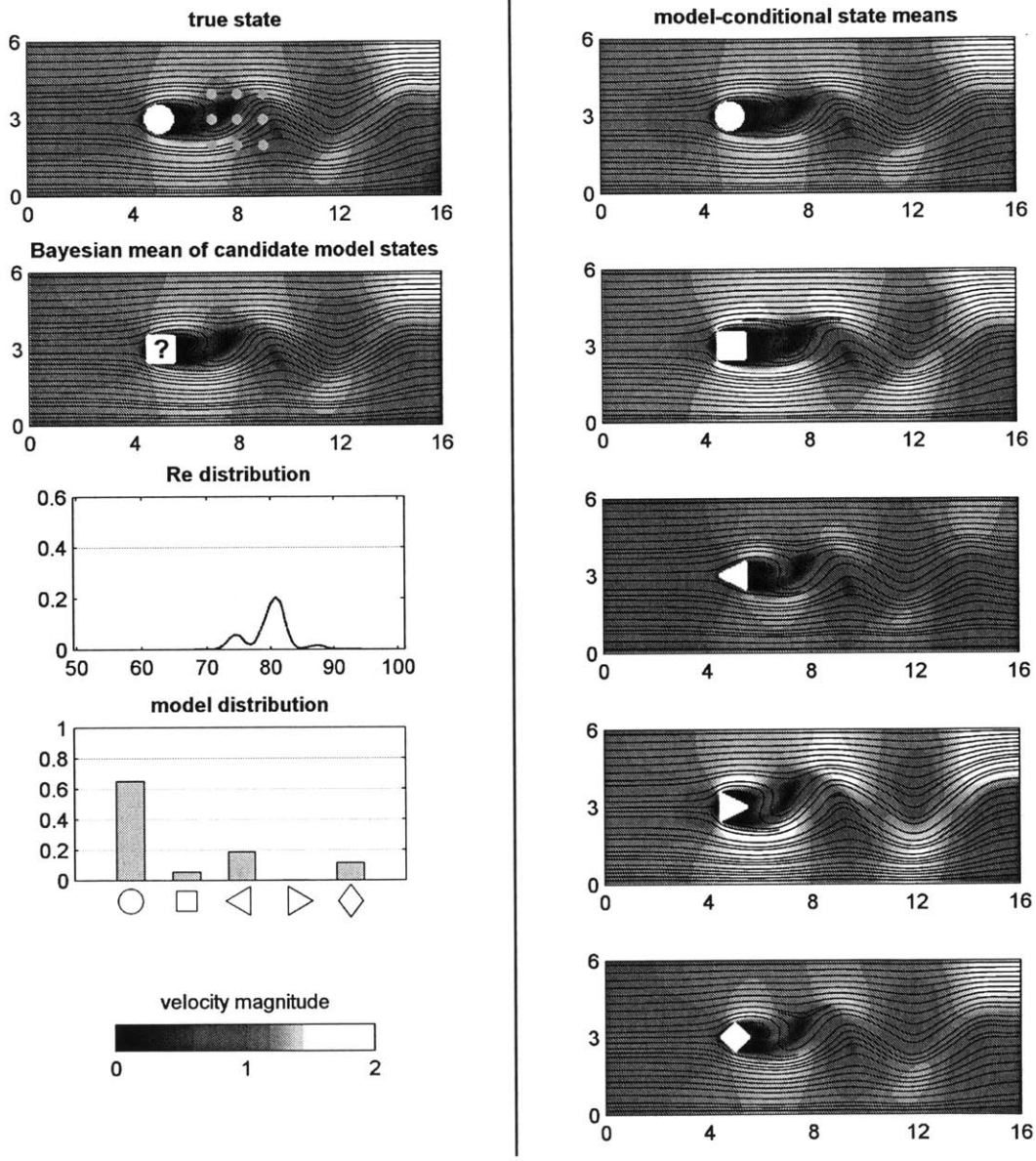

Figure 4-15: As Figure 4-9 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

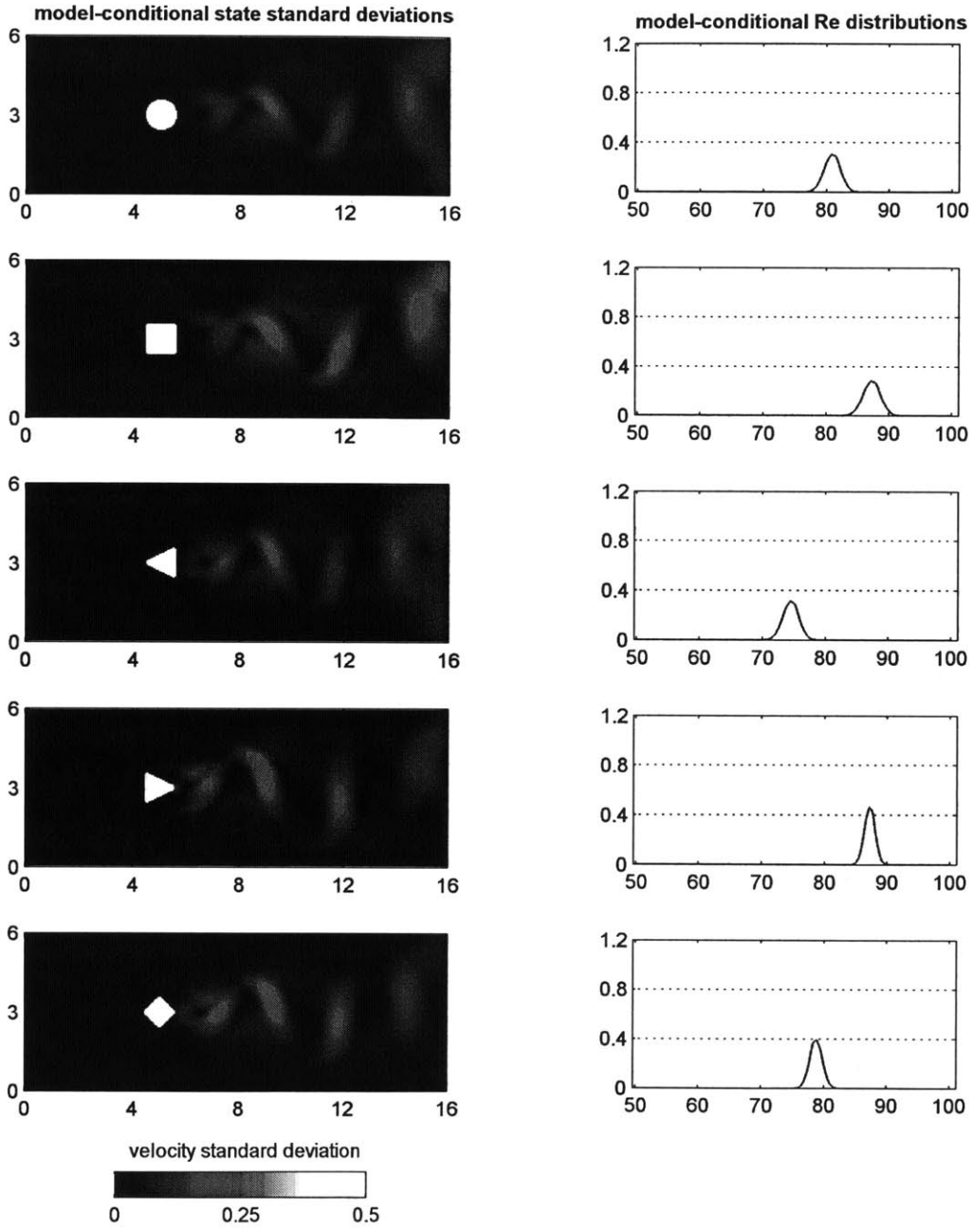

Figure 4-16: As Figure 4-10 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

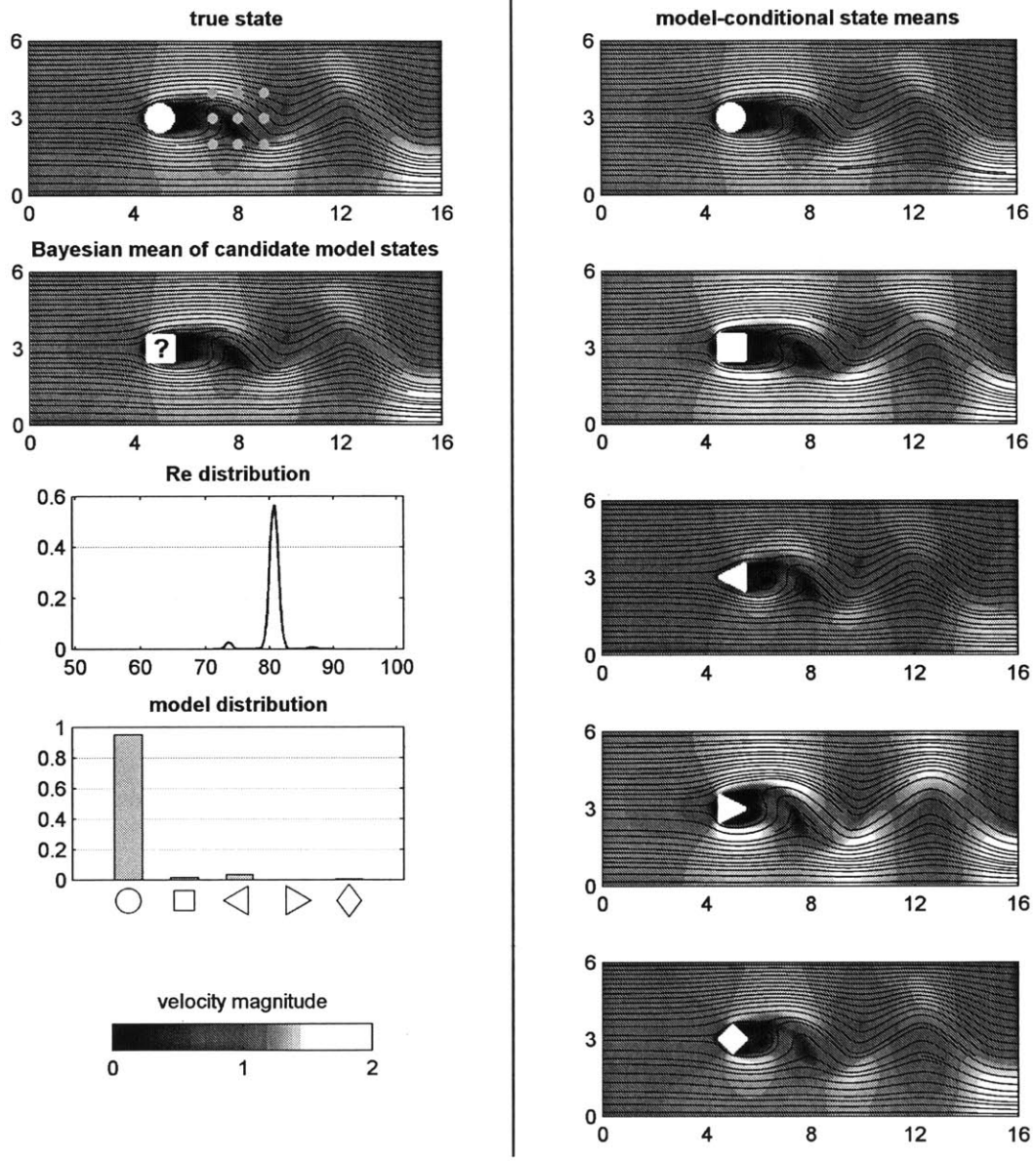

Figure 4-17: As Figure 4-9 but at non-dimensional time $t = 50$ (i.e. after 50 observation episodes).

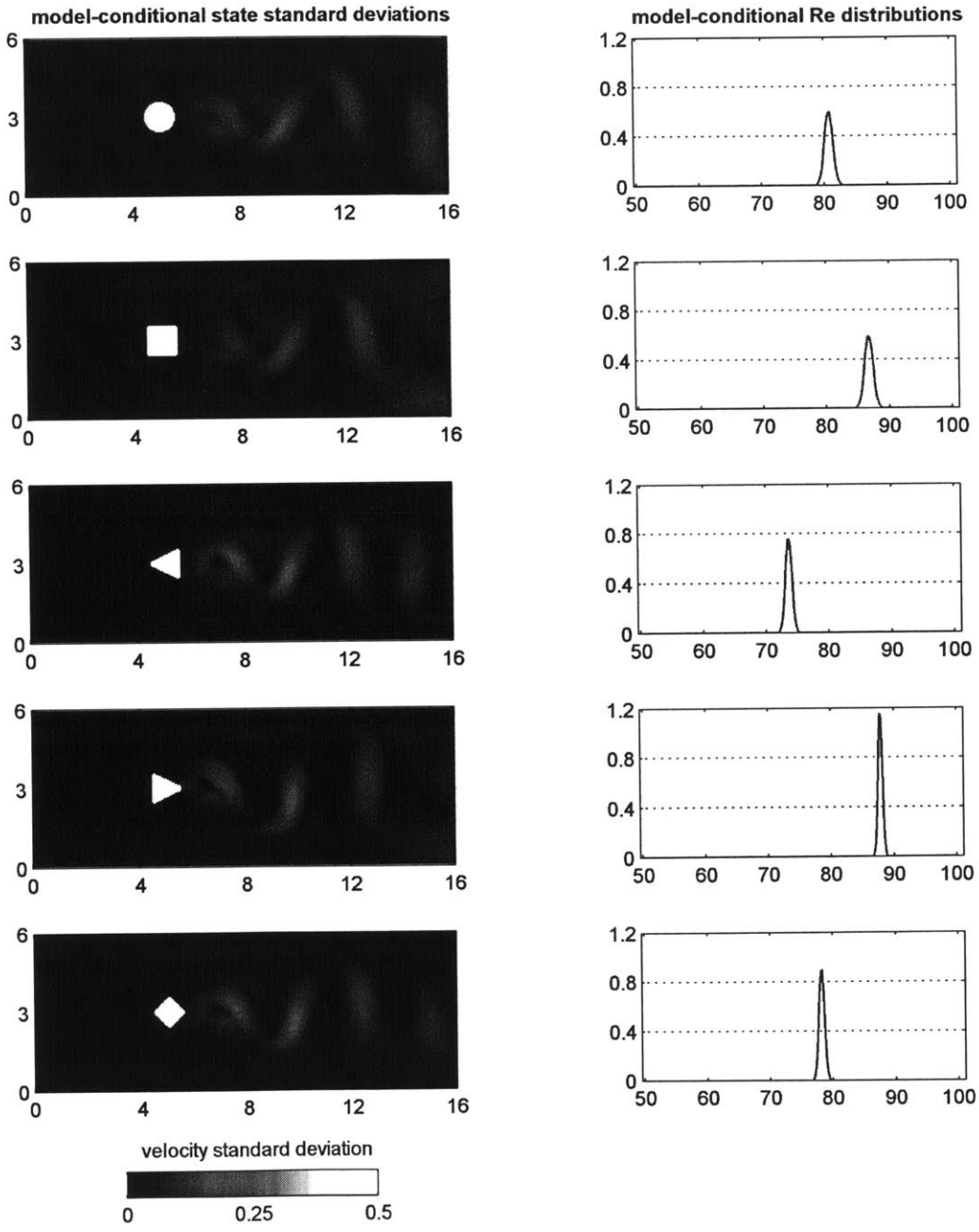

Figure 4-18: As Figure 4-10 but at non-dimensional time $t = 50$ (i.e. after 50 observation episodes).

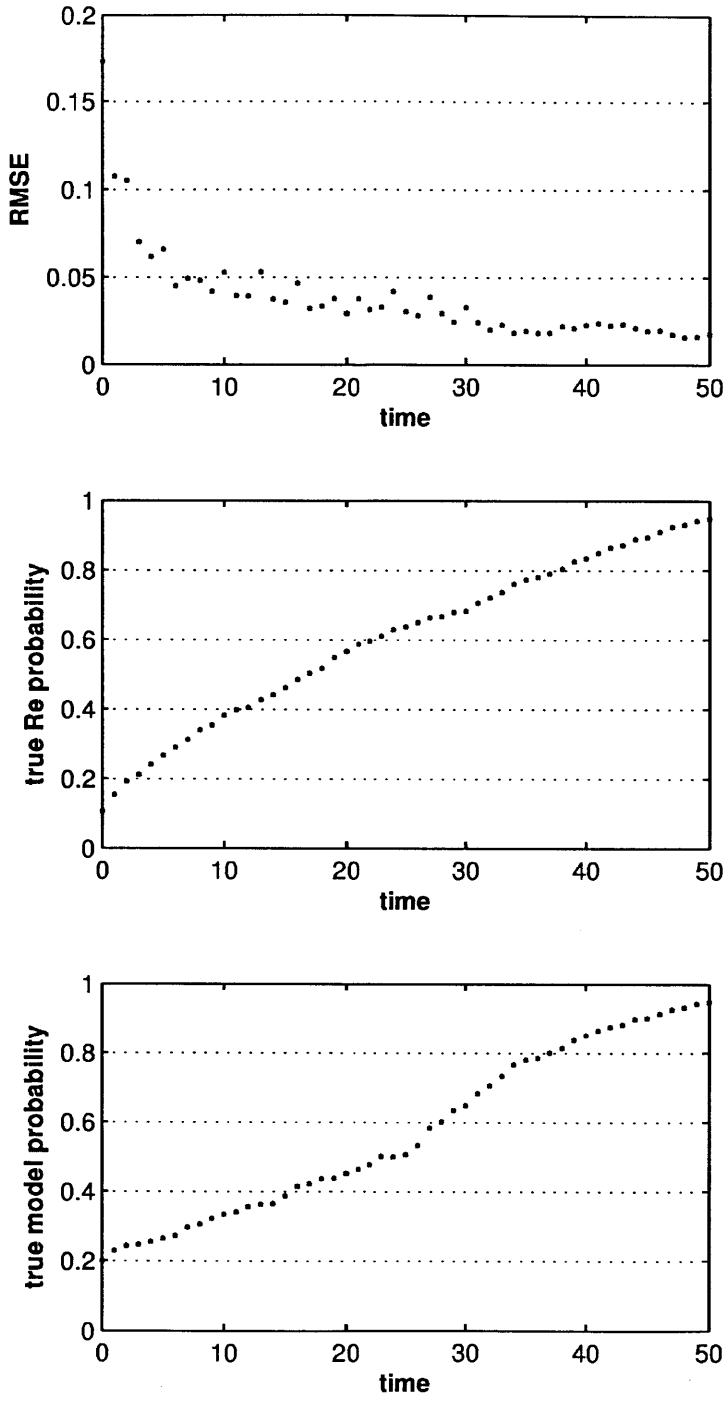

Figure 4-19: Time progression of inference metrics for the stochastic flow past an obstacle system (as defined at the beginning of Section 4.4) with observations made from the deterministic simulated truth featuring the circular obstacle (Experiment A1). Metrics are plotted at every 1 non-dimensional time unit immediately following observation and inference (i.e. posterior values from the GMM-DO filter).

4.4.2 Experiment A2: Downstream-Pointing Triangular Obstacle

The second experiment we consider, Experiment A2, is that of a simulated truth defined by the deterministic simulation featuring the downstream-pointing triangular obstacle. The goal is again to jointly infer the velocity state, Reynolds number, and obstacle shape when noisy observations are made from this simulated truth. The evolution of this inference is illustrated in Figures 4-20 to 4-23.

Figure 4-20 illustrates the system at the initial non-dimensional time ($t = 0$), before any observations are made. The true velocity field is again in a time-periodic state and the conditional velocity field means for all the candidate models are symmetric about the $r_2 = 3$ centerline. The Reynolds number distribution and model distribution are uniform.

Figure 4-21 illustrates the system after 5 non-dimensional time units and 5 observation episodes. The conditional velocity field means are beginning to align in phase with the true field and the Reynolds number distribution is beginning to shift towards the true Reynolds number of 80. The probabilities of the three candidate models featuring the circular obstacle, the square obstacle, and the upstream-pointing triangular obstacle have all fallen quickly.

Figure 4-22 illustrates the system after 10 non-dimensional time units and 10 observation episodes. All the candidate models other than the one featuring the downstream-pointing triangular obstacle are struggling to match the vortex shedding pattern of the true field. The Reynolds number distribution is concentrating around the true Reynolds number of 80 and the candidate model featuring the downstream-pointing triangular obstacle is now clearly favored by the model distribution.

Figure 4-23 illustrates the system after 30 non-dimensional time units and 30 observation episodes. The Bayesian velocity field mean is now consistent with the true velocity field, the Reynolds number distribution is concentrated around the true Reynolds number of 80, and the candidate model featuring the downstream-pointing triangular obstacle is now the only plausible model, as indicated by the highly peaked

model distribution.

Figure 4-24 illustrates the time progression of the three inference metrics introduced at the beginning of Section 4.4. Again, the RMSE approaches 0, the true Reynolds number probability approaches 1, and the true model probability approaches 1, all indicating successful inference.

Remarks

First, it is interesting to note that model inference proceeds much more quickly for the deterministic simulation featuring the downstream-pointing triangular obstacle (Experiment A2) than for the simulation featuring the circular obstacle (Experiment A1), suggesting that the downstream-pointing triangular obstacle is more distinct amongst the set of five possible obstacles than the circular obstacle. Of course, this speed of inference result depends on the dynamics, uncertainty, and observation locations selected. Second, this exemplifies how the methodology developed in Chapter 3 for Bayesian model inference can also be used as a statistical determinant of the similarity or dissimilarity between various stochastic dynamical models. For example, we can say that for the given Reynolds number range and initial state uncertainty considered, the diamond obstacle is the 'closest' to the downstream-pointing triangular obstacle. Similar remarks could have been made for Experiment A1: for example, that experiment indicated which obstacles were closest/furthest to a circle for the given Reynolds number range, initial uncertainty, and observation locations.

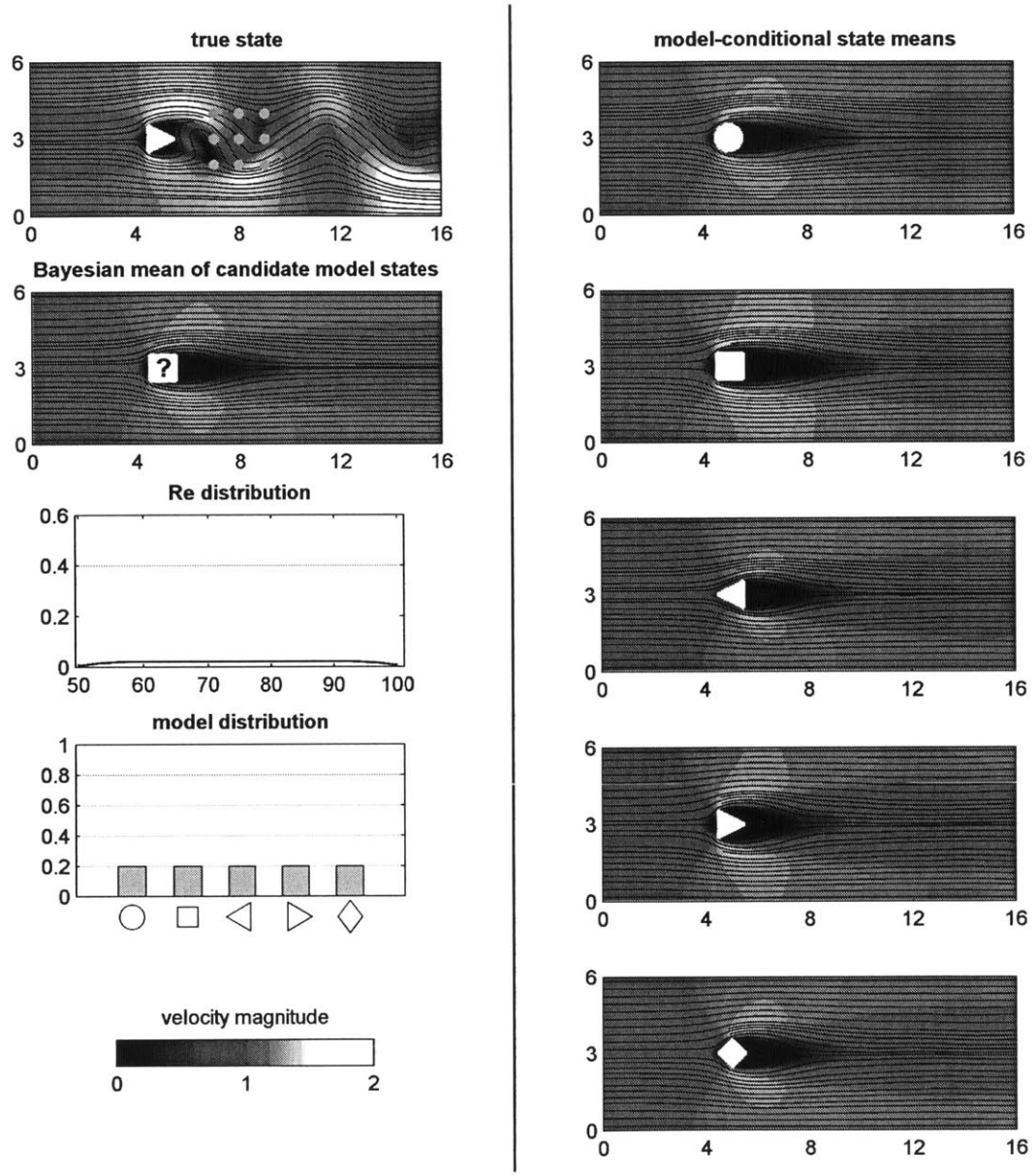

Figure 4-20: As Figure 4-9 but for the second experiment featuring the downstream-pointing triangular obstacle (Experiment A2).

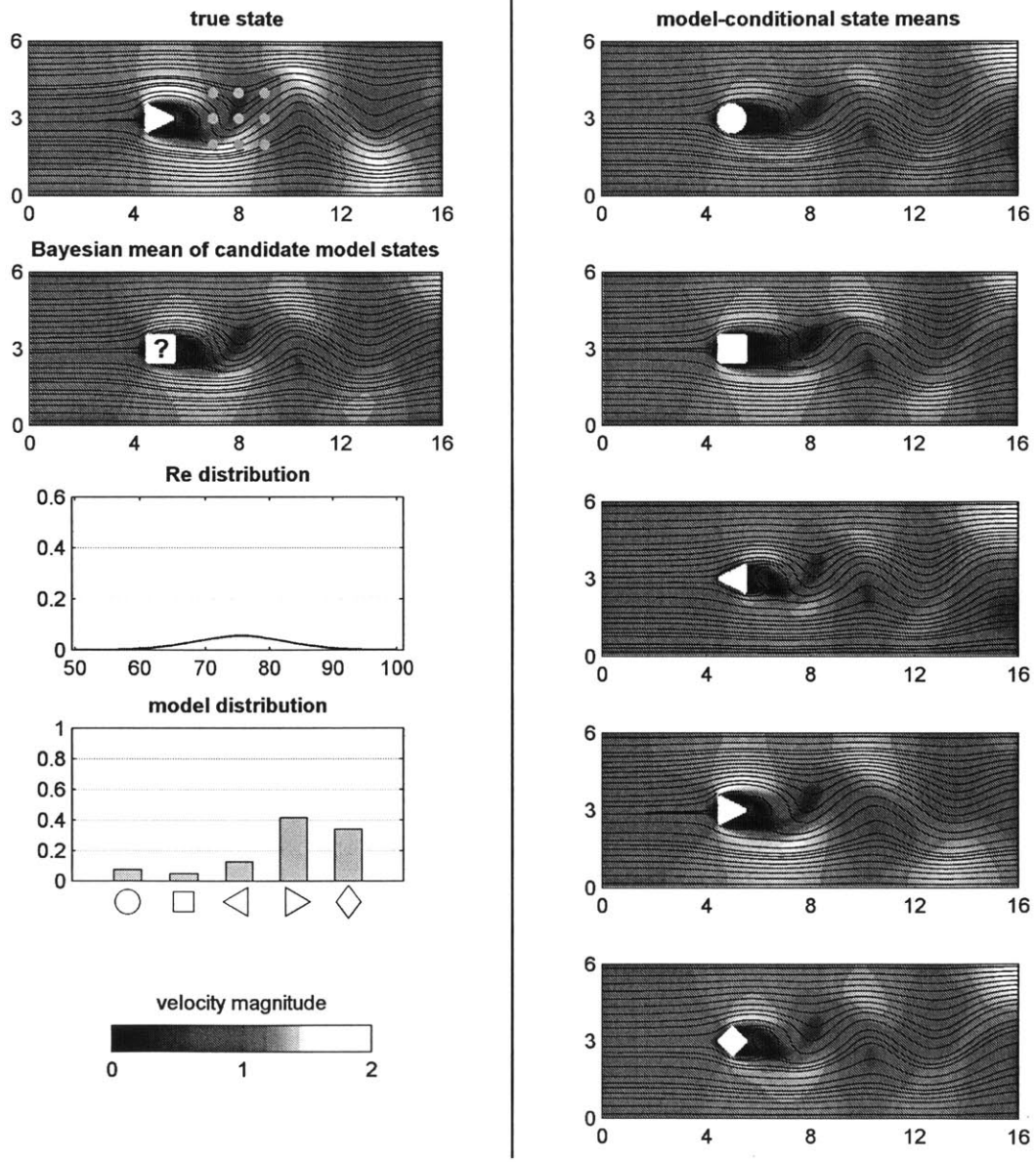

Figure 4-21: As Figure 4-20 but at non-dimensional time $t = 5$ (i.e. after 5 observation episodes).

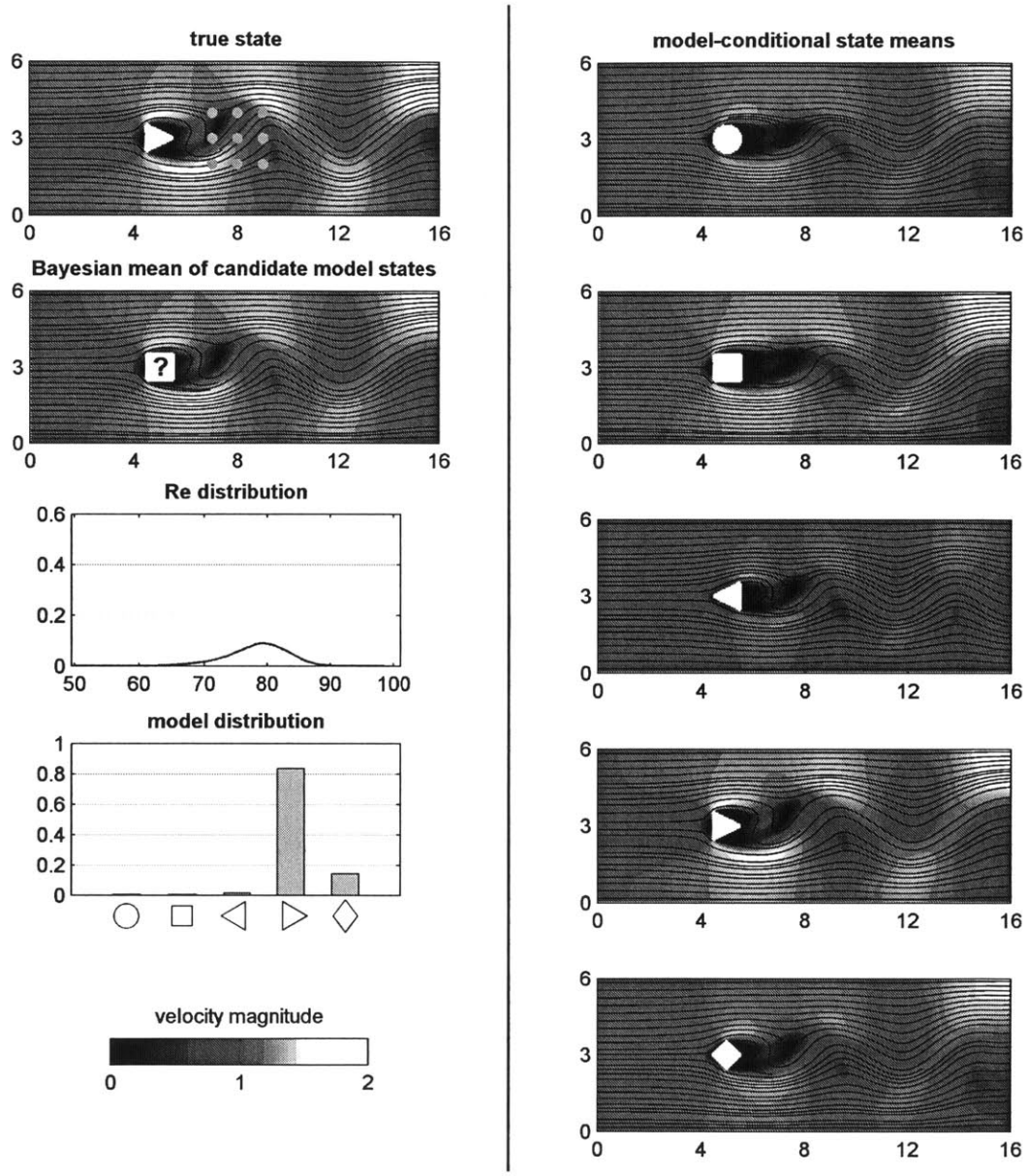

Figure 4-22: As Figure 4-20 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

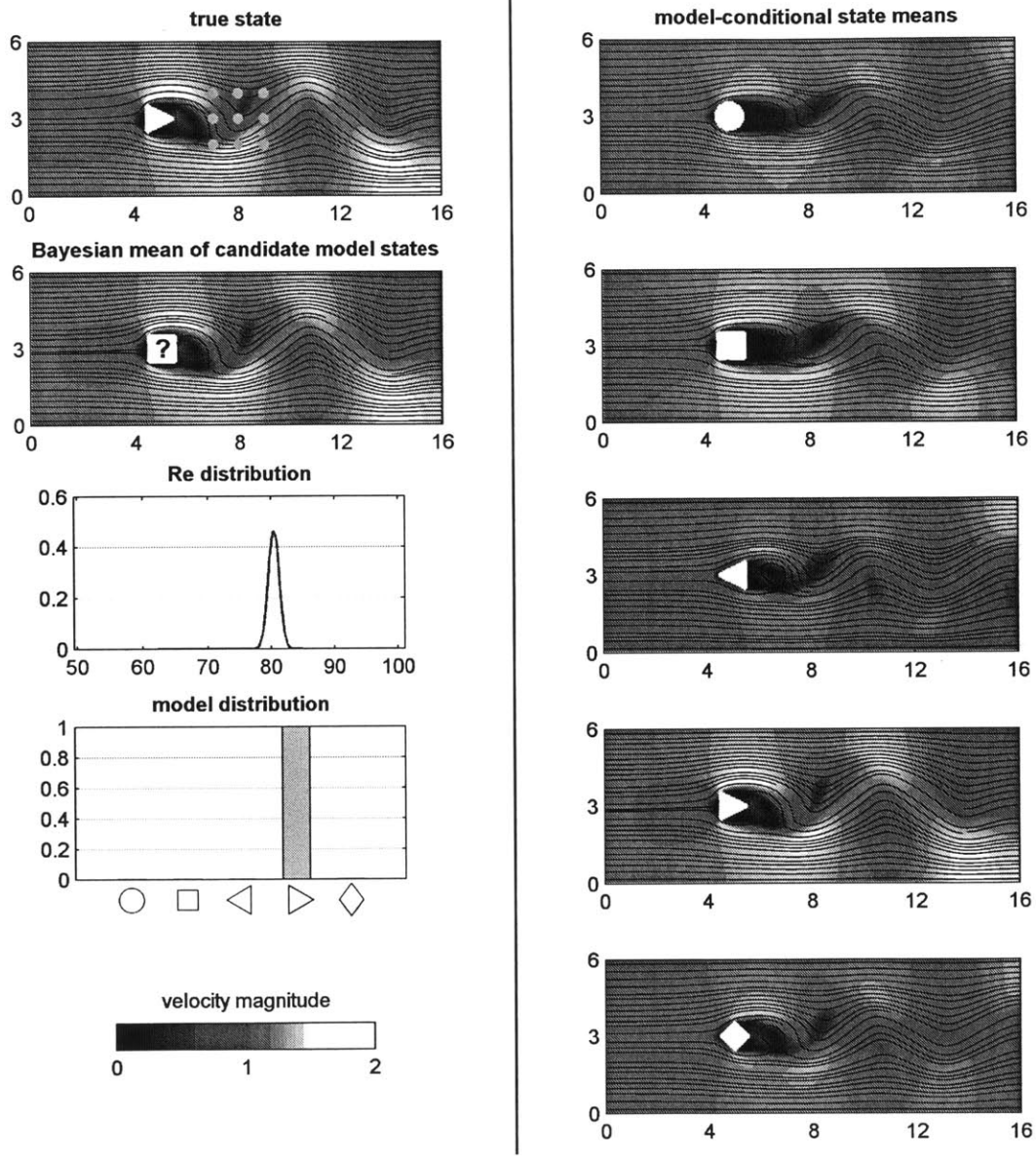

Figure 4-23: As Figure 4-20 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

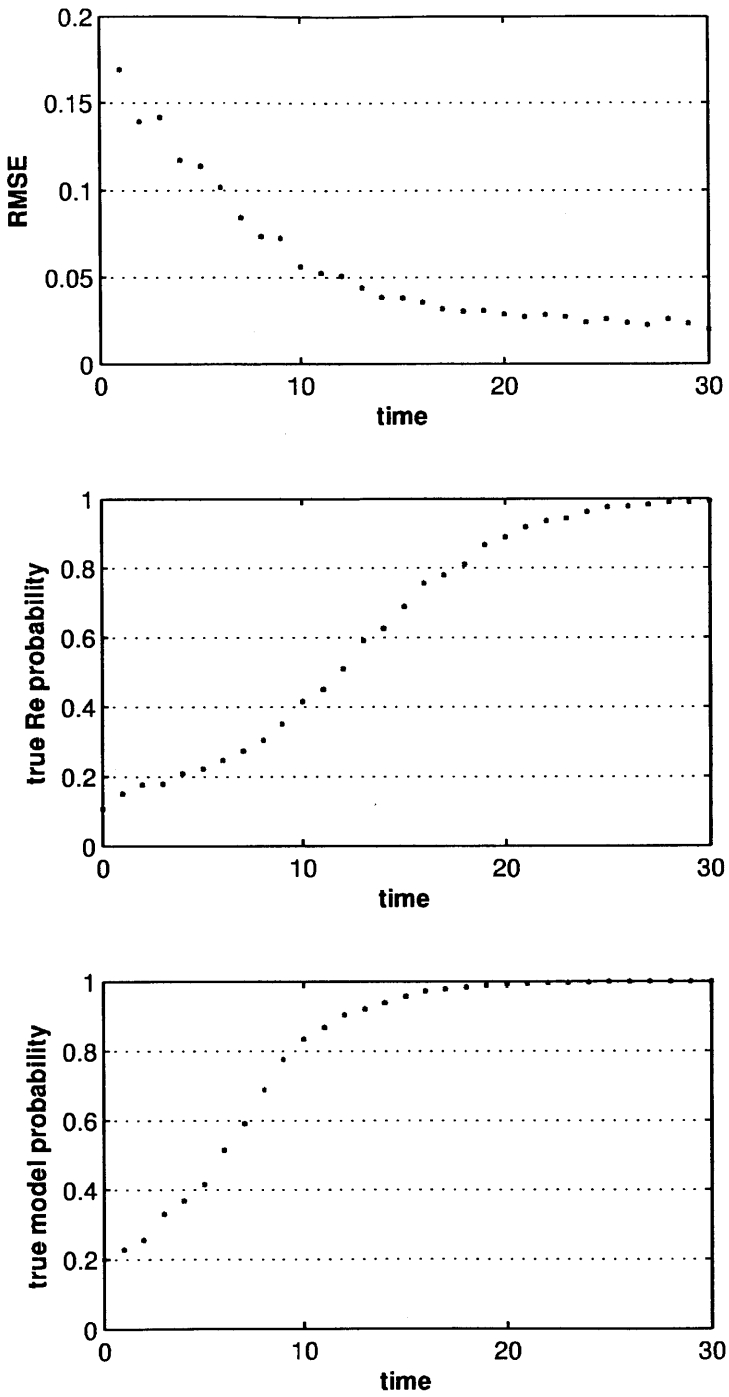

Figure 4-24: As Figure 4-19 but for the second experiment featuring the downstream-pointing triangular obstacle (Experiment A2).

Chapter 5

Microorganism Tracer

5.1 Marine Microorganisms

Microscopic marine organisms abound in the oceans of the world. These microorganisms generally fall within two functional groups: phytoplankton, which harvest solar energy through photosynthesis, and zooplankton, which feed on other microorganisms. Countless species of both phytoplankton and zooplankton exist, which collectively play major roles in the Earth's biogeochemical cycles. Concentrations of basic chemical compounds such as oxygen and carbon dioxide in both the Earth's oceans and atmosphere are closely tied to the abundance of these marine microorganisms, rendering their study of great importance to researchers in geochemistry and climate science [14, 39, 119]. Furthermore, phytoplankton and zooplankton constitute the foundations of the oceans' biological food networks. The growth and decay of these microorganisms thus have far-flung consequences for the health of many macroscale marine species, which garners them significant interest from marine biologists and ecologists [38, 56, 57, 80].

The submillimeter scale of phytoplankton and zooplankton often renders their self-directed movements within their fluid domains negligible compared to their bulk transport by convection. When this holds, these microorganisms are well-modeled as fluid flow tracers, immotile particles whose spatial evolution is governed entirely by the physical dynamics of their fluid environments [1, 2]. A consequence of this is that

species with salient optical properties can produce stunning natural visualizations of fluid flow within large bodies of water, as exhibited in Figure 5-1.

While they share a means of physical transport, the many distinct species of phytoplankton and zooplankton in the world's oceans each grow and decay according to unique biological (reaction) dynamics. Assuming the continuum hypothesis holds, one can define a microorganism concentration. The stochastic dynamical evolution of a microorganism species within a fluid domain can then be described by the general equation

$$\frac{\partial \rho}{\partial t} = \underbrace{\underbrace{-\boldsymbol{v} \cdot \nabla \rho}_{\text{advection}} + \underbrace{\kappa \nabla^2 \rho}_{\text{diffusion}}}_{\text{convection}} + \underbrace{F_\rho(\rho, \boldsymbol{r}, t; \omega)}_{\text{biology(reaction)}} , \tag{5.1}$$

where $\rho = \rho(\boldsymbol{r}, t; \omega)$ represents the spatial field of microorganism concentration defined as either mass or number of entities per fluid volume, κ represents the diffusivity of the microorganism within the fluid, and F_ρ represents the reaction equation for the microorganism, which captures all nonconvective influences [139]. Variables within the reaction equation for a particular species typically include the spatial distributions of various chemical compounds, which may be necessary for growth, and the spatial distributions of other microorganism species, which may act as predators, preys, or competitors for resources. Physical factors may also be included, such as fluid pressure, fluid temperature, and light abundance. If the microorganism is not entirely immotile, motility terms can also be included in the reaction (biological behavior) term. The identification of the reaction equation for individual species as well as the elucidation of general reaction equation features for groups of species are central focuses of marine microorganism research [136].

5.2 Stochastic Microorganism Tracer System

The second stochastic dynamical system considered for application of the Bayesian model inference methodology developed in Chapter 3 is a simulation of a single marine microorganism species convected through a stochastic fluid domain. The two-dimensional spatial domain of the system is shown in Figure 5-2, with fluid flow

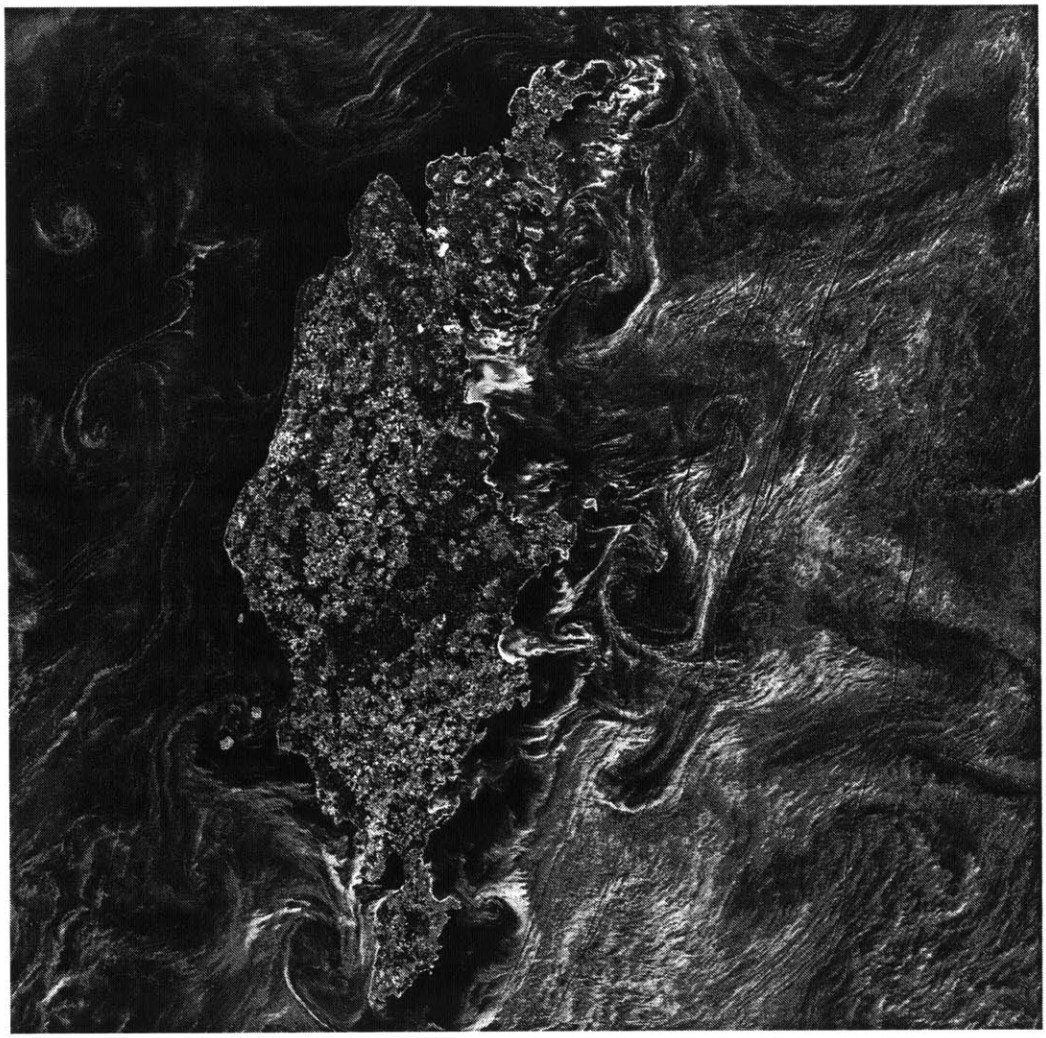

Figure 5-1: Image of intense concentrations of phytoplankton around the Swedish island of Gotland in the Baltic Sea taken by the Landsat 7 satellite on July 13, 2005 [105]. Green chlorophyll in the photosynthesizing phytoplankton renders them optically visible.

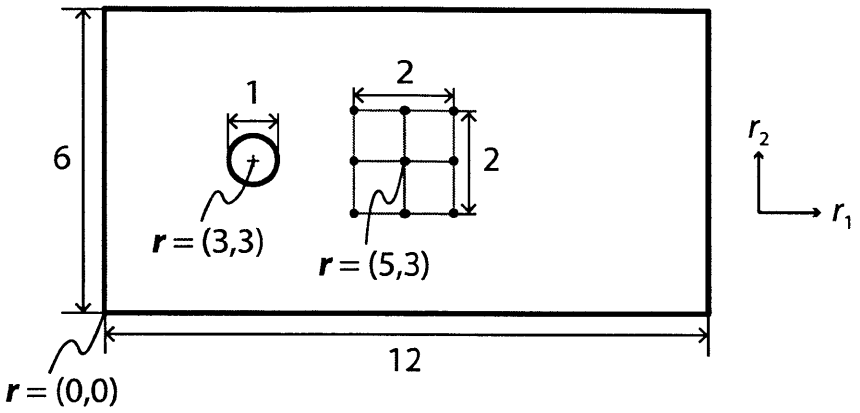

Figure 5-2: Two-dimensional spatial domain of the stochastic microorganism tracer system. All lengths and coordinates are given in nondimensionalized length units. The observation array is indicated by the nine points situated at the center of the domain.

proceeding left to right in the positive r_1 direction. The domain features the familiar circular obstacle that defines the classic flow past a cylinder system introduced in Section 4.1. This is a simple idealization of flows around an island, as exhibited in Figure 5-1, and these flows affect the microorganism's fluid environment. A Reynolds number of 50 is chosen for the system. The microorganism is assumed to be immotile and thus represents a fluid flow tracer. Uncertainty in the fluid velocity and microorganism concentration fields is introduced in two ways: 1) uncertainty in the initial velocity field and 2) uncertainty in the reaction equation for the microorganism.

As in the stochastic flow past an obstacle system explored in Chapter 4, the phase of vortex shedding behind the obstacle in the stochastic microorganism tracer system is highly sensitive to the initial fluid velocity field. Since the concentration field is closely coupled to the velocity field through (5.1), uncertainty in the initial velocity field translates directly to uncertainty in the concentration field as well. Figure 5-3 illustrates that different phase shifts of the periodic fluid flow pattern downstream of the obstacle can correspond to substantially different fluid velocity and microorganism concentration fields.

In addition to uncertainty arising from the fluid velocity field, uncertainty in the reaction equation for the microorganism can significantly increase the uncertainty

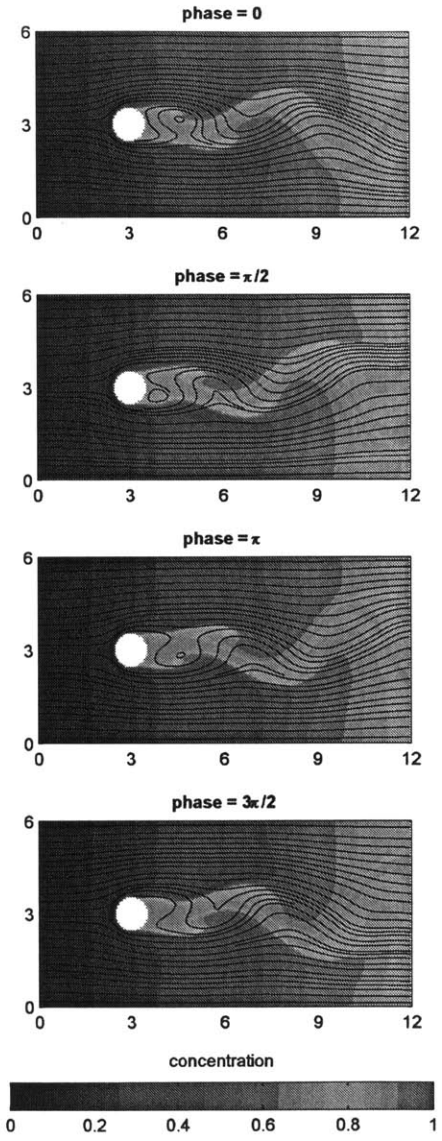

Figure 5-3: Microorganism concentrations overlaid on fluid flow patterns, for realizations of the stochastic microorganism tracer system featuring a basic microorganism reaction equation and phase shifts of 0, $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$. Directions of fluid flow are indicated with streamlines. Note that the flow patterns for phase shifts of 0 and π are a mirrored pair about the $r_2 = 3$ centerline, as are the flow patterns for phase shifts of $\frac{\pi}{2}$ and $\frac{3\pi}{2}$.

in the microorganism concentration field. Basic phytoplankton reaction terms are assumed for the microorganism behavior, i.e.

$$F_\rho(\rho, \boldsymbol{r}, t; \omega) = \underbrace{g(\boldsymbol{r}, t)\frac{\rho(\rho_{\max} - \rho)}{\rho_{\max} - \rho + k_g}}_{\text{growth}} - \underbrace{d(\boldsymbol{r}, t)\rho}_{\text{decay}} , \tag{5.2}$$

where $g(\boldsymbol{r}, t)$ represents the microorganism growth factor, $\rho_{\max}$ represents the maximum microorganism concentration, k_g represents the growth regularization parameter, and $d(\boldsymbol{r}, t)$ represents the microorganism decay factor [136]. The values of $\rho_{\max}$ and k_g are both chosen to be 1.0. The formulation of the growth and decay factors however are assumed to be uncertain, with three possible formulations for $g(\boldsymbol{r}, t)$

$$g_+(\boldsymbol{r}, t) = g_+(t) = (1 + A_T \sin(t/T_\rho))\, G , \tag{5.3}$$

$$g_0(\boldsymbol{r}, t) = g_0 = G , \tag{5.4}$$

$$g_-(\boldsymbol{r}, t) = g_-(t) = (1 - A_T \sin(t/T_\rho))\, G \tag{5.5}$$

and three possible formulations for $d(\boldsymbol{r}, t)$

$$d_+(\boldsymbol{r}, t) = d_+(\boldsymbol{r}) = \left(1 + A_L \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D , \tag{5.6}$$

$$d_0(\boldsymbol{r}, t) = d_0 = D , \tag{5.7}$$

$$d_-(\boldsymbol{r}, t) = d_-(\boldsymbol{r}) = \left(1 - A_L \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D , \tag{5.8}$$

where $\|\bullet\|$ represents the L_2 norm operator for any given vector and $\boldsymbol{r}_0 = (3, 3)$ represents the center of the circular obstacle illustrated in Figure 5-2. The temporal dependence in (5.3) and (5.5) is used to simulate possible periodic influences in real ocean environments, such as solar light cycles or tides, while the spatial dependence in (5.6) and (5.8) is used to simulate possible near-coast influences, such as bathymetric changes or pollution. The growth factor parameters G, A_T, and T_ρ are chosen to be 0.8, 0.5, and 0.7 respectively while the decay factor parameters D, A_L, and L_ρ are chosen to be 0.1, 1.8, and 1.0 respectively. The combination of the three growth factor forms (5.3)–(5.5) and three decay factor forms (5.6)–(5.8) leads to a total of

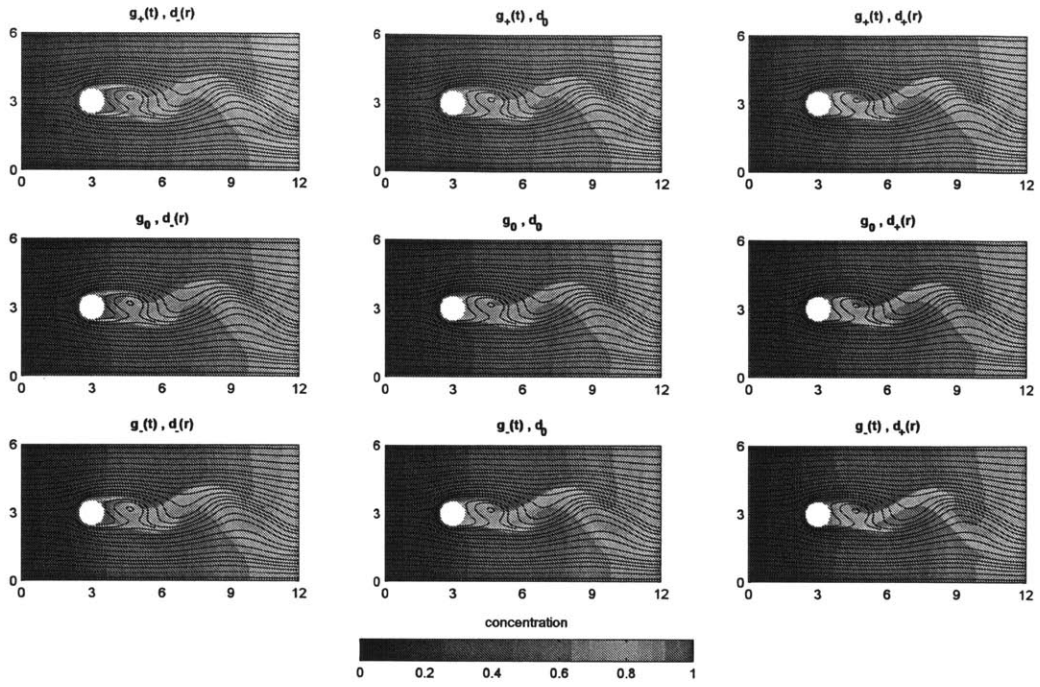

Figure 5-4: Microorganism concentrations overlaid on fluid flow patterns, for realizations of the stochastic microorganism tracer system featuring nine different microorganism reaction equations at non-dimensional time $t = \pi \times T_\rho \approx 2.2$ (i.e. when periodic temporal effects are either maximal or minimal).

nine possible reaction equations for the microorganism. These reaction equations result in subtle but significant variations of the microorganism concentration field, as exhibited in Figure 5-4.

5.3 Description of the Experiments

5.3.1 Model Formulation

Nine possible reaction equations are considered for the microorganism in the system, leading to nine possible sets of stochastic dynamical equations. Each of these equation sets represents a distinct candidate model for the system, with each candidate assumed to be equally likely a priori (i.e. a uniform initial model distribution is used). We proceed to describe each component of (1.2) for each of these nine candidate models.

$\mathcal{D}_n$ — As in the stochastic flow past an obstacle system, the evolution of the fluid velocity field in the stochastic microorganism tracer system for all nine candidate models is modeled using the two-dimensional Navier-Stokes equations for an incompressible, Newtonian fluid in nondimensional form (4.2)–(4.4). These equations are coupled to the equation for the evolution of the microorganism concentration field (5.1), with the general form of the reaction equation given by (5.2). Considering the three possible forms for the growth factor (5.3)–(5.5) and the three possible forms for the decay factor (5.6)–(5.8), nine distinct reaction equations are possible, one for each of the nine candidate models. The flow parameters are here assumed to be determined. Specifically, Re in (4.2) and (4.3) is chosen to be 50 while κ in (5.1) is chosen to be 0 for simplicity.

SG_n — Unlike the stochastic flow past an obstacle system, the geometries of the spatial domains for the nine candidate models here are all the same, with the domain for each candidate represented by Figure 5-2. For purposes of numerical implementation, these spatial domains are discretized using a regular rectangular grid of 240 and 120 elements along the r_1 and r_2 directions respectively ($\Delta r_1 = \Delta r_2 = \frac{1}{20}$). This spatial discretization results in 28,349 r_1-velocity, 28,229 r_2-velocity, and 28,489 microorganism concentration state variables for each of the candidate models. The state vector for any of the candidate models is then

$$\boldsymbol{x}(t;\omega) = \begin{bmatrix} \boldsymbol{v_1}(t;\omega) \\ \boldsymbol{v_2}(t;\omega) \\ \boldsymbol{\rho}(t;\omega) \end{bmatrix} ,$$

where $\boldsymbol{\rho}(t;\omega)$ represents the spatially-discretized vector of concentration state variables (state augmentation is not needed for this system because all parameters are assumed to be determined). $N_{\boldsymbol{X}}$, the dimension of the state vector, is thus 85,067.

BC_n — Deterministic Dirichlet boundary conditions for the velocity and concentration fields are used for the left domain limits for all nine candidate models

$$\boldsymbol{v} = (1,0) \ , \ \rho = 0.1 \text{ for } r_1 = 0 \ , \tag{5.9}$$

while deterministic Neumann boundary conditions are used for the top, bottom, and right domain limits

$$\frac{\partial \boldsymbol{v}}{\partial r_2} = 0 \ , \ \frac{\partial \rho}{\partial r_2} = 0 \text{ for } r_2 = 0 \text{ and } r_2 = 6 \ ,$$
$$\frac{\partial \boldsymbol{v}}{\partial r_1} = 0 \ , \ \frac{\partial \rho}{\partial r_1} = 0 \text{ for } r_1 = 12 \ .$$

Additional deterministic Dirichlet boundary conditions are used to impose no-slip conditions on the surface of the circular obstacle for all the candidate models.

IC_n — For each of the nine candidate models, the state vector mean as represented in the expansion (3.23) is initialized as follows. The means of the velocity state variables are obtained by numerically solving the continuity equation (4.4), given the boundary condition (5.9). The means of the microorganism concentration states variables are subsequently found by numerically solving (5.1) for the steady state concentrations corresponding to these mean velocities. For each of the candidate models, a covariance matrix is then constructed for both the velocity and concentration state variables that respects symmetry about the $r_2 = 3$ centerline using again the boundary-mollified spatial covariance method suggested in [121]. Discrete Karhunen-Loève transforms are performed on these covariance matrices to initialize eight pairs of modes and mode coefficients for each candidate model, for a DO subspace dimension $N_{\mathrm{DO}} = 8$. As for the stochastic flow past an obstacle system, the symmetry properties of this procedure lead to unbiased initialization of uncertainty in vortex shedding phase, as exhibited in Figures 5-6 and 5-7.

5.3.2 True Solution Generation

We generate four different true solutions for the stochastic microorganism tracer system. Each of these simulations has its own fixed formulation for the microorganism: each represents the 'truth' for the corresponding experiment. The first—Experiment B1—features the constant formulations for growth factor, g_0, and decay factor, d_0, representing a spatially and temporally invariant microorganism reaction equation.

The second—Experiment B2—features the constant growth factor formulation g_0 and the spatially-variable decay factor formulation $d_-(\boldsymbol{r})$, representing a spatially variant but temporally invariant reaction equation. The third—Experiment B3—features the time-dependent growth factor formulation $g_+(t)$ and the spatially-variable decay factor formulation $d_-(\boldsymbol{r})$, representing a reaction equation that is both spatially and temporally variant. Finally, the fourth—Experiment B4—features the growth factor formulation

$$g'(\boldsymbol{r},t) = g'(t) = (1 + (1/2)A_T \sin(t/T_\rho))\, G \tag{5.10}$$

and the decay factor formulation

$$d'(\boldsymbol{r},t) = d'(\boldsymbol{r}) = \left(1 - (1/2)A_L \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D\ , \tag{5.11}$$

representing a reaction equation with intermediate spatial and temporal variance that, importantly, is not explicitly included in the set of candidate models formulated for the system.

Deterministic initial velocity fields satisfying the continuity equation (4.4) are constructed for all four simulated truths, with arbitrary asymmetries introduced to induce vortex shedding. Microorganism concentration fields are initialized at 0.1. Deterministic evolution of the velocity and concentration fields is performed by numerically solving the Navier-Stokes equations (4.2)–(4.4) and microorganism evolution equation (5.1) using a finite-volume approach employing a regular rectangular grid of 240 and 120 elements along the r_1 and r_2 directions respectively and a non-dimensional time-step of of $\Delta t = \frac{1}{120}$ up to a final non-dimensional time $T = 80$. These deterministic simulations represent four possible realizations of the stochastic microorganism tracer system, each with a true microorganism reaction equation, true dynamical velocity field, and true dynamical concentration field that are to be inferred. Each of these four solutions constitutes an independent experiment conducted with this system and results for all four will be described in the following section.

5.3.3 Observations and Inference

Every 1 non-dimensional time unit (i.e. every 120 numerical time-steps), noisy observations of microorganism concentration are made at the nine locations indicated in Figure 5-2, for a total of $N_{\boldsymbol{Y}} = 9$ observations at every observation time. A single non-dimensional time unit for this system corresponds to a phase shift of approximately $\frac{\pi}{3}$ when $\mathrm{Re} = 50$. Unbiased Gaussian noise is applied to the observations with a standard deviation equal to approximately 5% of the maximum microorganism concentration, which are then used by the GMM-DO filter to perform joint model-conditional inference of the velocity and concentration state variables for each of the candidate models. The linear observation matrix $\mathbf{H}$ appearing in (3.16) is specified so as to identify the concentration state variables corresponding to the nine observation locations while the matrix of observation covariances $\mathbf{R}$ is specified as a diagonal matrix with diagonal elements equal to the variance of the observation noise.

Following the procedure outlined in [129], the Bayesian information criterion (BIC), in conjunction with the expectation-maximization (EM) algorithm for GMMs, is used to select the optimal number of GMM components N_{GMM} at every observation time. Typical BIC-optimized values for N_{GMM} were found to lie between 40 and 60 for experiments conducted with the stochastic microorganism tracer system. Marginal likelihoods for the candidate models are calculated using (3.31) following every iteration of the GMM-DO filter and the model distribution is updated according to (1.9).

5.3.4 Numerical Method

Numerical evolution of the state vector mean, modes, and mode coefficients for each of the candidate models is performed using the DO evolution equations (3.4)–(3.6) and the Navier-Stokes equations (4.2)–(4.4) with a fluid flow tracer (5.1) according to the finite-volume methodology developed in [141]. To obtain explicit differential equations for the evolution of the DO decomposition for the non-polynomial nonlinearities of the microorganism reaction terms (5.2), a local linearization is employed [91]. These

locally linearized terms are derived in Appendix D. A non-dimensional time-step of $\Delta t = \frac{1}{120}$ is used. $N_{\text{MC}} = 10^4$ Monte Carlo samples are used for the stochastic evolution of the mode coefficients. For numerical robustness, the logarithms of the marginal likelihoods calculated using (3.31) are again used to perform model inference rather than the marginal likelihoods themselves according to (4.7).

A summary of the experimental properties used for the stochastic microorganism tracer system is provided in Table 5.1.

Table 5.1: Experimental properties for the stochastic flow past an obstacle system.

Property	Value
$N_{\mathcal{M}}$	9
$N_{\boldsymbol{X}}$	85,067
$N_{\boldsymbol{\Theta}}$	0
$N_{\boldsymbol{Y}}$	9
N_{DO}	8
N_{MC}	10^4
Δr_1	1/20
Δr_2	1/20
Δt	1/120
T	80

5.4 Results

As for the stochastic flow past an obstacle system, Bayesian state variable mean RMSE and true model probability are used to evaluate the success of the Bayesian model inference methodology as applied to the stochastic microorganism tracer system. Since their numerical values were all normalized to be of order 1, both fluid velocity and microorganism concentration state variables are included in RMSE calculations without weighting. A RMSE approaching 0 and a true model probability approaching 1 again indicate successful state and model inference respectively.

5.4.1 Experiment B1: Constant Growth and Decay

The first experiment we consider, Experiment B1, is that of a simulated truth defined by the deterministic simulation featuring the constant growth and decay factor formulations g_0 and d_0 respectively. The goal is to infer the state of the velocity fields and microorganism concentration jointly with the formulation of the reaction equation when noisy observations are made from this simulated truth. The evolution of this inference is illustrated in Figures 5-5 to 5-10.

Figures 5-5, 5-6, and 5-7 illustrate the system at the initial non-dimensional time ($t = 0$), before any observations are made. The fluid velocity and microorganism concentration fields of the deterministic simulated truth are in a time-periodic state. Regular asymmetric vortex shedding is occurring downstream of the obstacle, with a period of approximately 6.0 non-dimensional time units. The model-conditional velocity and concentration field means for all nine candidate models are symmetric about the $r_2 = 3$ centerline, indicating no bias towards any particular vortex shedding phase. The model distribution is uniform. The model-conditional velocity and concentration standard deviations are also all symmetric about the $r_2 = 3$ centerline.

Figure 5-8 illustrates the system after 10 non-dimensional time units and 10 observation episodes. The model-conditional microorganism concentration field means for all nine candidate models (each mean is estimated using the GMM-DO filter) are beginning to align in phase with the true concentration field. Despite the absence of direct velocity field observations, the conditional velocity field means for the candidates are also beginning to align with the true velocity field, as a result of the joint inference of microorganism concentration and fluid velocity. Inference of the vortex shedding phase behind the obstacle however is far from complete, as evidenced by the blurred features in the Bayesian mean of these conditional concentration field means, when compared to those in the true field. The true model of the system is favored by the model distribution, with 'adjacent' models—i.e. models featuring smaller deviations in their growth and decay factor forms—maintaining higher probabilities than distant models.

Figure 5-9 illustrates the system after 30 non-dimensional time units and 30 observation episodes. Agreement between the true velocity and concentration fields and the inferred velocity and concentration field means has improved significantly, though some concentration field features are still slightly blurred in the Bayesian mean. The true model is now strongly favored by the model distribution, with only adjacent models remaining as plausible alternatives.

Figure 5-10 illustrates the system after 80 non-dimensional time units and 80 observation episodes. The inferred velocity and concentration field means are now consistent with the true velocity and concentration fields. The true model is now also the only plausible model for the system, as indicated by the highly peaked model distribution.

Figure 5-11 illustrates the time progression of the two inference metrics selected above. The RMSE approaches 0 and the true model probability approaches 1, both indicating successful inference.

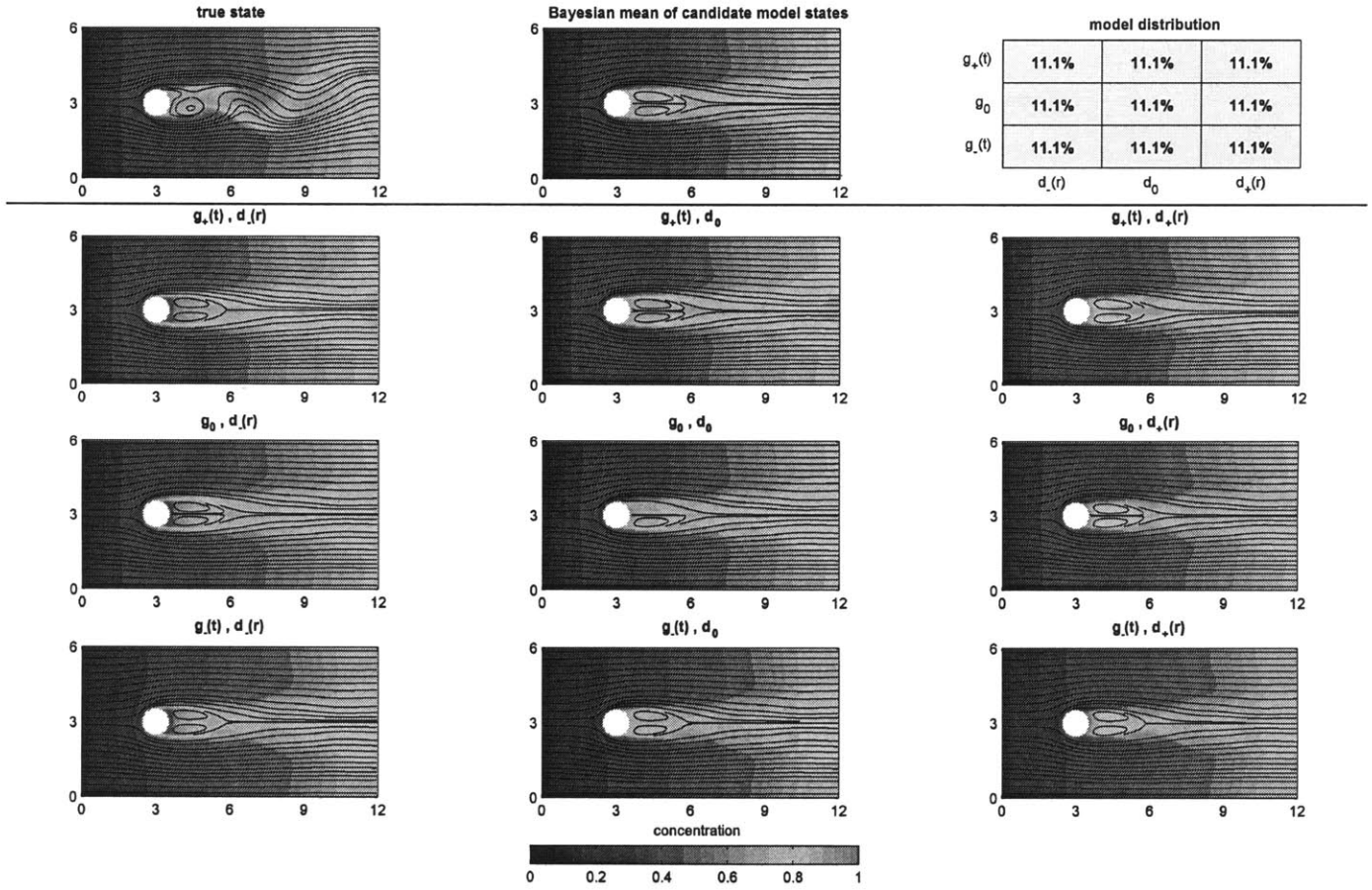

Figure 5-5: The stochastic microorganism tracer system with observations made from the deterministic simulated truth featuring the growth factor g_0 and the decay factor d_0 (Experiment B1) at non-dimensional time $t = 0$. Observation locations are indicated on the true state field in pink. Model-conditional velocity and concentration field means for each of the nine candidate models are displayed on the bottom, arranged in an array corresponding to the model distribution illustration in the top right corner.

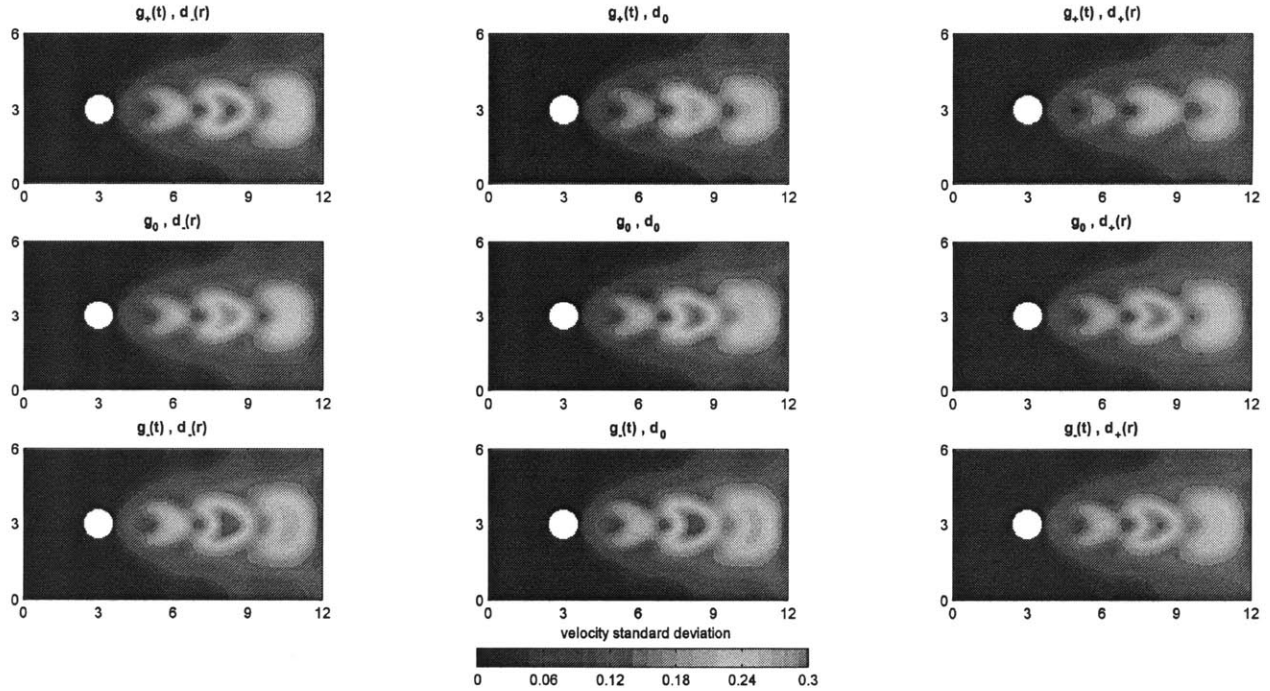

Figure 5-6: As Figure 5-5 but showing model-conditional uncertainties at non-dimensional time $t = 0$. Specifically, model-conditional velocity field standard deviations for each of the nine candidate models are displayed, arranged in an array corresponding to the model distribution illustration in the top right corner of Figure 5-5.

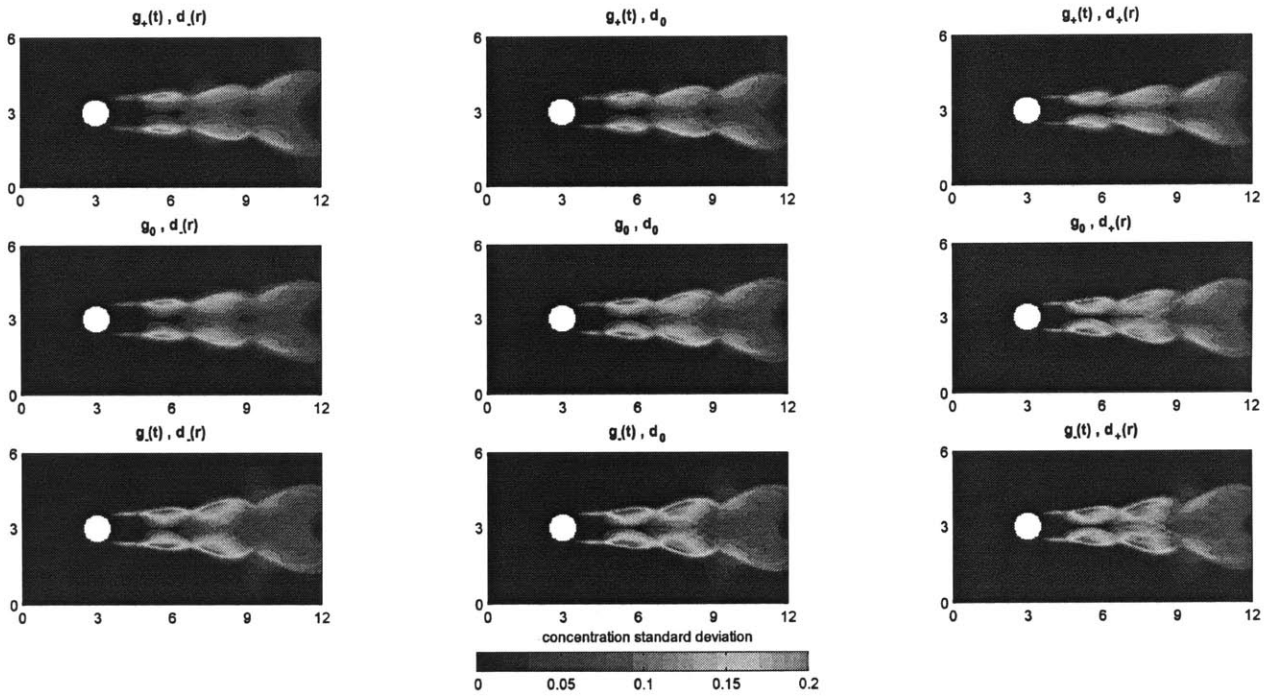

Figure 5-7: As Figure 5-6 but with model-conditional concentration field standard deviations displayed.

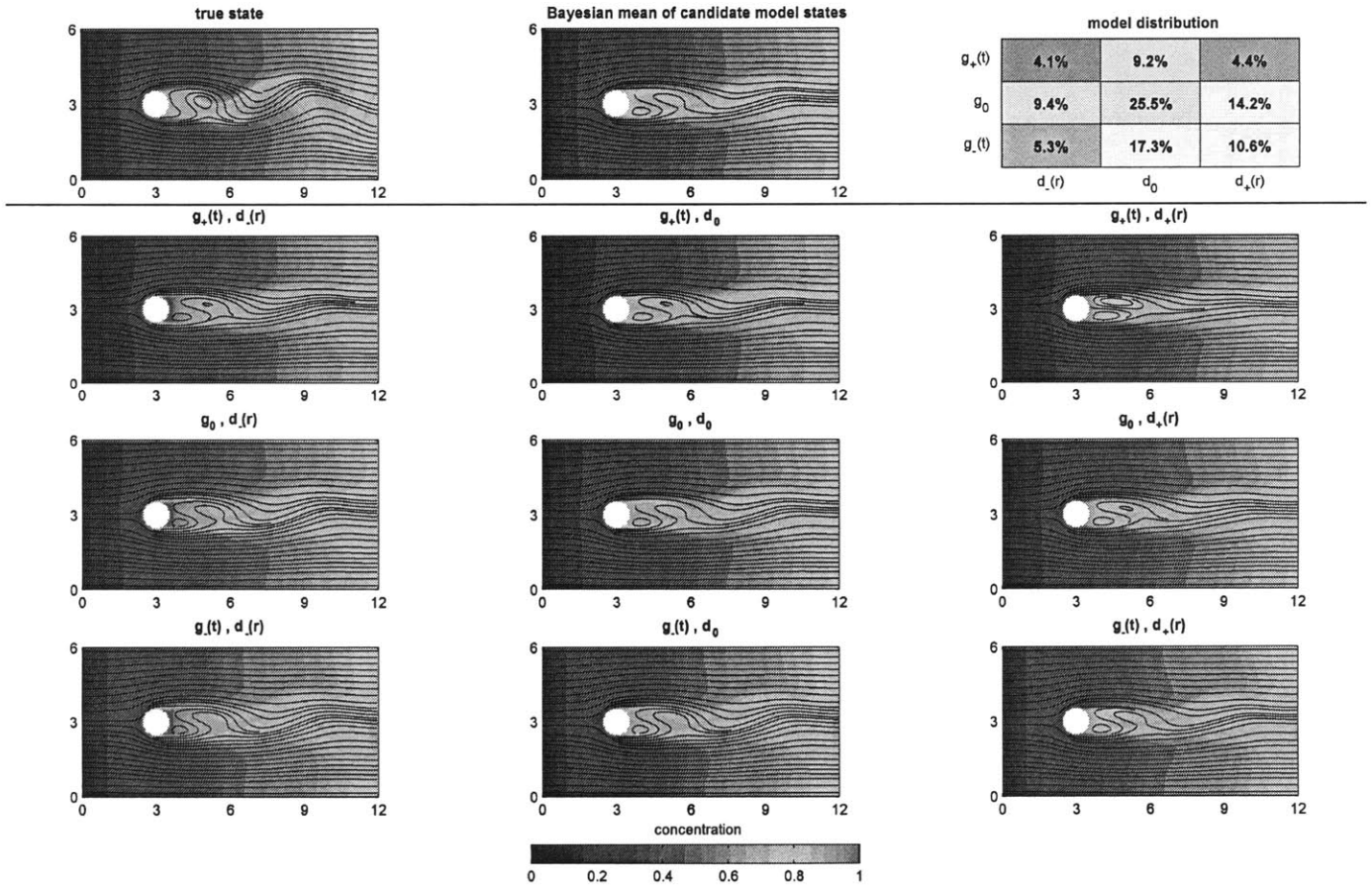

Figure 5-8: As Figure 5-5 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

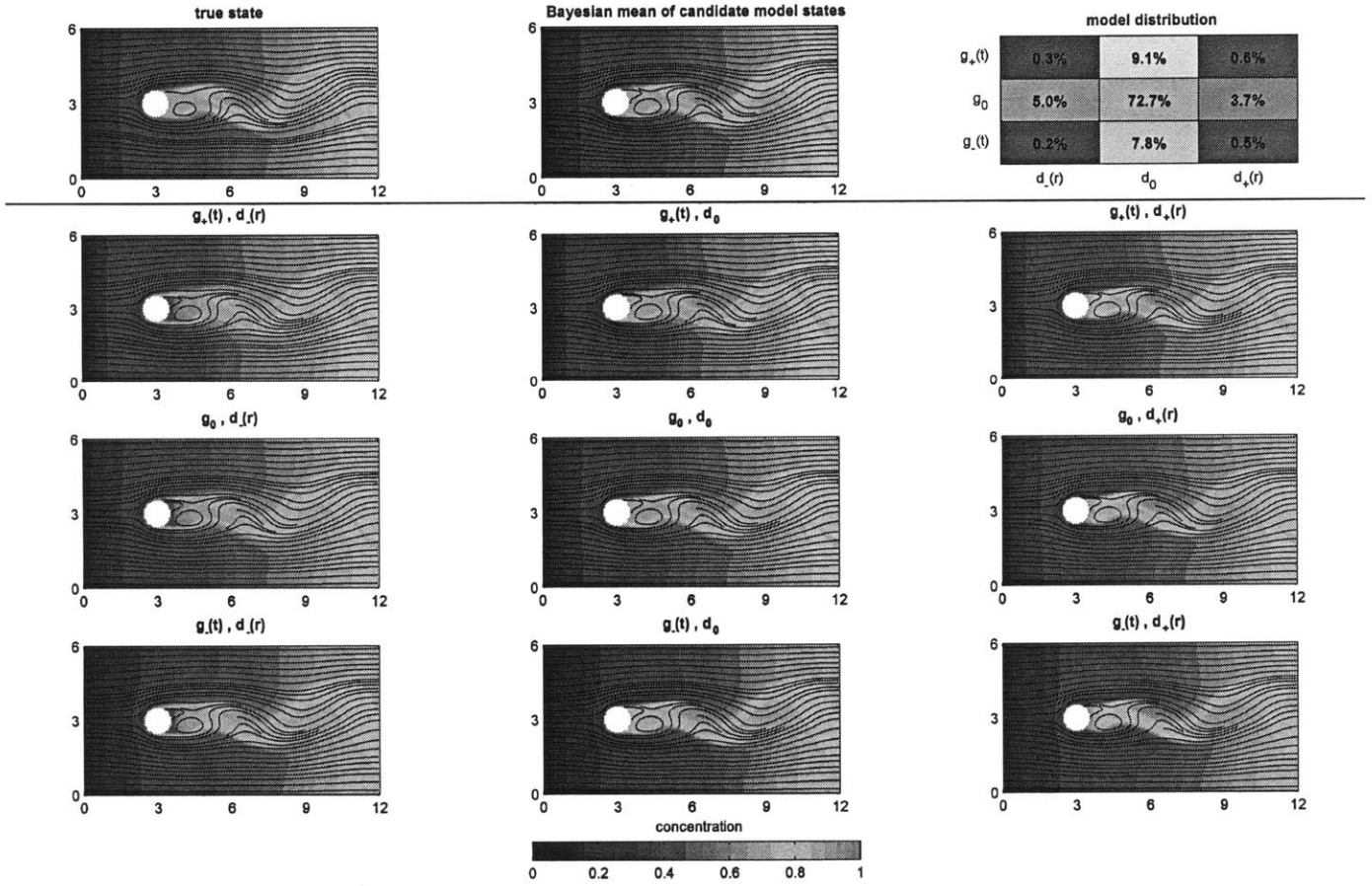

Figure 5-9: As Figure 5-5 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

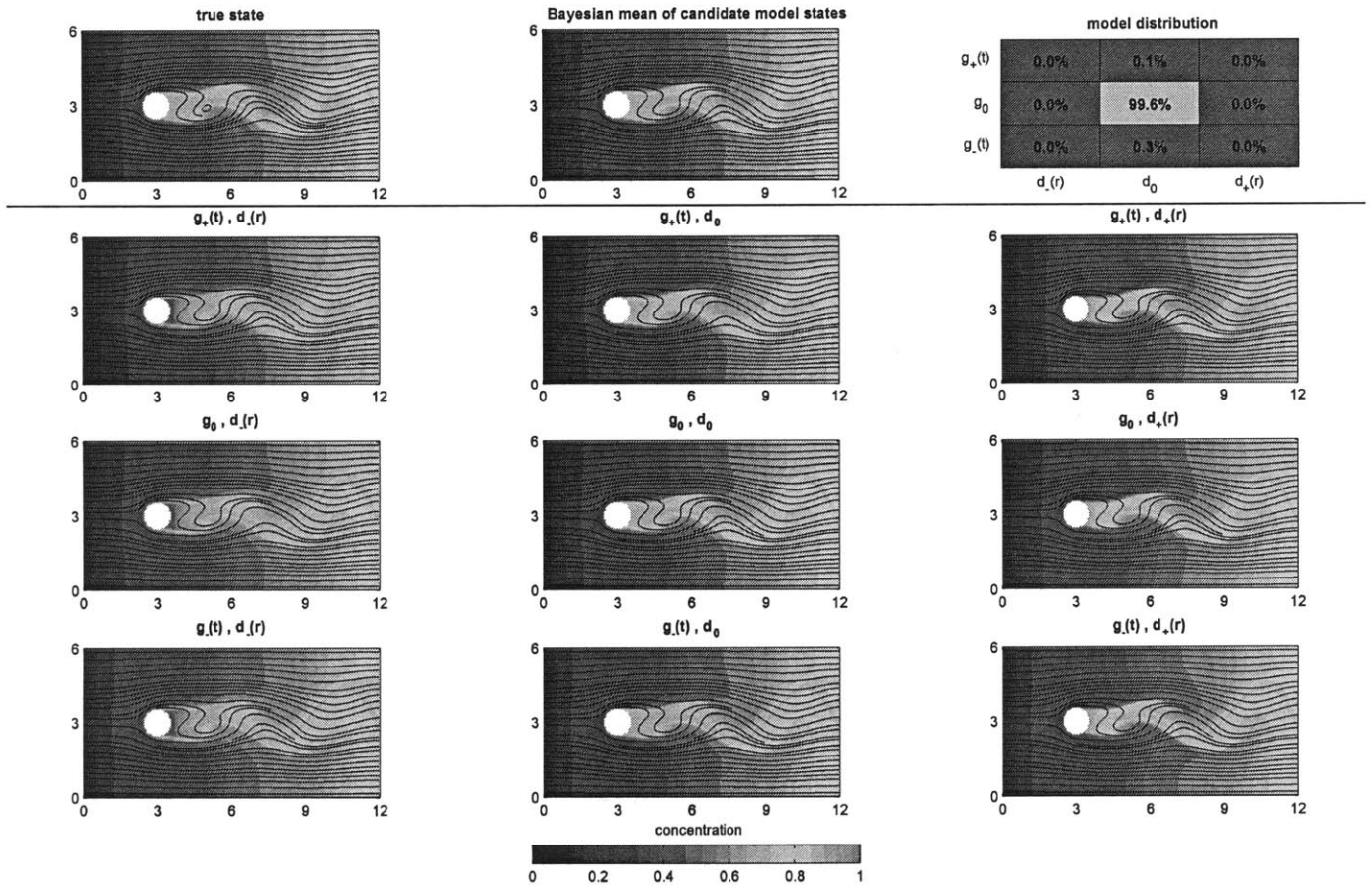

Figure 5-10: As Figure 5-5 but at non-dimensional time $t = 80$ (i.e. after 80 observation episodes).

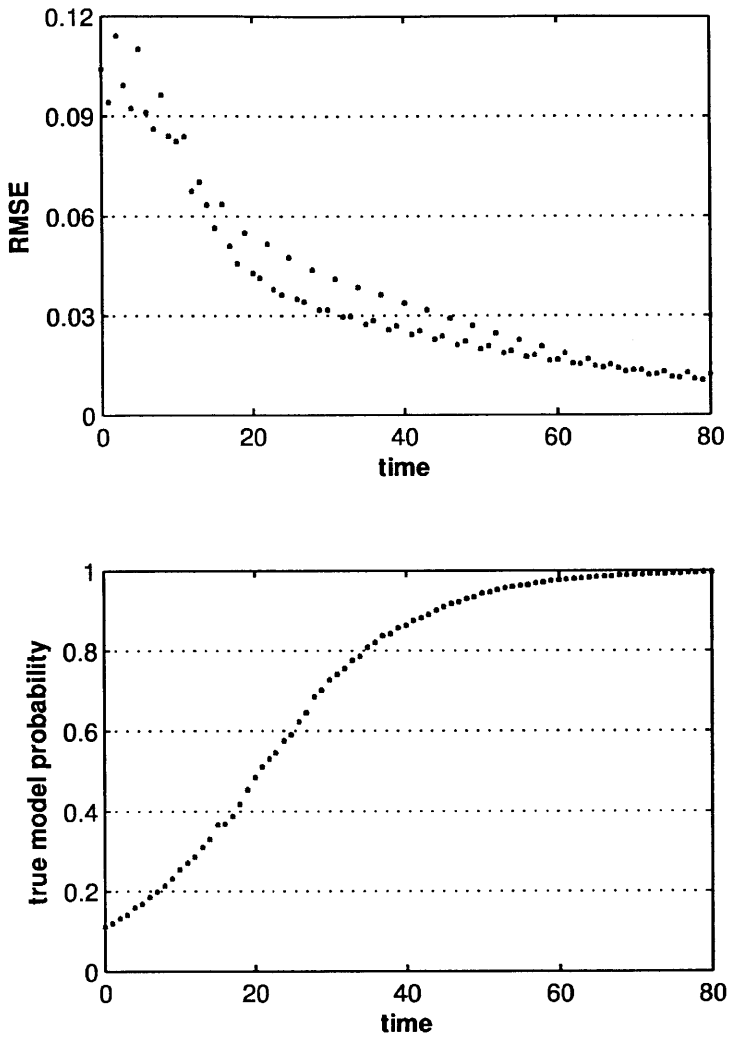

Figure 5-11: Time progression of inference metrics for the stochastic microorganism tracer system (as defined at the beginning of Section 4.4) with observations made from the deterministic simulated truth featuring the constant growth factor g_0 and decay factor d_0 formulations (Experiment B1). Metrics are plotted at every 1 non-dimensional time unit immediately following observation and inference (i.e. posterior values from the GMM-DO filter).

5.4.2 Experiment B2: Constant Growth and Spatially-Variable Decay

The second experiment we consider, Experiment B2, is that of a simulated truth defined by the deterministic simulation featuring the constant growth and spatially-variable decay factor formulations g_0 and $d\_(\boldsymbol{r})$ respectively. The evolution of this inference is illustrated in Figures 5-5 to 5-10. The results are analogous to those illustrated in Figures 5-5 to 5-10, with successful inference indicated by the time progression of the inference metrics in Figure 5-16. Note that the true model probability rises to 1 slightly faster in Experiment B2 than in Experiment B1. This is due to the fact that the true model in Experiment B2 resides on the periphery of the model space (i.e. the set of candidate models/feasible model formulations) while the true model in Experiment B1 resides at the center. RMSE consequently decreases slightly faster for Experiment B2 than for Experiment B1.

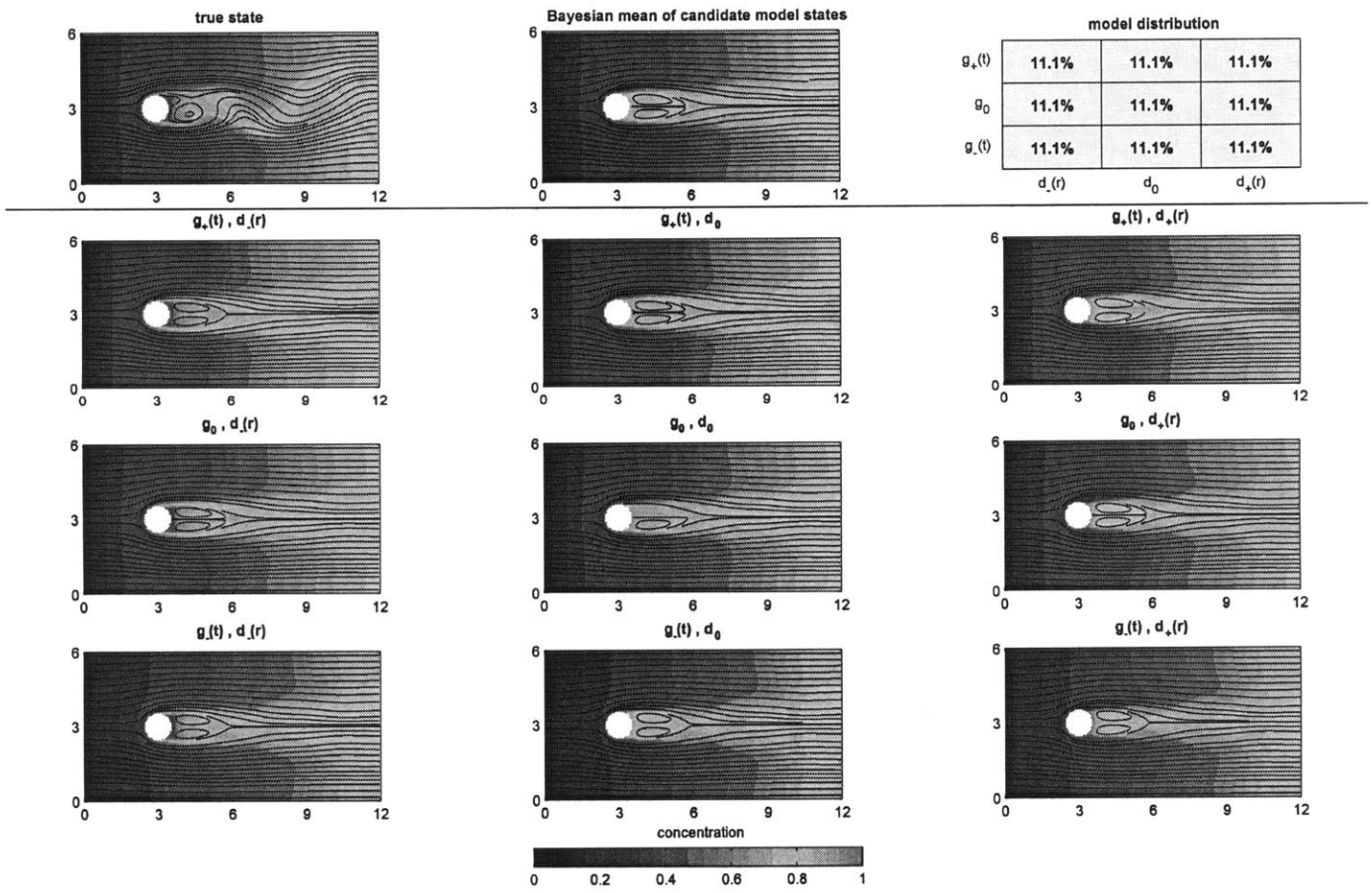

Figure 5-12: As Figure 5-5 but for the second experiment featuring the growth factor g_0 and decay factor $d_-(\boldsymbol{r})$ formulations (Experiment B2).

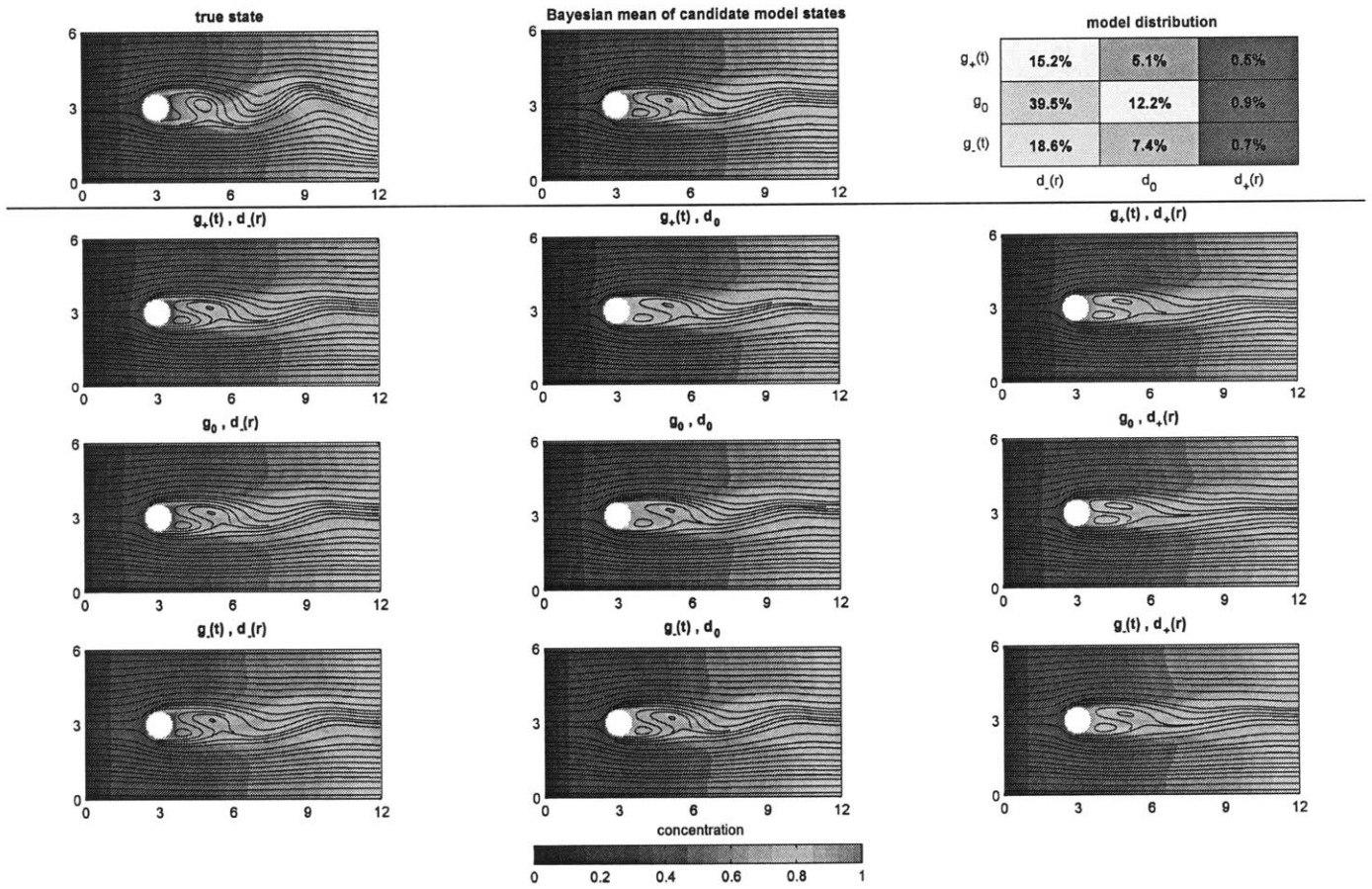

Figure 5-13: As Figure 5-12 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

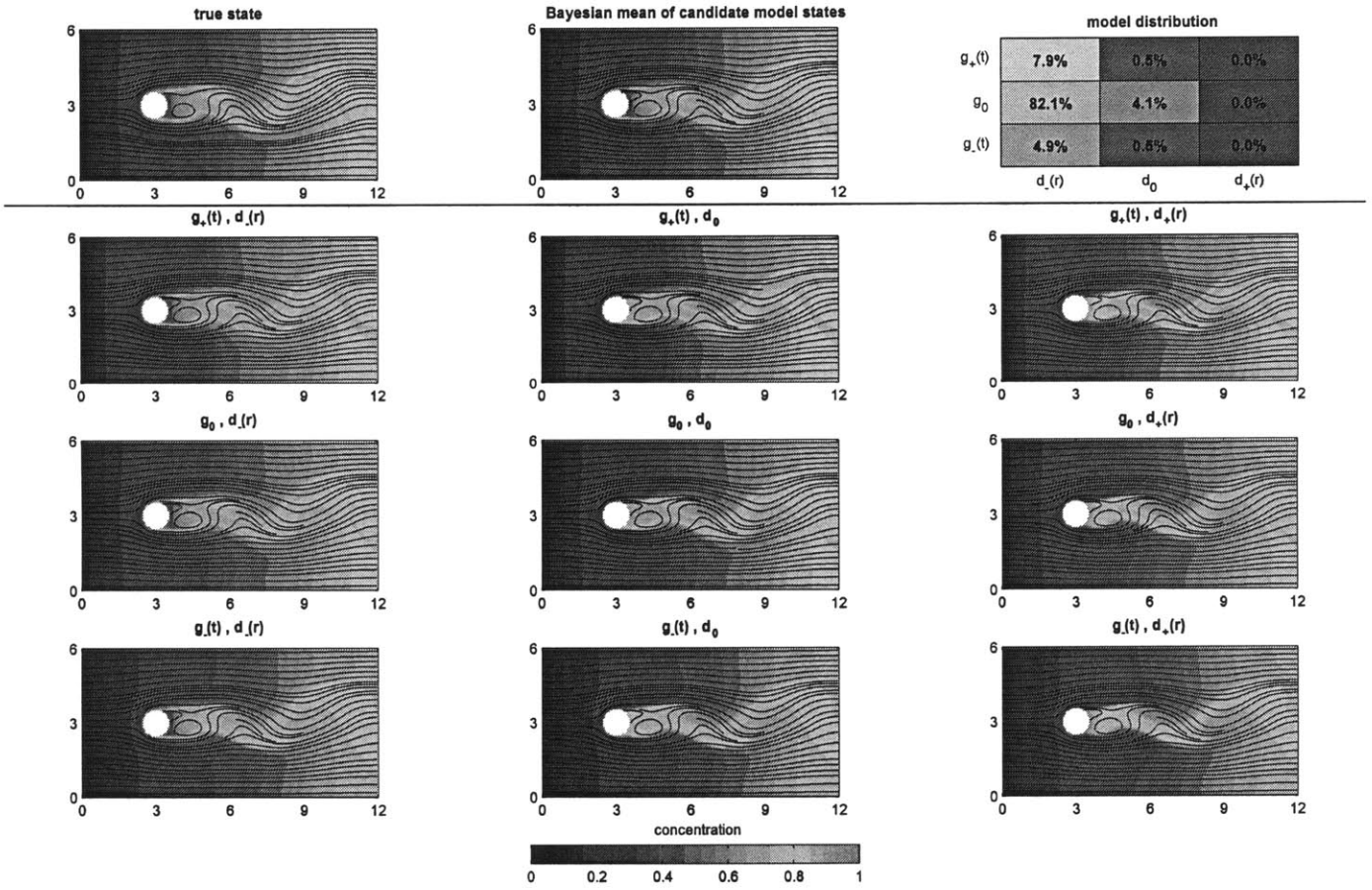

Figure 5-14: As Figure 5-12 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

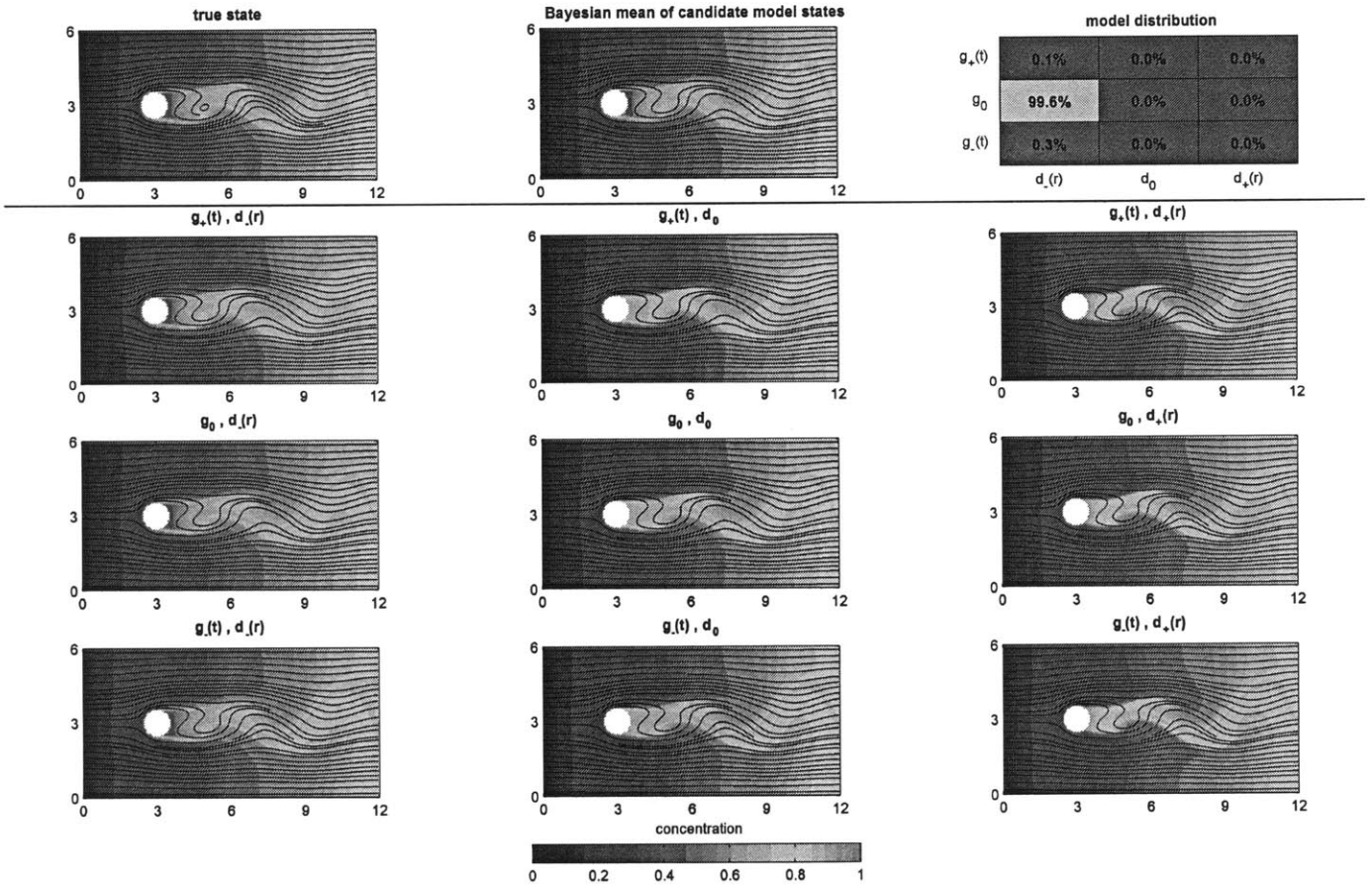

Figure 5-15: As Figure 5-12 but at non-dimensional time $t = 80$ (i.e. after 80 observation episodes).

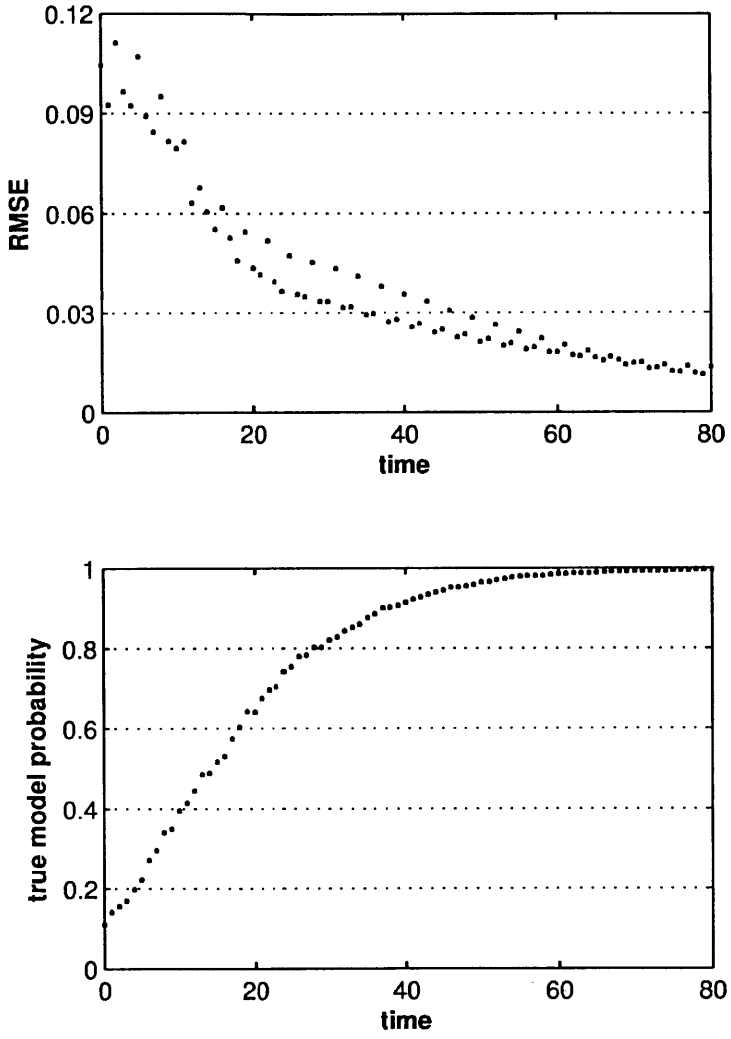

Figure 5-16: As Figure 5-11 but for the second experiment featuring the growth factor g_0 and decay factor $d_{-}(\boldsymbol{r})$ formulations (Experiment B2).

5.4.3 Experiment B3: Time-Dependent Growth and Spatially-Variable Decay

The third experiment we consider, Experiment B3, is that of a simulated truth defined by the deterministic simulation featuring the time-dependent growth and spatially-variable decay factor formulations $g_+(t)$ and $d_-(\boldsymbol{r})$ respectively. The evolution of this inference is illustrated in Figures 5-17 to 5-20. The results are again analogous to those illustrated in Figures 5-5 to 5-10, with successful inference indicated by the time progression of the inference metrics in Figure 5-21. Note that the true model probability rises to 1 even faster in Experiment B3 than in Experiment B2. This is again due to the fact that the true model in Experiment B3 resides on the periphery of the model space, with even fewer 'adjacent' candidate models than the true model in Experiment B2. RMSE consequently decreases even faster for Experiment B3 than for Experiment B2. This is analogous to the different rates of inference for Experiments A1 and A2 discussed in Section 4.4.2: the true model in Experiment B3 is more dynamically distinct from the other candidate models than the true models in Experiments B1 and B2.

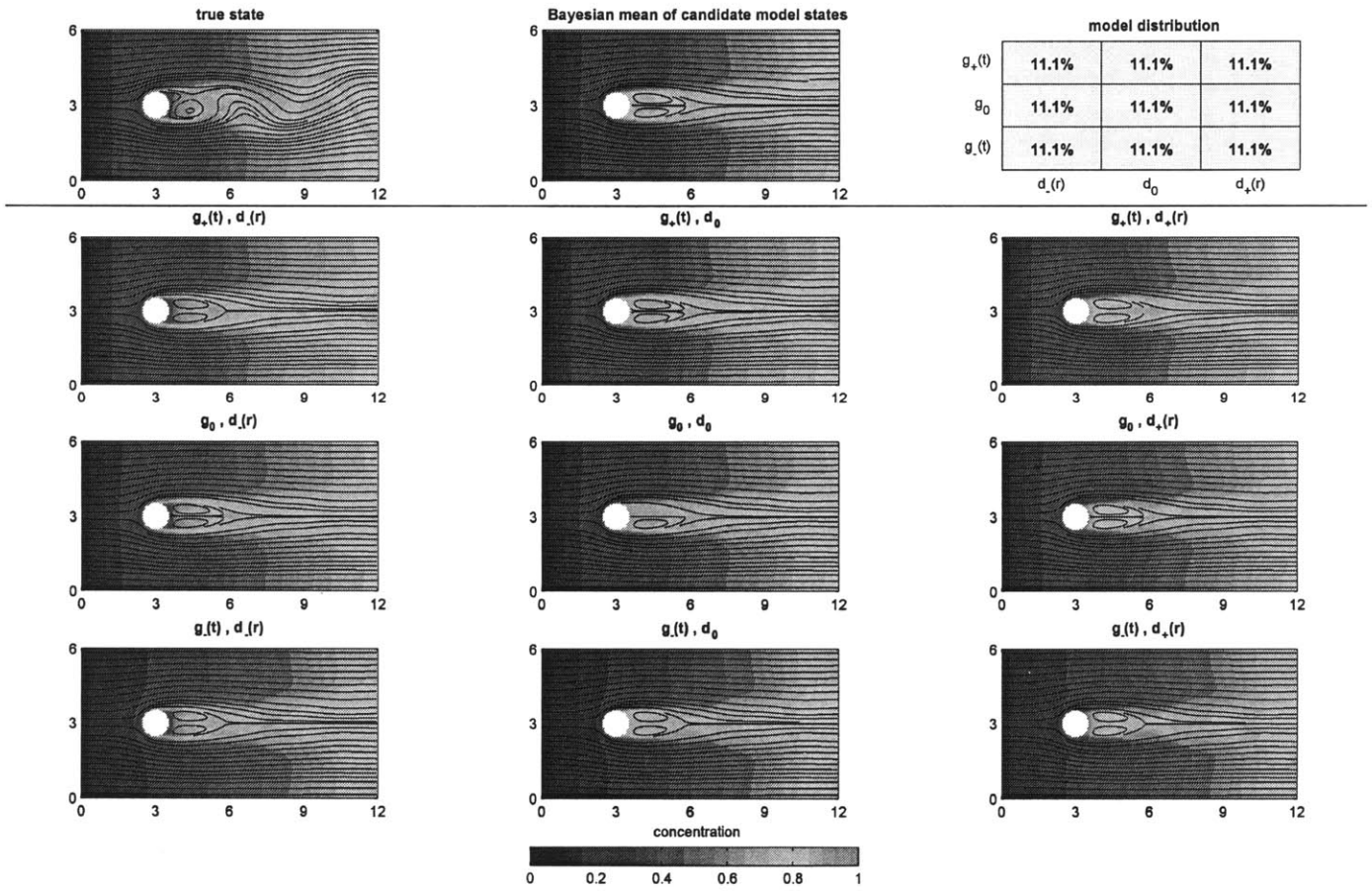

Figure 5-17: As Figure 5-5 but for the third experiment featuring the growth factor $g_+(t)$ and decay factor $d_-(\boldsymbol{r})$ formulations (Experiment B3).

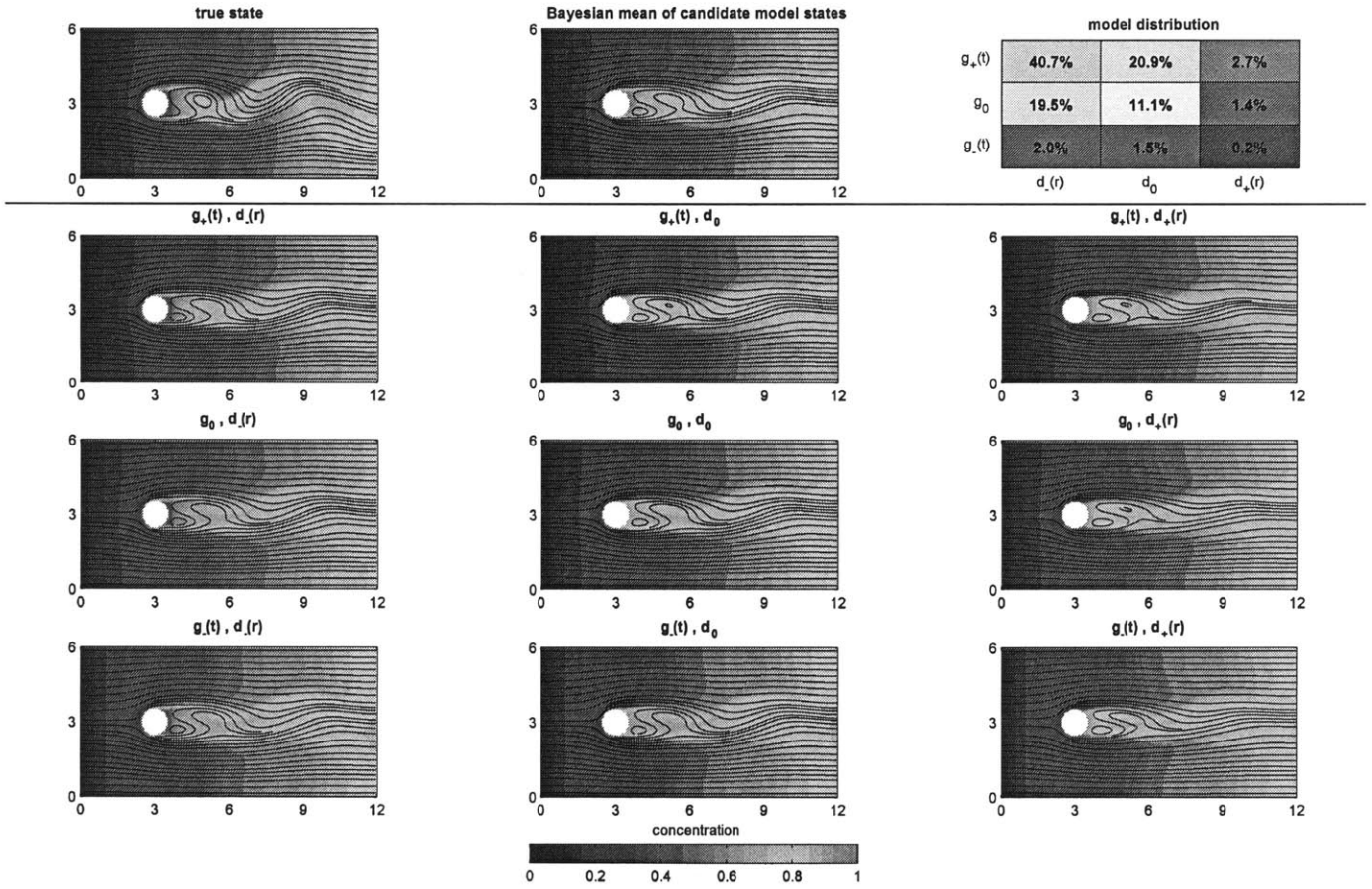

Figure 5-18: As Figure 5-17 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

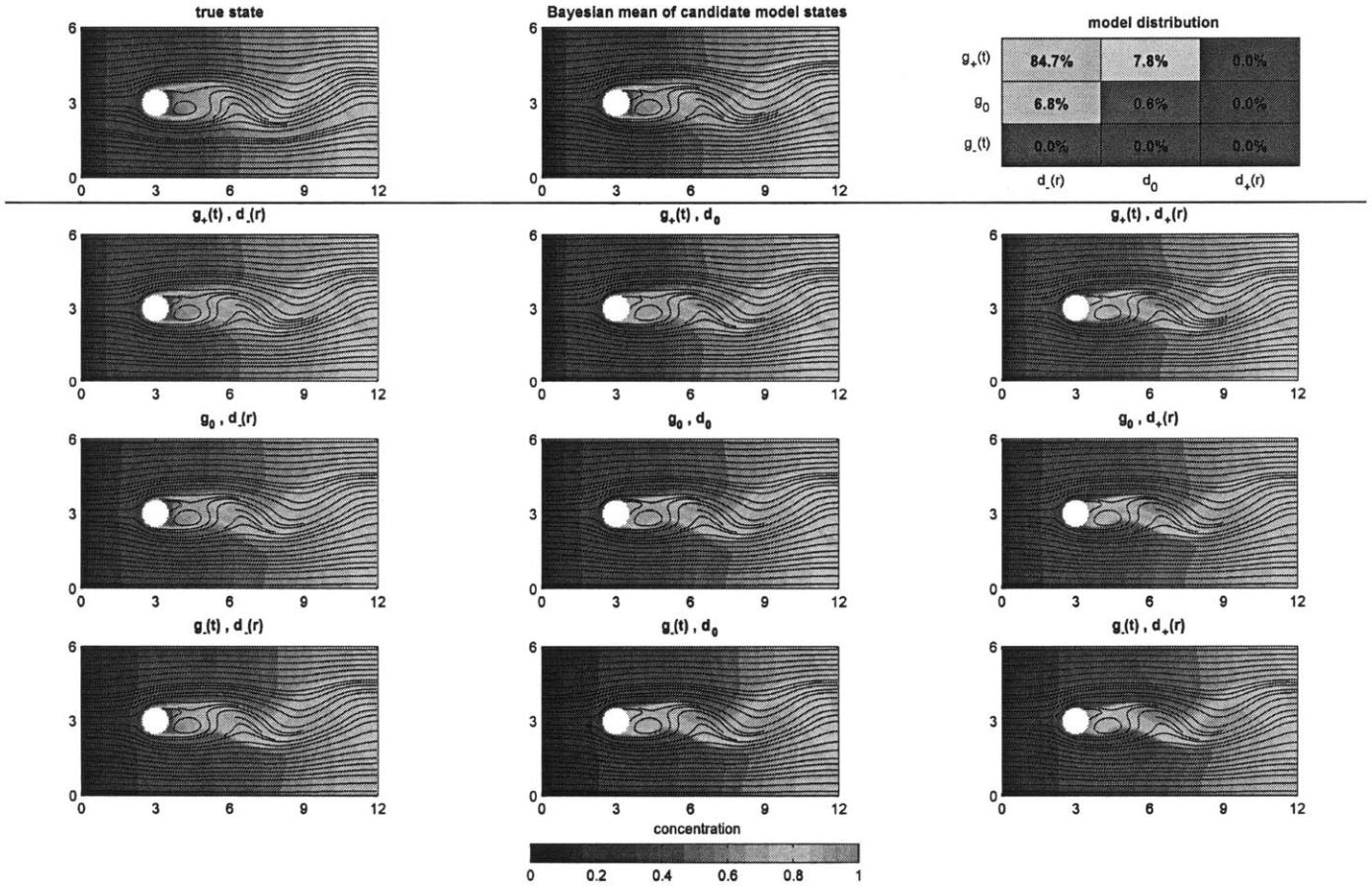

Figure 5-19: As Figure 5-17 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

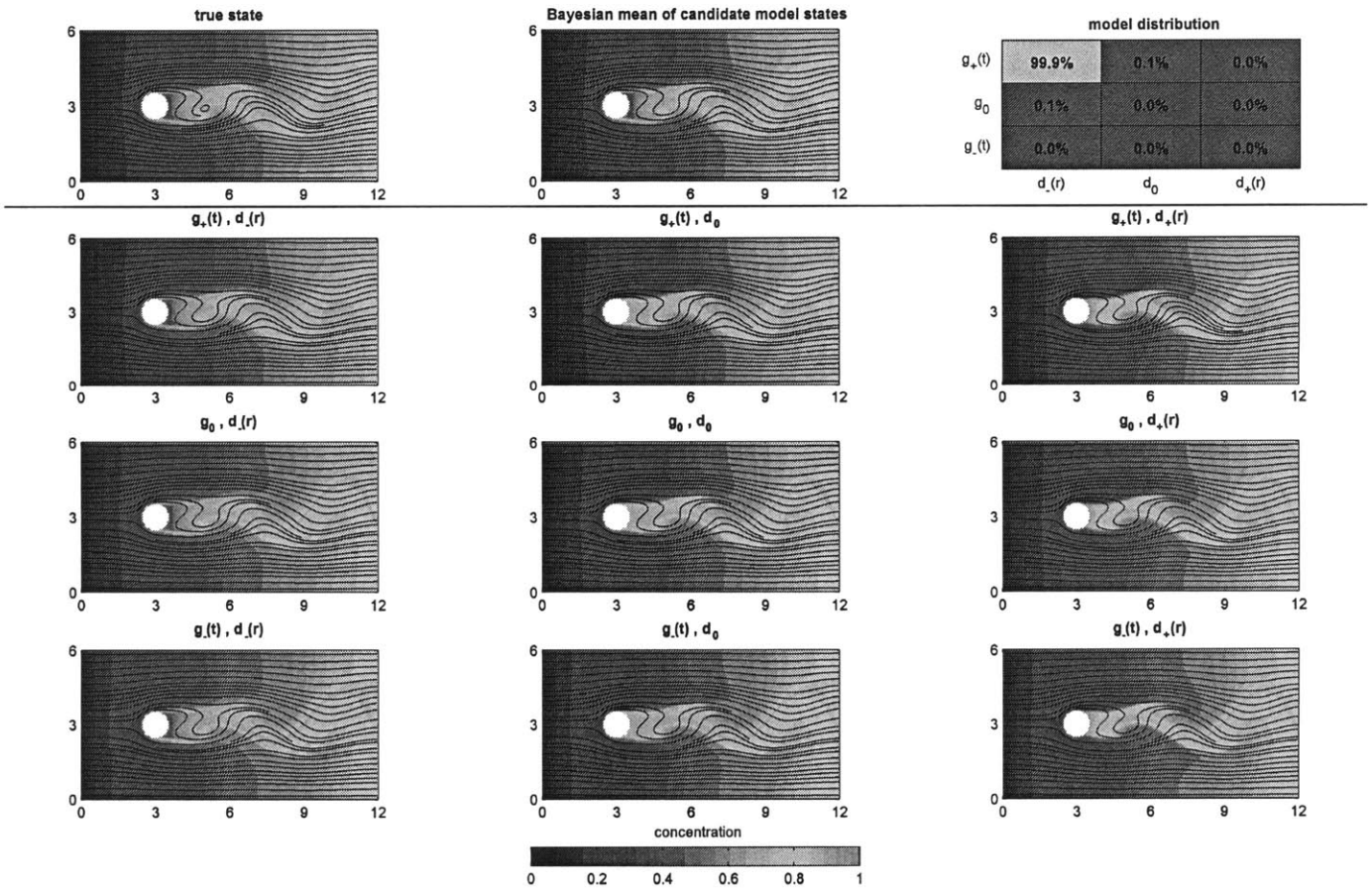

Figure 5-20: As Figure 5-17 but at non-dimensional time $t = 80$ (i.e. after 80 observation episodes).

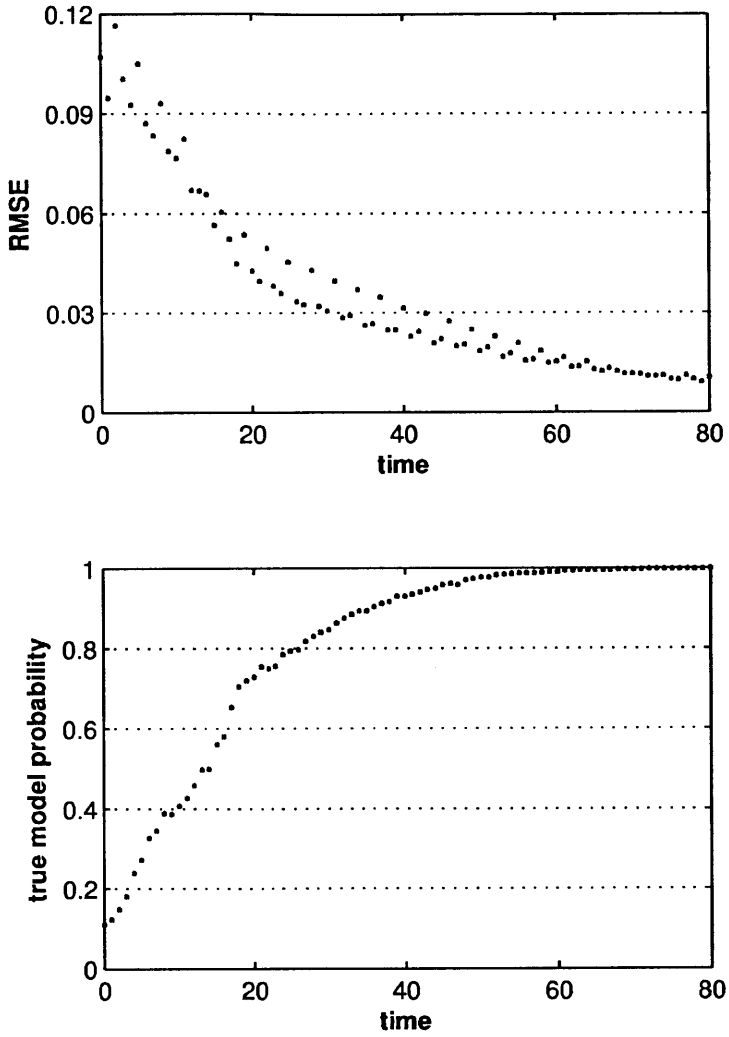

Figure 5-21: As Figure 5-11 but for the third experiment featuring the growth factor $g_+(t)$ and decay factor $d_-(\boldsymbol{r})$ formulations (Experiment B3).

5.4.4 Experiment B4: Intermediate Time-Dependent Growth and Spatially-Variable Decay

The fourth experiment we consider, Experiment B4, is that of a simulated truth defined by the deterministic simulation featuring the intermediate time-dependent growth and spatially-variable decay factor formulations $g'(t)$ and $d'(\boldsymbol{r})$ respectively. This combination of growth and decay factor formulations is not explicitly represented as one of the nine candidate models for the system, but is represented implicitly as an intermediary between the four candidate models formed by combining the two growth factor formulations $g_+(t)$ and g_0 and the two decay factor formulations d_0 and $d_-(\boldsymbol{r})$. The goal of this experiment is to find out if the formulation of the reaction terms (in general, their probabilities) can still be inferred jointly with the state of the velocity and microorganism concentration fields from noisy observations made from the simulated truth. The evolution of this inference is illustrated in Figures 5-22 to 5-25.

Figure 5-22 illustrates the system at the initial non-dimensional time ($t = 0$), before any observations are made. The fluid velocity and microorganism concentration fields of the deterministic simulated truth are again in a time-periodic state. As for the other three experiments considered above, the model-conditional velocity and concentration field means for all nine candidate models are symmetric about the $r_2 = 3$ centerline and the model distribution is uniform.

Figure 5-23 illustrates the system after 10 non-dimensional time units and 10 observation episodes. The model-conditional microorganism concentration field means for all nine candidate models are beginning to exhibit slight asymmetries, indicating partial inference. The model distribution is clearly favoring the candidate models that simultaneously feature one of the two growth factor forms $g_+(t)$ and g_0 and one of the two decay factor forms d_0 and $d_-(\boldsymbol{r})$—i.e. the four candidate models that encircle the true model. The probabilities of the other candidate models are falling quickly.

Figure 5-24 illustrates the system after 30 non-dimensional time units and 30

observation episodes. Agreement between the true velocity and concentration fields and the inferred velocity and concentration field means has improved significantly, though some concentration field features are still slightly blurred in the Bayesian mean. The model distribution now indicates that the only plausible models are the four candidate models that encircle the true model.

Figure 5-25 illustrates the system after 80 non-dimensional time units and 80 observation episodes. The inferred velocity and concentration field means are now consistent with the true velocity and concentration fields. No clear favorite amongst the four encircling candidate models has emerged, strongly suggesting that the true model is an intermediary.

Figure 5-26 illustrates the time progression of the Bayesian state variable mean RMSE and the probabilities of the four candidate models most dynamically similar to the true model for this experiment. Surprisingly, the RMSE still approaches 0, even though the true model of the system is not explicitly represented by a candidate model. This indicates that the Bayesian model inference methodology developed in Chapter 3 is at least partially robust to inadequacies in the initial representation of a system's model uncertainty. The four model probabilities all float about 0.25 (uniform distribution), with no one showing persistent dominance over the others. The non-monotonic nature of the model probability time-series is likely attributable to non-linear interactions between the experiment's various periodic features: vortex shedding downstream of the obstacle (with a period of approximately 6.0 non-dimensional time units), the time-dependent growth factor formulation (with time constant $T_\rho = 0.7$), and temporally-discrete observations (with a frequency of 1 non-dimensional time unit).

The above results indicate that interpolation within the functional space of model formulations is possible. To further refine the Bayesian learning, the inference would need to fully account for the possibility of intermediary models represented by varied combinations of growth and decay factors. An approach to do so is to render A_T and A_L appearing in $g_+(t)$, $g_-(t)$, $d_+(\boldsymbol{r})$, and $d_-(\boldsymbol{r})$ uncertain. Specifically, defining initial uniform distributions between -1 and 1 for both A_T and A_L would allow for

the encapsulation of all nine explicit reaction equations considered above along with their intermediaries in the general reaction equation

$$F_\rho(\rho, \boldsymbol{r}, t; \omega) = \Big[\left(1 + A_T(\omega) \sin(t/T_\rho)\right) G \Big] \frac{\rho(\rho_{\max} - \rho)}{\rho_{\max} - \rho + k_g} - \Big[\left(1 + A_L(\omega) \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D \Big] \rho \, . \tag{5.12}$$

This general reaction equation would then represent a single 'super' candidate model for the stochastic microorganism tracer system, with other candidates featuring different functional forms for the growth and decay terms in (5.12)—i.e. alternatives to

$$f_g(\rho, \boldsymbol{r}, t; \omega) = \frac{\rho(\rho_{\max} - \rho)}{\rho_{\max} - \rho + k_g}$$

and

$$f_d(\rho, \boldsymbol{r}, t; \omega) = \rho \, .$$

In essence, this approach transforms the inference of formulations using a limited set of candidate models into an infinite-dimensional parameter estimation inference. This is possible if the different candidate models are compatible (at least numerically) such that they can be combined (e.g. linearly added) into a coherent all-encompassing model formulation with stochastic parameters. Further discussion of the utility and economy of such expansions of the model space is provided in Chapter 6.

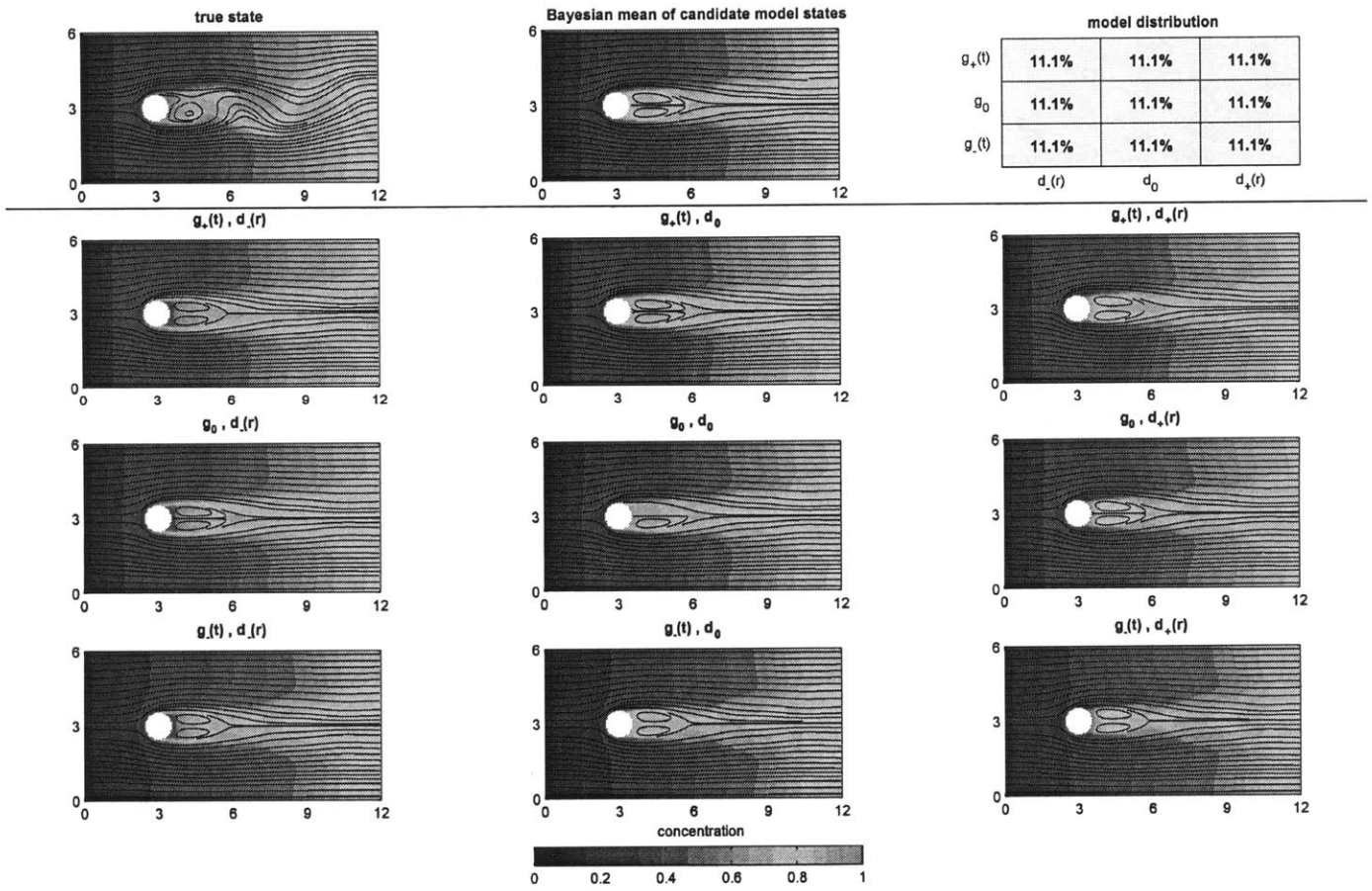

Figure 5-22: As Figure 5-5 but for the fourth experiment featuring the growth factor $g'(t)$ and decay factor $d'(\boldsymbol{r})$ formulations (Experiment B4).

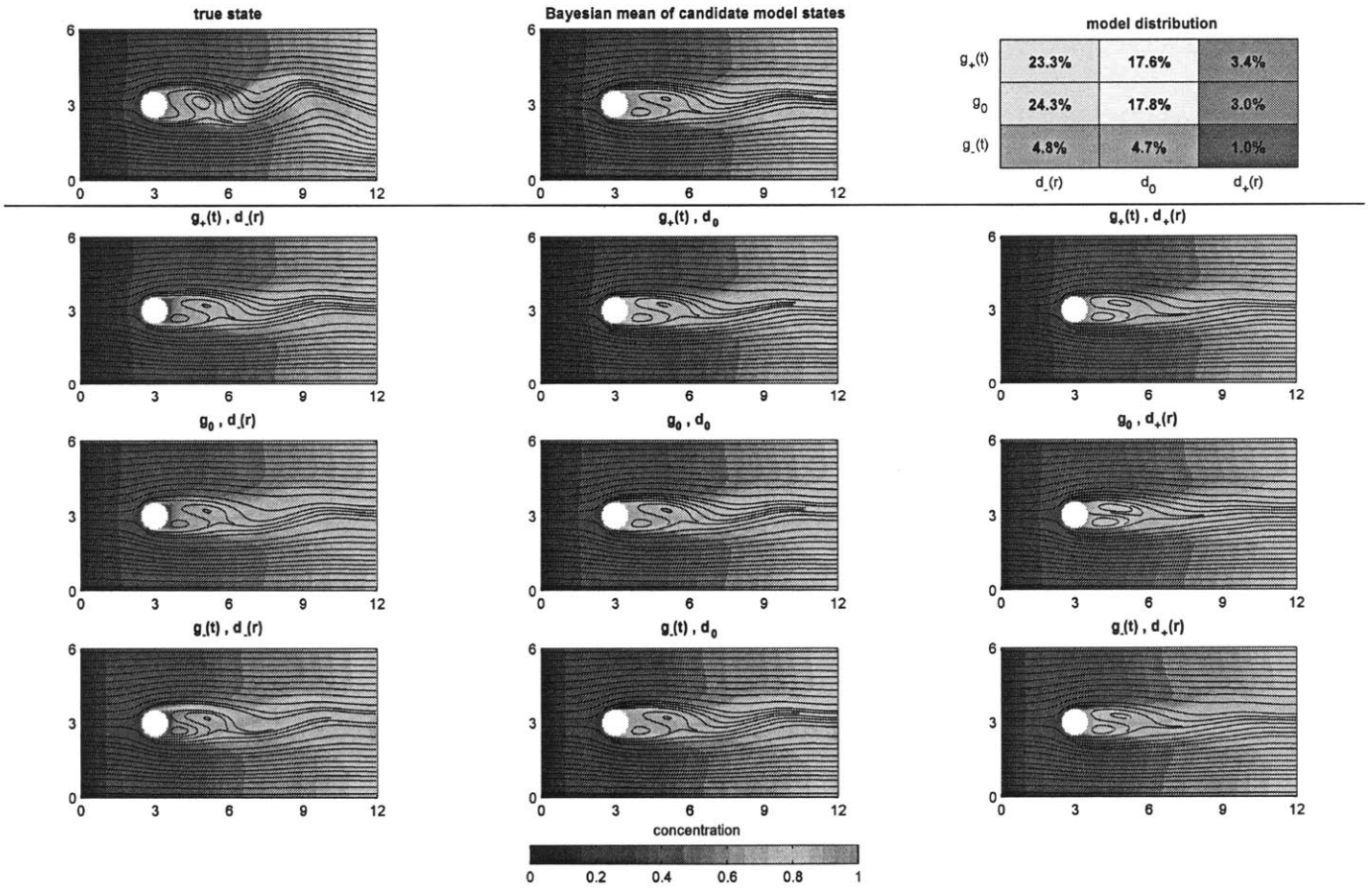

Figure 5-23: As Figure 5-22 but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

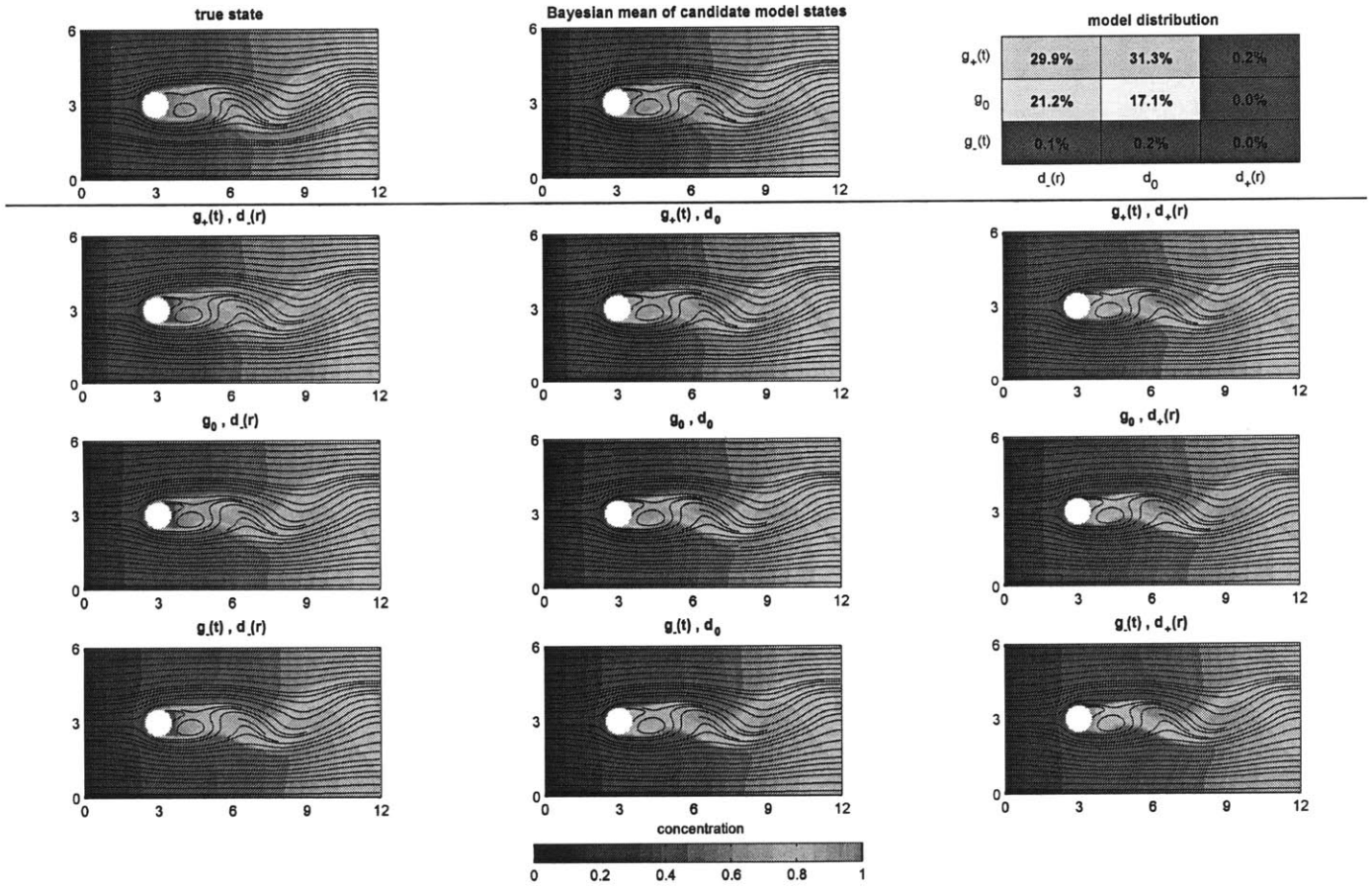

Figure 5-24: As Figure 5-22 but at non-dimensional time $t = 30$ (i.e. after 30 observation episodes).

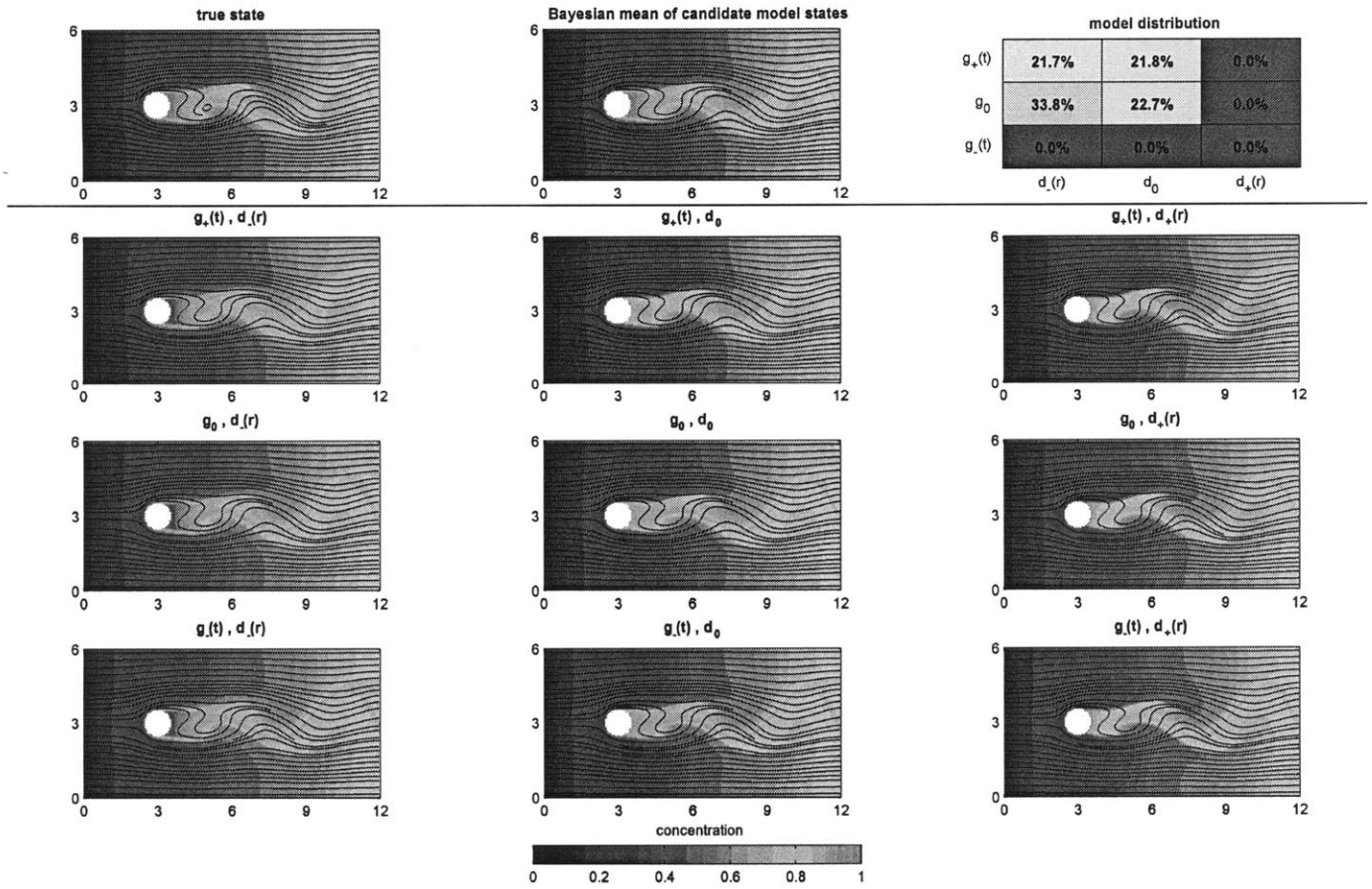

Figure 5-25: As Figure 5-22 but at non-dimensional time $t = 80$ (i.e. after 80 observation episodes).

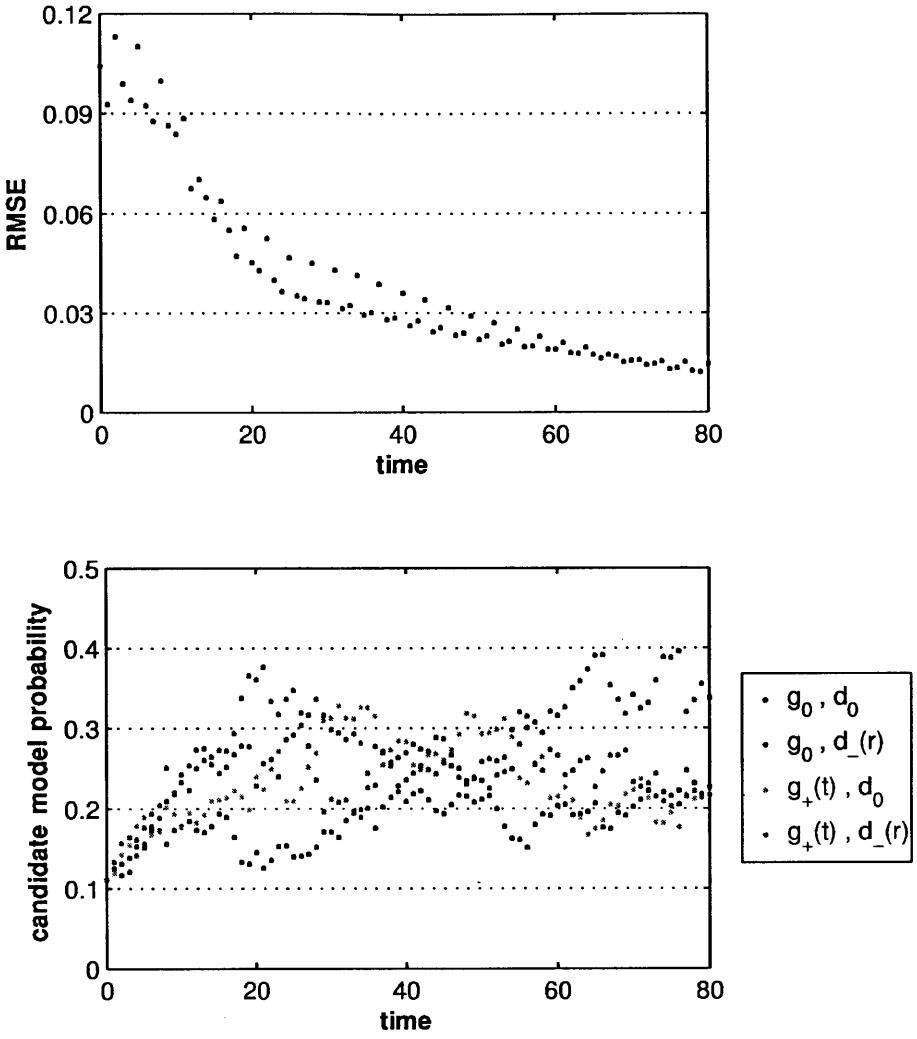

Figure 5-26: As Figure 5-11 but for the fourth experiment featuring the growth factor $g'(t)$ and decay factor $d'(\boldsymbol{r})$ formulations (Experiment B4). Since the true model for this experiment is not explicitly represented within the set of candidate models, the probabilities of the four candidates that are most dynamically similar to the true model are plotted in place of the true model probability.

Chapter 6

Discussion

6.1 Adaptive Modeling

For the two classes of simulated experiments presented so far, the candidate model sets for the systems were assumed to be immutable. The candidate models formulated for the systems at initialization were the only ones ever assessed for validity. However, this does not need to be the case and our approach can be extended to variable model formulations within the context of adaptive modeling [85].

To summarize the results presented from this perspective, for the two experiments conducted with the stochastic flow past an obstacle system (Experiments A1 and A2) and for the first three experiments conducted with the stochastic microorganism tracer system (Experiments B1 to B3), one of the candidate models was the true model of the system. The Bayesian model inference methodology developed in Chapter 3 was able to correctly infer the true models in each of these cases. For the fourth experiment conducted with the stochastic microorganism tracer system however (Experiment B4), the set of formulated candidate models did not explicitly include the true model for the system. The result of our Bayesian model inference methodology was then to assign higher probabilities to candidate models that have formulations likely to resemble that of the true model; in other words, the methodology found an implicit representation of the true model using the available candidate models. To allow further and more precise inference in that scenario, we would need to allow

the evolution of the candidate model set itself. The model distribution would then be updated through inference and new models would be considered as the inference progresses: this process can be referred to as adaptive modeling.

Adaptive modeling can be viewed as the intelligent allocation of computational resources over the space of stochastic dynamical models. An ideal accounting of model uncertainty would assign finite probability density to all plausible models at all times, an approach that is theoretically in line with the epistemological tenets of science but is practically infeasible for the vast majority of systems. Given a computational budget, it is sensible to limit computational resources to high probability regions of the model space. Consequently, as inference proceeds with a finite set of candidate models, low probability candidates should be replaced with new candidates that are in the vicinity of high probability candidates, a recursive process that could be performed either automatically or in conjunction with a human subject matter expert. When little information is available regarding the stochastic dynamical model governing a particular system, such a recursion may make it unnecessary to explicitly represent the entirety of the plausible model space using the initial set of candidate models, a computationally burdensome approach. It may instead be sufficient to use the initial candidates to sparsely populate the highest probability density regions of the model space and rely on the adaptive modeling scheme to refine the candidate model set as inference proceeds. One approach to achieve this consists of defining classes of (non-compatible) candidate models and then, for each class of (compatible) models, utilizing principal components in the model functional space and adaptive parameter estimation of their coefficients. This direct extension of what we have presented was touched upon at the end of Chapter 5. In general, the fundamental trade-offs between computational expediency and explicitness of model uncertainty representation, especially when the possibility of uncertain parameters is considered, are worthy of further study.

6.2 Adaptive Sampling

For both the simulated systems considered in the present work, the locations of the observations used for inference were assumed to be fixed within the spatial domains of the systems. For many real-world stochastic dynamical systems however, several alternative sets of observations are commonly available, which correspond to several possible choices for the observation likelihood function $\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})$ defined in (1.3). In oceanographic settings for example, mobile observation platforms typically enable a virtually countless number of options for the spatial and temporal configuration of observations [54]. The selection of a particular observation set is usually guided by a combination of practical accessibility, economic cost, and, critically, the ability of the target observations to reduce uncertainty in the system under investigation. The process of choosing observations based on an uncertainty reduction criterion is known as adaptive sampling [26, 85].

One measure of the uncertainty in a stochastic dynamical system is the differential entropy of the system's state vector

$$\mathcal{H}\,[\,\boldsymbol{X}\,] \equiv -\int p_{\boldsymbol{X}}(\boldsymbol{x}) \log p_{\boldsymbol{X}}(\boldsymbol{x})\; \mathrm{d}\boldsymbol{x}\;. \tag{6.1}$$

If the system features model uncertainty, the state variable distribution in (6.1) is taken to be the state variable distribution from the Bayesian mean of the model-conditional distributions (1.5), in which case

$$\mathcal{H}\,[\,\boldsymbol{X}\,] = -\int \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n) \log \left(\sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x}\,|\,\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n) \right) \mathrm{d}\boldsymbol{x}\;. \tag{6.2}$$

The value of a set of tentative observations can be quantified as the expected reduction in the differential entropy of the system's state vector

$$\mathrm{E}_{\boldsymbol{y}}\,[\,\mathcal{H}\,[\,\boldsymbol{X}\,|\,\boldsymbol{y}\,] - \mathcal{H}\,[\,\boldsymbol{X}\,]\,] = \mathrm{E}_{\boldsymbol{y}}\,[\,\mathcal{H}\,[\,\boldsymbol{X}\,|\,\boldsymbol{y}\,]\,] - \mathcal{H}\,[\,\boldsymbol{X}\,]\;, \tag{6.3}$$

where $\mathrm{E}_{\boldsymbol{y}}\,[\,\bullet\,]$ represents an expectation over all possible realizations of the observation

vector. Using an identity from information theory [128], we then have

$$\mathrm{E}_{\boldsymbol{y}}\left[\mathcal{H}\left[\boldsymbol{X}\,|\,\boldsymbol{y}\right]\right]-\mathcal{H}\left[\boldsymbol{X}\right]=\mathrm{E}_{\boldsymbol{x}}\left[\mathcal{H}\left[\boldsymbol{Y}\,|\,\boldsymbol{x}\right]\right]-\mathcal{H}\left[\boldsymbol{Y}\right]\;, \tag{6.4}$$

where $\mathrm{E}_{\boldsymbol{x}}\left[\bullet\right]$ represents an expectation over all possible realizations of the state vector. A complete derivation of (6.4) is provided in Appendix E.

The conditional differential entropy of the observation vector $\mathcal{H}\left[\boldsymbol{Y}\,|\,\boldsymbol{x}\right]$ can be found directly using the observation likelihood function associated with the target observations

$$\begin{aligned}\mathcal{H}\left[\boldsymbol{Y}\,|\,\boldsymbol{x}\right]&\equiv\int p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\log p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\;\mathrm{d}\boldsymbol{y}\\&=\int\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})\log\mathcal{L}(\boldsymbol{y}\,|\,\boldsymbol{x})\;\mathrm{d}\boldsymbol{y}\;.\end{aligned} \tag{6.5}$$

The differential entropy of the observation vector $\mathcal{H}\left[\boldsymbol{Y}\right]$ meanwhile needs to be found through

$$\mathcal{H}\left[\boldsymbol{Y}\right]\equiv\int p_{\boldsymbol{Y}}(\boldsymbol{y})\log p_{\boldsymbol{Y}}(\boldsymbol{y})\;\mathrm{d}\boldsymbol{y}\;. \tag{6.6}$$

For systems featuring model uncertainty, $p_{\boldsymbol{Y}}(\boldsymbol{y})$ in (6.6) can be found using (1.12), yielding

$$\mathcal{H}\left[\boldsymbol{Y}\right]=\int\sum_{n=1}^{N_{\mathcal{M}}}p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)\,p_{\mathcal{M}}(\mathcal{M}_n)\log\left(\sum_{n=1}^{N_{\mathcal{M}}}p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n)\,p_{\mathcal{M}}(\mathcal{M}_n)\right)\mathrm{d}\boldsymbol{y}\qquad\forall\boldsymbol{y}\in\mathbb{R}^{N_{\boldsymbol{Y}}}\;. \tag{6.7}$$

GMM representations of the conditional observation distributions appearing in (6.7) could be found using (3.31), the same equation used to perform the critical reduced-dimension marginal likelihood calculation for the Bayesian model inference methodology developed in Chapter 3. The linear combination of these GMM conditional observation distributions in (6.7) then renders the observation distribution $p_{\boldsymbol{Y}}(\bullet)$ a GMM distribution as well. Thus, the entropy of the observation vector $\mathcal{H}\left[\boldsymbol{Y}\right]$ could be found using one of the numerous approximation techniques available

for calculating the entropy of a random vector with a GMM distribution [63]. This potential adaptive sampling extension of the Bayesian model inference methodology developed in the present work is a promising topic for future work.

6.3 Conclusion

A new methodology for Bayesian inference of stochastic dynamical models was developed in the present work. The methodology was applied successfully to two high-dimensional, nonlinear simulated systems. Potential extensions of the methodology to incorporate both adaptive modeling and adaptive sampling are promising avenues for future investigation. Applications of the methodology to real systems are also certainly worthwhile endeavors. In the grand scheme, the present work represents a step towards the comprehensively adaptive investigation of stochastic dynamical systems envisioned in [85].

Appendix A

Convolution of Gaussian Distributions

Consider the convolution of two Gaussian distributions

$$\begin{aligned} p_{\boldsymbol{Y}}(\boldsymbol{y}) &= \Big[\mathcal{N}(\bullet;\, \boldsymbol{\mu}_1,\, \boldsymbol{\Sigma}_1) * \mathcal{N}(\bullet;\, \boldsymbol{\mu}_2,\, \boldsymbol{\Sigma}_2)\Big](\boldsymbol{y}) \\ &= \int \mathcal{N}(\boldsymbol{y}-\boldsymbol{x};\, \boldsymbol{\mu}_1,\, \boldsymbol{\Sigma}_1)\, \mathcal{N}(\boldsymbol{x};\, \boldsymbol{\mu}_2,\, \boldsymbol{\Sigma}_2)\, \mathrm{d}\boldsymbol{x}\,. \end{aligned} \tag{A.1}$$

Ignoring normalization constants, the product of the Gaussian distributions in (A.1) is

$$\begin{aligned} &\mathcal{N}(\boldsymbol{y}-\boldsymbol{x};\, \boldsymbol{\mu}_1,\, \boldsymbol{\Sigma}_1)\, \mathcal{N}(\boldsymbol{x};\, \boldsymbol{\mu}_2,\, \boldsymbol{\Sigma}_2) \\ &\quad \propto \exp\Big[-\frac{1}{2}\big[(\boldsymbol{y}-\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}-\boldsymbol{x}-\boldsymbol{\mu}_1) \\ &\qquad\qquad + (\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\big]\Big]\,. \end{aligned} \tag{A.2}$$

Let $W_1 = \Sigma_1^{-1}$, $W_2 = \Sigma_2^{-1}$, and $W_s = W_1 + W_2$ define symmetric matrices. Expanding the quadratic terms in the exponent in (A.2) and using symmetry,

$$
\begin{aligned}
&(\boldsymbol{y} - \boldsymbol{x} - \boldsymbol{\mu}_1)^{\mathrm{T}} \Sigma_1^{-1} (\boldsymbol{y} - \boldsymbol{x} - \boldsymbol{\mu}_1) + (\boldsymbol{x} - \boldsymbol{\mu}_2)^{\mathrm{T}} \Sigma_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2) \\
&\quad = (\boldsymbol{y} - \boldsymbol{\mu}_1)^{\mathrm{T}} W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + \boldsymbol{x}^{\mathrm{T}} W_1 \boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}} W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) \\
&\qquad + \boldsymbol{x}^{\mathrm{T}} W_2 \boldsymbol{x} + \boldsymbol{\mu}_2^{\mathrm{T}} W_2 \boldsymbol{\mu}_2 - 2\boldsymbol{x}^{\mathrm{T}} W_2 \boldsymbol{\mu}_2 \\
&\quad = \boldsymbol{x}^{\mathrm{T}} W_s \boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}} \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right) \\
&\qquad + (\boldsymbol{y} - \boldsymbol{\mu}_1)^{\mathrm{T}} W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2^{\mathrm{T}} W_2 \boldsymbol{\mu}_2 \\
&\quad = \Big[\boldsymbol{x}^{\mathrm{T}} W_s \boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}} \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right) \\
&\qquad\quad + \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right)^{\mathrm{T}} W_s^{-1} \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right) \Big] \\
&\qquad + \Big[(\boldsymbol{y} - \boldsymbol{\mu}_1)^{\mathrm{T}} W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2^{\mathrm{T}} W_2 \boldsymbol{\mu}_2 \\
&\qquad\quad - \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right)^{\mathrm{T}} W_s^{-1} \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right) \Big] \qquad \text{(A.3)} \\
&\quad = \Big[\mathbf{A} \Big] + \Big[\mathbf{B} \Big] , \qquad \text{(A.4)}
\end{aligned}
$$

where $\mathbf{A}$ and $\mathbf{B}$ have been defined so as to encapsulate the first three and last three terms in (A.3) respectively. Let

$$
\widehat{\boldsymbol{x}} = W_s^{-1} \left(W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + W_2 \boldsymbol{\mu}_2 \right) . \qquad \text{(A.5)}
$$

Then, $\mathbf{A}$ reduces to

$$
\begin{aligned}
\mathbf{A} &= (\boldsymbol{x} - \widehat{\boldsymbol{x}})^{\mathrm{T}} W_s (\boldsymbol{x} - \widehat{\boldsymbol{x}}) \\
&= (\boldsymbol{x} - \widehat{\boldsymbol{x}})^{\mathrm{T}} (\Sigma_1^{-1} + \Sigma_2^{-1}) (\boldsymbol{x} - \widehat{\boldsymbol{x}}) . \qquad \text{(A.6)}
\end{aligned}
$$

Expanding the terms in $\mathbf{B}$,

$$
\begin{aligned}
\mathbf{B} = &(\boldsymbol{y} - \boldsymbol{\mu}_1)^{\mathrm{T}} (W_1 + W_2) W_s^{-1} W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2^{\mathrm{T}} (W_1 + W_2) W_s^{-1} W_2 \boldsymbol{\mu}_2 \\
&- (\boldsymbol{y} - \boldsymbol{\mu}_1)^{\mathrm{T}} W_1^{\mathrm{T}} W_s^{-1} W_1 (\boldsymbol{y} - \boldsymbol{\mu}_1) - \boldsymbol{\mu}_2^{\mathrm{T}} W_2^{\mathrm{T}} W_s^{-1} W_2 \boldsymbol{\mu}_2 \\
&- 2(\boldsymbol{y} - \boldsymbol{\mu}_1)^{\mathrm{T}} W_1^{\mathrm{T}} W_s^{-1} W_2 \boldsymbol{\mu}_2 . \qquad \text{(A.7)}
\end{aligned}
$$

By the symmetry of W_1 and W_2, $\mathbf{B}$ reduces to

$$\begin{aligned}\mathbf{B} &= (\boldsymbol{y}-\boldsymbol{\mu}_1)^{\mathrm{T}} W_2 W_s^{-1} W_1 (\boldsymbol{y}-\boldsymbol{\mu}_1) + \boldsymbol{\mu}_2^{\mathrm{T}} W_1 W_s^{-1} W_2 \boldsymbol{\mu}_2 \\ &\quad - 2(\boldsymbol{y}-\boldsymbol{\mu}_1)^{\mathrm{T}} W_1^{\mathrm{T}} W_s^{-1} W_2 \boldsymbol{\mu}_2 \ . \end{aligned} \tag{A.8}$$

Using the identity $W_1(W_1+W_2)^{-1}W_2 = W_2(W_1+W_2)^{-1}W_1 = (W_1^{-1}+W_2^{-1})^{-1}$,

$$\begin{aligned}\mathbf{B} &= (\boldsymbol{y}-\boldsymbol{\mu}_1)^{\mathrm{T}} (W_1^{-1}+W_2^{-1})^{-1} (\boldsymbol{y}-\boldsymbol{\mu}_1) + \boldsymbol{\mu}_2^{\mathrm{T}} (W_1^{-1}+W_2^{-1})^{-1} \boldsymbol{\mu}_2 \\ &\quad - 2(\boldsymbol{y}-\boldsymbol{\mu}_1)^{\mathrm{T}} (W_1^{-1}+W_2^{-1})^{-1} \boldsymbol{\mu}_2 \\ &= (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}} (W_1^{-1}+W_2^{-1})^{-1} (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\ &= (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}} (\Sigma_1+\Sigma_2)^{-1} (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \ . \end{aligned} \tag{A.9}$$

Substituting (A.6) and (A.9) into (A.4),

$$\begin{aligned}&(\boldsymbol{y}-\boldsymbol{x}-\boldsymbol{\mu}_1)^{\mathrm{T}} \Sigma_1^{-1} (\boldsymbol{y}-\boldsymbol{x}-\boldsymbol{\mu}_1) + (\boldsymbol{x}-\boldsymbol{\mu}_2)^{\mathrm{T}} \Sigma_2^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_2) \\ &\quad = (\boldsymbol{x}-\widehat{\boldsymbol{x}})^{\mathrm{T}} (\Sigma_1^{-1}+\Sigma_2^{-1}) (\boldsymbol{x}-\widehat{\boldsymbol{x}}) \\ &\qquad + (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}} (\Sigma_1+\Sigma_2)^{-1} (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \ . \end{aligned} \tag{A.10}$$

The right-hand side of (A.10) represents the quadratic terms in the exponent of the product of two new Gaussian distributions

$$\begin{aligned}&\mathcal{N}\big(\boldsymbol{x}\,;\,\widehat{\boldsymbol{x}}\,,\,(\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}\big)\,\mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\,,\,\Sigma_1+\Sigma_2\big) \\ &\quad \propto \exp\Big[-\frac{1}{2}\big[(\boldsymbol{x}-\widehat{\boldsymbol{x}})^{\mathrm{T}} (\Sigma_1^{-1}+\Sigma_2^{-1}) (\boldsymbol{x}-\widehat{\boldsymbol{x}}) \\ &\qquad\qquad + (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathrm{T}} (\Sigma_1+\Sigma_2)^{-1} (\boldsymbol{y}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big]\Big] \ . \end{aligned} \tag{A.11}$$

Using (A.10), we can thus equate (A.2) and (A.11), up to a multiplicative constant, to obtain

$$\begin{aligned}&\mathcal{N}\big(\boldsymbol{y}-\boldsymbol{x}\,;\,\boldsymbol{\mu}_1\,,\,\Sigma_1\big)\,\mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_2\,,\,\Sigma_2\big) \\ &\quad = c\,\mathcal{N}\big(\boldsymbol{x}\,;\,\widehat{\boldsymbol{x}}\,,\,(\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}\big)\,\mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\,,\,\Sigma_1+\Sigma_2\big) \ , \end{aligned} \tag{A.12}$$

where c is a constant of proportionality. Substituting (A.12) into (A.1),

$$\begin{aligned} p_{\boldsymbol{Y}}(\boldsymbol{y}) &= \int c\mathcal{N}\big(\boldsymbol{x}\,;\,\widehat{\boldsymbol{x}}\,,\,(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})^{-1}\big)\,\mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\big)\,\mathrm{d}\boldsymbol{x} \\ &= c\mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\big)\int \mathcal{N}\big(\boldsymbol{x}\,;\,\widehat{\boldsymbol{x}}\,,\,(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})^{-1}\big)\,\mathrm{d}\boldsymbol{x} \\ &= c\mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\big)\,. \end{aligned} \tag{A.13}$$

Using (A.1) again,

$$\begin{aligned} \int p_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} &= \iint \mathcal{N}\big(\boldsymbol{y}-\boldsymbol{x}\,;\,\boldsymbol{\mu}_1\,,\,\boldsymbol{\Sigma}_1\big)\,\mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_2\big)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\ &= \iint \mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{x}+\boldsymbol{\mu}_1\,,\,\boldsymbol{\Sigma}_1\big)\,\mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_2\big)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\ &= \int \mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_2\big)\left(\int \mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{x}+\boldsymbol{\mu}_1\,,\,\boldsymbol{\Sigma}_1\big)\,\mathrm{d}\boldsymbol{y}\right)\mathrm{d}\boldsymbol{x} \\ &= \int \mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_2\big)\,\mathrm{d}\boldsymbol{x} \\ &= 1\,. \end{aligned} \tag{A.14}$$

Then, since the Gaussian distribution in (A.13) is equal to (A.1), and (A.1) itself integrates to unity, c must be equal to one. Therefore,

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathcal{N}\big(\boldsymbol{y}\,;\,\boldsymbol{\mu}_1+\boldsymbol{\mu}_2\,,\,\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\big)\,. \tag{A.15}$$

The convolution of two Gaussian distributions is thus another Gaussian distribution whose mean and variance are equal to the sums of the means and variances of the two component distributions respectively. The derivation presented here is an extension of the derivation presented in [93].

Appendix B

Derivation of the DO Evolution Equations

Starting from the reduced dimensionality stochastic subspace expansion

$$\boldsymbol{x}(t;\omega) \approx \overline{\boldsymbol{x}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\, \tilde{\boldsymbol{x}}_i(t) \ , \tag{B.1}$$

the stochastic dynamical model

$$\frac{\mathrm{d}\boldsymbol{x}(t;\omega)}{\mathrm{d}t} = \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] \ , \tag{B.2}$$

and the DO condition for mode evolution

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \perp \mathbf{V}_{\mathrm{DO}} \Leftrightarrow \left\langle \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \, , \, \tilde{\boldsymbol{x}}_j(t) \right\rangle = 0 \quad \forall i,j \in \{1,\ldots,N_{\mathrm{DO}}\} \ , \tag{B.3}$$

the DO evolution equations

$$\frac{\mathrm{d}\overline{\boldsymbol{x}}(t)}{\mathrm{d}t} = \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] , \tag{B.4}$$

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N_{\mathrm{DO}}} \mathbf{C}^{-1}_{(i,j)}\, \mathcal{P}_{\mathbf{V}^{\perp}_{\mathrm{DO}}}\left[\mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right]$$

$$\forall i \in \{1,\ldots,N_{\mathrm{DO}}\} , \tag{B.5}$$

$$\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t} = \left\langle \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] - \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] , \tilde{\boldsymbol{x}}_i(t) \right\rangle$$

$$\forall i \in \{1,\ldots,N_{\mathrm{DO}}\} , \tag{B.6}$$

which govern the evolution of the state vector mean, modes, and mode coefficients in (B.1) respectively can be derived.

Substituting (B.1) into (B.2),

$$\frac{\mathrm{d}\overline{\boldsymbol{x}}(t;\omega)}{\mathrm{d}t} + \sum_{i=1}^{N_{\mathrm{DO}}}\left[\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t}\,\tilde{\boldsymbol{x}}_i(t) + \phi_i(t;\omega)\,\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\right] = \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] . \tag{B.7}$$

To derive the governing equation for the mean (B.4), we take expectations of (B.7) to obtain

$$\frac{\mathrm{d}\overline{\boldsymbol{x}}(t;\omega)}{\mathrm{d}t} + \sum_{i=1}^{N_{\mathrm{DO}}}\left[\underbrace{\mathrm{E}\left[\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t}\right]}_{0}\tilde{\boldsymbol{x}}_i(t) + \underbrace{\mathrm{E}\left[\phi_i(t;\omega)\right]}_{0}\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\right]$$

$$= \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] . \tag{B.8}$$

Since $\phi_i(t;\omega)$ are zero-mean stochastic processes, the expectations on the left-hand side are equal to zero and (B.8) reduces to (B.4), the DO evolution equation for the state vector mean.

To derive the governing equations for the modes (B.5), we multiply (B.7) by the

jth mode coefficient,

$$\phi_j(t;\omega)\,\frac{\mathrm{d}\overline{\boldsymbol{x}}(t;\omega)}{\mathrm{d}t}+\sum_{i=1}^{N_{\mathrm{DO}}}\left[\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t}\,\phi_j(t;\omega)\,\tilde{\boldsymbol{x}}_i(t)+\phi_i(t;\omega)\phi_j(t;\omega)\,\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\right] = \phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\ , \quad \text{(B.9)}$$

then taking expectations,

$$\underbrace{\mathrm{E}\left[\,\phi_j(t;\omega)\,\right]}_{0}\,\frac{\mathrm{d}\overline{\boldsymbol{x}}(t;\omega)}{\mathrm{d}t}+\sum_{i=1}^{N_{\mathrm{DO}}}\left[\mathrm{E}\left[\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t}\,\phi_j(t;\omega)\right]\tilde{\boldsymbol{x}}_i(t)+\mathrm{E}\left[\,\phi_i(t;\omega)\phi_j(t;\omega)\,\right]\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\right] = \mathrm{E}\left[\,\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\ , \quad \text{(B.10)}$$

where the first expectation on the left-hand side is again equal to zero because $\phi_j(t;\omega)$ is a zero-mean stochastic process. Taking the inner product of (B.10) with the kth mode,

$$\sum_{i=1}^{N_{\mathrm{DO}}}\left[\mathrm{E}\left[\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t}\,\phi_j(t;\omega)\right]\underbrace{\left\langle\,\tilde{\boldsymbol{x}}_i(t)\,,\,\tilde{\boldsymbol{x}}_k(t)\,\right\rangle}_{\delta_{ik}}+\mathrm{E}\left[\,\phi_i(t;\omega)\phi_j(t;\omega)\,\right]\underbrace{\left\langle\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\,,\,\tilde{\boldsymbol{x}}_k(t)\right\rangle}_{0}\right] = \left\langle\,\mathrm{E}\left[\,\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\,,\,\tilde{\boldsymbol{x}}_k(t)\,\right\rangle\ , \quad \text{(B.11)}$$

where the first inner product on the left-hand side is equal to the Kronecker delta δ_{ik} due to the orthonormality of the modes and the second is equal to zero due to the DO condition (B.3). (B.11) then reduces to

$$\mathrm{E}\left[\frac{\mathrm{d}\phi_k(t;\omega)}{\mathrm{d}t}\,\phi_j(t;\omega)\right] = \left\langle\,\mathrm{E}\left[\,\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\,,\,\tilde{\boldsymbol{x}}_k(t)\,\right\rangle\ . \quad \text{(B.12)}$$

Switching k indices for i indices and substituting back into (B.10),

$$\sum_{i=1}^{N_{\mathrm{DO}}} \left[\langle \mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] , \tilde{\boldsymbol{x}}_i(t)\rangle \, \tilde{\boldsymbol{x}}_i(t) + \mathbf{C}_{(i,j)} \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \right] = \mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] , \quad \text{(B.13)}$$

where the definition of the mode coefficient covariance matrix

$$\mathbf{C}_{(i,j)} = \mathrm{E}\left[\phi_i(t;\omega)\phi_j(t;\omega)\right] \quad \forall i,j \in \{1,\ldots,N_{\mathrm{DO}}\}$$

has been used for the second expectation in (B.10). Rearranging,

$$\begin{aligned}
\sum_{i=1}^{N_{\mathrm{DO}}} \mathbf{C}_{(i,j)} \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} &= \mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] \\
&\quad - \sum_{i=1}^{N_{\mathrm{DO}}} \langle \mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] , \tilde{\boldsymbol{x}}_i(t)\rangle \, \tilde{\boldsymbol{x}}_i(t) \\
&= \mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] - \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}}\left[\mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] \\
&= \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathrm{E}\left[\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] . \qquad \text{(B.14)}
\end{aligned}$$

In matrix form,

$$\mathbf{C} \begin{bmatrix} \frac{\mathrm{d}\tilde{\boldsymbol{x}}_1(t)}{\mathrm{d}t} \\ \vdots \\ \frac{\mathrm{d}\tilde{\boldsymbol{x}}_{N_{\mathrm{DO}}}(t)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathrm{E}\left[\phi_1(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] \\ \vdots \\ \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathrm{E}\left[\phi_{N_{\mathrm{DO}}}(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] \end{bmatrix} . \quad \text{(B.15)}$$

Note that the elements of the vectors in (B.15) are themselves vectors. Since the mode coefficient covariance matrix is necessarily positive definite, it is always invertible

$$\begin{bmatrix} \frac{\mathrm{d}\tilde{\boldsymbol{x}}_1(t)}{\mathrm{d}t} \\ \vdots \\ \frac{\mathrm{d}\tilde{\boldsymbol{x}}_{N_{\mathrm{DO}}}(t)}{\mathrm{d}t} \end{bmatrix} = \mathbf{C}^{-1} \begin{bmatrix} \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathrm{E}\left[\phi_1(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] \\ \vdots \\ \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}\left[\mathrm{E}\left[\phi_{N_{\mathrm{DO}}}(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] \end{bmatrix} . \quad \text{(B.16)}$$

The elements of (B.16) are equivalent to (B.5), the DO evolution equations for the modes.

To derive the governing equations for the mode coefficients (B.6), we take the inner product of (B.7) with the jth mode,

$$\left\langle \frac{\mathrm{d}\overline{\boldsymbol{x}}(t;\omega)}{\mathrm{d}t} \, , \, \tilde{\boldsymbol{x}}_j(t;\omega) \right\rangle + \sum_{i=1}^{N_{\mathrm{DO}}} \left[\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t} \underbrace{\left\langle \tilde{\boldsymbol{x}}_i(t) \, , \, \tilde{\boldsymbol{x}}_j(t) \right\rangle}_{\delta_{ij}} + \phi_i(t;\omega) \underbrace{\left\langle \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \, , \, \tilde{\boldsymbol{x}}_j(t) \right\rangle}_{0} \right] = \left\langle \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] \, , \, \tilde{\boldsymbol{x}}_j(t;\omega) \right\rangle \, , \quad \text{(B.17)}$$

where, as for (B.11), the second inner product on the left-hand side is equal to the Kronecker delta δ_{ij} due to the orthonormality of the modes and the third is equal to zero due to the DO condition (B.3). Substituting (B.4),

$$\left\langle \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] \, , \, \tilde{\boldsymbol{x}}_j(t;\omega) \right\rangle + \frac{\mathrm{d}\phi_j(t;\omega)}{\mathrm{d}t} = \left\langle \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] \, , \, \tilde{\boldsymbol{x}}_j(t;\omega) \right\rangle \, , \quad \text{(B.18)}$$

and rearranging,

$$\frac{\mathrm{d}\phi_j(t;\omega)}{\mathrm{d}t} = \left\langle \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] - \mathrm{E}\left[\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right] \, , \, \tilde{\boldsymbol{x}}_j(t;\omega) \right\rangle \, . \quad \text{(B.19)}$$

Switching j indices for i indices yields (B.6), the DO evolution equations for the mode coefficients.

Appendix C

Subspace Equivalency in the GMM-DO Filter

If the prior probability distribution for state variables is represented by a GMM distribution and the observation likelihood function is Gaussian (as in (3.16)), the posterior state variable distribution is given by

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) = \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{X}|\boldsymbol{Y},j} \times \mathcal{N}\big(\boldsymbol{x}\,;\,\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j}\,,\,\boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j}\big) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,, \tag{C.1}$$

where

$$\pi_{\boldsymbol{X}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{X},j} \times \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},j}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\big)}{\sum_{k=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{X},k} \times \mathcal{N}\big(\boldsymbol{y}\,;\,\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},k}\,,\,\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},k}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\big)}\,, \tag{C.2}$$

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j} = \boldsymbol{\mu}_{\boldsymbol{X},j} + \mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},j}\right)\,, \tag{C.3}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j} = \left(\mathbf{I} - \mathbf{K}_j\mathbf{H}\right)\boldsymbol{\Sigma}_{\boldsymbol{X},j} \tag{C.4}$$

$$\forall j \in \{1,\ldots,N_{\text{GMM}}\}$$

and

$$\mathbf{K}_j = \boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}}\big(\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\big)^{-1} \quad \forall j \in \{1,\ldots,N_{\text{GMM}}\}\,. \tag{C.5}$$

The equivalent posterior GMM coefficient distribution is proposed to be

$$p_{\boldsymbol{\Phi}|\boldsymbol{Y}}(\boldsymbol{\phi}|\boldsymbol{y}) = \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} \times \mathcal{N}\left(\boldsymbol{\phi}\,;\, \boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\right) \quad \forall \boldsymbol{\phi} \in \mathbb{R}^{N_{\text{DO}}}\,, \tag{C.6}$$

where

$$\pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{\Phi},j} \times \mathcal{N}\left(\tilde{\boldsymbol{y}}\,;\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\,,\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\right)}{\sum_{k=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi},k} \times \mathcal{N}\left(\tilde{\boldsymbol{y}}\,;\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},k}\,,\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},k}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\right)}\,, \tag{C.7}$$

$$\boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j} - \sum_{k=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \times \boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},k}{}'\,, \tag{C.8}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \left(\mathbf{I} - \tilde{\mathbf{K}}_j\tilde{\mathbf{H}}\right)\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j} \tag{C.9}$$

$$\forall j \in \{1, \ldots, N_{\text{GMM}}\}$$

and

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}}\,, \tag{C.10}$$

$$\tilde{\mathbf{H}} = \mathbf{H}\boldsymbol{\mathcal{X}}\,, \tag{C.11}$$

$$\tilde{\mathbf{K}}_j = \boldsymbol{\mathcal{X}}^{\mathrm{T}}\mathbf{K}_j \quad \forall j \in \{1, \ldots, N_{\text{GMM}}\}\,, \tag{C.12}$$

$$\boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \tilde{\mathbf{K}}_j\left(\tilde{\boldsymbol{y}} - \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right) \quad \forall j \in \{1, \ldots, N_{\text{GMM}}\}\,. \tag{C.13}$$

In [129], the authors show that (C.1) is equivalent to (C.6) through the affine transformation

$$\boldsymbol{x}(t;\omega) = \overline{\boldsymbol{x}}(t) + \boldsymbol{\mathcal{X}}(t)\boldsymbol{\phi}(t;\omega) \tag{C.14}$$

given a concurrent state vector mean update

$$\overline{\boldsymbol{x}}(t^+) = \overline{\boldsymbol{x}}(t^-) + \boldsymbol{\mathcal{X}} \sum_{k=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \times \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \tag{C.15}$$

by deriving (C.6) starting from (C.1). Here, we take a slightly different approach to demonstrate the same. We first propose an intermediate mode coefficient distribution:

$$p_{\boldsymbol{\Phi}|\boldsymbol{Y}}(\boldsymbol{\phi}|\boldsymbol{y})' = \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} \times \mathcal{N}\big(\boldsymbol{\phi}\,;\, \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\big) \quad \forall \boldsymbol{\phi} \in \mathbb{R}^{N_{\text{DO}}}\,, \tag{C.16}$$

with parameters as defined above. Then we note that due to the affine transformation (C.14), to demonstrate the equivalence between (C.1) and (C.16), it is sufficient to show that the parameters of the two GMMs are related through

$$\pi_{\boldsymbol{X}|\boldsymbol{Y},j} = \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,, \tag{C.17}$$

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j} = \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,, \tag{C.18}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j} = \boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \tag{C.19}$$

$$\forall j \in \{1, \ldots, N_{\text{GMM}}\}\,.$$

This can be accomplished by considering the analogous relations between the parameters of the prior GMM state variable distribution and the prior GMM mode coefficient distribution

$$\pi_{\boldsymbol{X},j} = \pi_{\boldsymbol{\Phi},j}\,, \tag{C.20}$$

$$\boldsymbol{\mu}_{\boldsymbol{X},j} = \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\,, \tag{C.21}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X},j} = \boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \tag{C.22}$$

$$\forall j \in \{1, \ldots, N_{\text{GMM}}\}\,.$$

Substituting (C.21) and (C.22) into (C.2),

$$\pi_{\boldsymbol{X}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{X},j} \times \mathcal{N}\big(\boldsymbol{y}\,;\, \mathbf{H}\left(\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right)\,,\, \mathbf{H}\left(\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big)}{\sum_{k=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{X},k} \times \mathcal{N}\big(\boldsymbol{y}\,;\, \mathbf{H}\left(\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},k}\right)\,,\, \mathbf{H}\left(\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},k}\boldsymbol{\mathcal{X}}^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}} + \mathbf{R}\big)}$$

$$= \frac{\pi_{\boldsymbol{X},j} \times \mathcal{N}\big(\boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}}\,;\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\,,\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\left(\mathbf{H}\boldsymbol{\mathcal{X}}\right)^{\mathrm{T}} + \mathbf{R}\big)}{\sum_{k=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{X},k} \times \mathcal{N}\big(\boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}}\,;\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},k}\,,\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},k}\left(\mathbf{H}\boldsymbol{\mathcal{X}}\right)^{\mathrm{T}} + \mathbf{R}\big)}\,. \tag{C.23}$$

Substituting (C.10) and (C.11) into (C.7),

$$\pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{X},j} \times \mathcal{N}\left(\boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}};\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j},\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\left(\mathbf{H}\boldsymbol{\mathcal{X}}\right)^{\mathrm{T}} + \mathbf{R}\right)}{\sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X},k} \times \mathcal{N}\left(\boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}};\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},k},\, \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},k}\left(\mathbf{H}\boldsymbol{\mathcal{X}}\right)^{\mathrm{T}} + \mathbf{R}\right)} \ . \quad \text{(C.24)}$$

The right-hand side of (C.23) is equal to the right-hand side of (C.24). Therefore, the posterior state and intermediate mode coefficient GMM component weights are related through (C.17).

Substituting (C.21) into (C.3),

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j} = \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\left(\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right)\right) \ . \quad \text{(C.25)}$$

Substituting (C.8) into the right-hand side of (C.18),

$$\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\left(\boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \tilde{\mathbf{K}}_j\left(\tilde{\boldsymbol{y}} - \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right)\right) \ . \quad \text{(C.26)}$$

Substituting (C.10), (C.11), and (C.12) into (C.26),

$$\begin{aligned}
\overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} &= \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\left(\boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \boldsymbol{\mathcal{X}}^{\mathrm{T}}\mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\overline{\boldsymbol{x}} - \mathbf{H}\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right)\right) \\
&= \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \underbrace{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^{\mathrm{T}}}_{\mathbf{I}}\mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right)\right) \\
&= \overline{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right)\right) \ , \quad \text{(C.27)}
\end{aligned}$$

where $\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^{\mathrm{T}}$ is equal to the identity matrix due to the orthonormality of the modes. The right-hand side of (C.25) is equal to the right-hand side of (C.27). Therefore, the posterior state and intermediate mode coefficient GMM component mean vectors are related through (C.18).

Substituting (C.22) into (C.4),

$$\boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j} = \left(\mathbf{I} - \mathbf{K}_j\mathbf{H}\right)\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \ . \quad \text{(C.28)}$$

Substituting (C.9) into the right-hand side of (C.19),

$$\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} = \boldsymbol{\mathcal{X}}\left(\mathbf{I} - \tilde{\mathbf{K}}_j\tilde{\mathbf{H}}\right)\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \ . \tag{C.29}$$

Substituting (C.11) and (C.12) into (C.29),

$$\begin{aligned}\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} &= \boldsymbol{\mathcal{X}}\left(\mathbf{I} - \boldsymbol{\mathcal{X}}^{\mathrm{T}}\mathbf{K}_j\mathbf{H}\boldsymbol{\mathcal{X}}\right)\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \\ &= \left(\boldsymbol{\mathcal{X}} - \underbrace{\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^{\mathrm{T}}}_{\mathbf{I}}\mathbf{K}_j\mathbf{H}\boldsymbol{\mathcal{X}}\right)\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \\ &= \left(\mathbf{I} - \mathbf{K}_j\mathbf{H}\right)\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \ ,\end{aligned} \tag{C.30}$$

where $\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^{\mathrm{T}}$ is again equal to the identity matrix due to the orthonormality of the modes. The right-hand side of (C.28) is equal to the right-hand side of (C.30). Therefore, the posterior state and intermediate mode coefficient GMM component covariance matrices are related through (C.19).

The intermediate GMM mode coefficient distribution (C.16) is thus equivalent to the posterior GMM state variable distribution (C.1) through the affine transformation (C.14). Applying the expectation operator to (C.16), the mean of the intermediate mode coefficient distribution is found to be

$$\begin{aligned}\mathrm{E}\left[\,\boldsymbol{\Phi}\,|\,\boldsymbol{y}\,\right]' &= \sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} \times \mathrm{E}\left[\,\mathcal{N}\left(\,\boldsymbol{\phi}\,;\,\boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,,\,\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\,\right)\,\right] \\ &= \sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} \times \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \ .\end{aligned} \tag{C.31}$$

Examining (C.8), we can then recognize that the proposed posterior mode coefficient distribution (C.6) is the zero-mean equivalent of the intermediate mode coefficient distribution (C.16). (C.6) is thus equivalent to (C.1) through the affine transformation (C.14) if the state vector mean is concurrently updated by (C.15). This validates the reduced dimensionality approach to Bayesian state variable inference employed by the GMM-DO filter.

Appendix D

Microorganism Reaction Equation Linearization

The numerical approach to solving the DO evolution equations (3.4)–(3.6) developed in [141] is most effective when the stochastic differential equations of the system under consideration are expressed as polynomials of the stochastic state variables. To achieve this condition for the stochastic microorganism tracer system, a local linearization of the non-polynomial microorganism reaction equation is performed, as suggested in [91].

Consider any single microorganism concentration state variable $\rho(t;\omega)$ within the concentration state variable vector $\boldsymbol{\rho}(t;\omega)$ for the stochastic microorganism tracer system. The expansion (3.1) dictates

$$\rho(t;\omega) = \overline{\rho}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\widetilde{\rho}_i(t) \ , \tag{D.1}$$

where $\overline{\rho}(t)$ represents the mean of the concentration state variable and $\widetilde{\rho}_i(t)$ represents the element in the ith mode corresponding to the concentration state variable.

Dropping temporal and stochastic indices and substituting into the reaction equa-

tion (5.2),

$$
\begin{aligned}
F_\rho(\rho) &= g\frac{(\overline{\rho}+\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i)(\rho_{\max}-\overline{\rho}-\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i)}{\rho_{\max}-\overline{\rho}-\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i+k_g} - d(\overline{\rho}+\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i) \\
&= g\frac{\overline{\rho}(\rho_{\max}-\overline{\rho})}{\rho_{\max}-\overline{\rho}+k_g-\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i} + g\frac{(\rho_{\max}-2\overline{\rho})\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i}{\rho_{\max}-\overline{\rho}+k_g-\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i} \\
&\quad - g\frac{(\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i)^2}{\rho_{\max}-\overline{\rho}+k_g-\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i} - d(\overline{\rho}+\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i)\ . \qquad \text{(D.2)}
\end{aligned}
$$

Let

$$
\varepsilon = \frac{\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i}{\rho_{\max}-\overline{\rho}+k_g}\ .
$$

Then, (D.2) becomes

$$
\begin{aligned}
F_\rho(\rho) &= g\frac{\overline{\rho}(\rho_{\max}-\overline{\rho})}{\rho_{\max}-\overline{\rho}+k_g}\frac{1}{1-\varepsilon} + g(\rho_{\max}-2\overline{\rho})\frac{\varepsilon}{1-\varepsilon} \\
&\quad - g(\rho_{\max}-\overline{\rho}+k_g)\frac{\varepsilon^2}{1-\varepsilon} - d(\overline{\rho}+(\rho_{\max}-\overline{\rho}+k_g)\varepsilon)\ . \qquad \text{(D.3)}
\end{aligned}
$$

Using a first-order Taylor approximation around $\varepsilon = 0$,

$$
\begin{aligned}
F_\rho(\rho) &\approx g\frac{\overline{\rho}(\rho_{\max}-\overline{\rho})}{\rho_{\max}-\overline{\rho}+k_g}(1+\varepsilon) + g(\rho_{\max}-2\overline{\rho})\varepsilon - d\overline{\rho} - d(\rho_{\max}-\overline{\rho}+k_g)\varepsilon \\
&= \left[g\frac{\overline{\rho}(\rho_{\max}-\overline{\rho})}{\rho_{\max}-\overline{\rho}+k_g} - d\overline{\rho}\right] \\
&\quad + \left[g\left(\rho_{\max}-2\overline{\rho}+\frac{\overline{\rho}(\rho_{\max}-\overline{\rho})}{\rho_{\max}-\overline{\rho}+k_g}\right) - d(\rho_{\max}-\overline{\rho}+k_g)\right]\varepsilon \\
&= F_\rho(\overline{\rho}) + \left[\frac{g}{\rho_{\max}-\overline{\rho}+k_g}\left(\rho_{\max}-2\overline{\rho}+\frac{\overline{\rho}(\rho_{\max}-\overline{\rho})}{\rho_{\max}-\overline{\rho}+k_g}\right) - d\right]\sum_{i=1}^{N_{\mathrm{DO}}}\phi_i\widetilde{\rho}_i\ . \qquad \text{(D.4)}
\end{aligned}
$$

(D.4) represents a linearized form of the microorganism reaction equation. It has been found that $\mathrm{E}[\varepsilon] < 0.1$ for all experiments conducted with the stochastic microorganism tracer system, justifying the first-order approximation employed in the derivation of (D.4). Experiments conducted with the linear term in (D.4) removed (i.e. using a zero-order approximation) showed minimal effects on the efficacy of the inference methodology presented in Chapter 5.

Appendix E

Conditional Entropy Identity

Starting from the expected value of the entropy of the posterior state variable distribution,

$$
\begin{aligned}
\mathrm{E}_{\boldsymbol{y}}\left[\mathcal{H}\left[\boldsymbol{X}\,|\,\boldsymbol{y}\right]\right] &= \int p_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathcal{H}\left[\boldsymbol{X}\,|\,\boldsymbol{y}\right]\,\mathrm{d}\boldsymbol{y} \\
&= \int p_{\boldsymbol{Y}}(\boldsymbol{y})\left(-\int p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\log p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\right)\mathrm{d}\boldsymbol{y} \\
&= -\iint p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,p_{\boldsymbol{Y}}(\boldsymbol{y})\log p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&= -\iint p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,p_{\boldsymbol{Y}}(\boldsymbol{y})\log\frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\,p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&= -\iint p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,p_{\boldsymbol{Y}}(\boldsymbol{y})\log p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&\quad -\iint p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,p_{\boldsymbol{Y}}(\boldsymbol{y})\log p_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&\quad +\iint p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,p_{\boldsymbol{Y}}(\boldsymbol{y})\log p_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&= -\iint p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\,p_{\boldsymbol{X}}(\boldsymbol{x})\log p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&\quad -\iint p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x})\,p_{\boldsymbol{X}}(\boldsymbol{x})\log p_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} \\
&\quad +\iint p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\,p_{\boldsymbol{Y}}(\boldsymbol{y})\log p_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
&= -\int p_{\boldsymbol{X}}(\boldsymbol{x}) \left(\int p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \log p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{y} \right) \mathrm{d}\boldsymbol{x} \\
&\quad - \int p_{\boldsymbol{X}}(\boldsymbol{x}) \log p_{\boldsymbol{X}}(\boldsymbol{x}) \left(\int p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{y} \right) \mathrm{d}\boldsymbol{x} \\
&\quad + \int p_{\boldsymbol{Y}}(\boldsymbol{y}) \log p_{\boldsymbol{Y}}(\boldsymbol{y}) \left(\int p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \right) \mathrm{d}\boldsymbol{y} \\
&= -\int p_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathcal{H}[\boldsymbol{Y}|\boldsymbol{x}] \, \mathrm{d}\boldsymbol{x} \\
&\quad - \int p_{\boldsymbol{X}}(\boldsymbol{x}) \log p_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\
&\quad + \int p_{\boldsymbol{Y}}(\boldsymbol{y}) \log p_{\boldsymbol{Y}}(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \\
&= \mathrm{E}_{\boldsymbol{x}}[\mathcal{H}[\boldsymbol{Y}|\boldsymbol{x}]] + \mathcal{H}[\boldsymbol{X}] - \mathcal{H}[\boldsymbol{Y}] \ .
\end{aligned}
$$

Rearranging,

$$
\mathrm{E}_{\boldsymbol{y}}[\mathcal{H}[\boldsymbol{X}|\boldsymbol{y}]] - \mathcal{H}[\boldsymbol{X}] = \mathrm{E}_{\boldsymbol{x}}[\mathcal{H}[\boldsymbol{Y}|\boldsymbol{x}]] - \mathcal{H}[\boldsymbol{Y}] \ .
$$

Bibliography

[1] L.A. Anderson and A.R. Robinson. Physical and biological modeling in the Gulf Stream region: Part II. Physical and biological processes. *Deep Sea Research I: Oceanographic Research Papers*, 48(5):1139–1168, 2001.

[2] L.A. Anderson, A.R. Robinson, and C.J. Lozano. Physical and biological modeling in the Gulf Stream region: I. Data assimilation methodology. *Deep-Sea Research I: Oceanographic Research Papers*, 47(10):1787–1827, 2000.

[3] C. Andrieu, N. De Freitas, A. Doucet, and M.I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1):5–43, 2003.

[4] H. Andrieu, M.N. French, W.F. Krajewski, and K.P. Georgakakos. Stochastic-dynamical rainfall simulation based on weather radar volume scan data. *Advances in Water Resources*, 26(5):581–593, 2003.

[5] M. Anghel, K.A. Werley, and A.E. Motter. Stochastic model for power grid dynamics. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, page 113. IEEE, 2007.

[6] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[7] P.J. Attar, E.H. Dowell, J.R. White, and J.P. Thomas. Reduced order nonlinear system identification methodology. *AIAA Journal*, 44:1895–1904, 2006.

[8] F. Auclair, P. Marsaleix, and P. De Mey. Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions. *Dynamics of Atmospheres and Oceans*, 36:309–346, 2003.

[9] A. Azevedo-Filho and R.D. Shachter. Laplace's method approximations for probabilistic inference in belief networks with continuous variables. In *Proceedings of the 10th International Conference on Uncertainty in Artificial Intelligence*, pages 28–36. Morgan Kaufmann Publishers, 1994.

[10] M.R. Bageant. Mechanically-assisted traversal of geological structures. Master's thesis, Massachusetts Institute of Technology, 2013.

[11] Y. Bar-Shalom, X.R. Li, T. Kirubarajan, and J. Wiley. *Estimation with Applications to Tracking and Navigation.* Wiley-Interscience, 2001.

[12] C. Barrett, H.B. Hunt III, M.V. Marathe, S.S. Ravi, D.J. Rosenkrantz, and R.E. Stearns. Modeling and analyzing social network dynamics using stochastic discrete graphical dynamical systems. *Theoretical Computer Science*, 2011.

[13] T. Bayes. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S. *Philosophical Transactions*, 53:370–418, 1763.

[14] M.J. Behrenfeld, R.T. OMalley, D.A. Siegel, C.R. McClain, J.L. Sarmiento, G.C. Feldman, A.J. Milligan, P.G. Falkowski, R.M. Letelier, and E.S. Boss. Climate-driven trends in contemporary ocean productivity. *Nature*, 444(7120):752–755, 2006.

[15] A. Bennett. *Inverse Methods in Physical Oceanography.* Cambridge University Press, 1992.

[16] L.M. Berliner, R.F. Milliff, and C.K. Wikle. Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research*, 108:3104–3121, 2003.

[17] J.A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. International Computer Science Institute, April 1998.

[18] H.M. Blackburn and R.D. Henderson. A study of two-dimensional flow past an oscillating cylinder. *Journal of Fluid Mechanics*, 385(1):255–286, 1999.

[19] J. Bongard and H. Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104:9943–9948, 2007.

[20] T. Braconnier, M. Ferrier, J-C. Jouhard, M. Montagnac, and P. Sagaut. Towards an adaptive POD/SVD surrogate model for aeronautic design. *Computers and Fluids*, 40:195–209, 2011.

[21] R.E. Brandt. Lunar power: An untapped renewable resource. Master's thesis, Massachusetts Institute of Technology, 2013.

[22] W. Bridewell, P. Langley, L. Todorovski, and S. Dzeroski. Inductive process modeling. *Machine Learning*, 71:1–32, 2008.

[23] J. Burkardt, M. Gunzburger, and H-C. Lee. POD and CVT-based reduced-order modeling of Navier-Stokes flows. *Computer Methods in Applied Mechanics and Engineering*, 196:337–355, 2006.

[24] R.P. Chhabra, A.A. Soares, and J.M. Ferreira. Steady non-Newtonian flow past a circular cylinder: A numerical study. *Acta Mechanica*, 172(1):1–16, 2004.

[25] K.C. Chhak and A.M. Moore. The North Atlantic Oscillation as a source of stochastic forcing of the wind-driven ocean circulation. *Dynamics of Atmospheres and Oceans*, 43(3-4):151–170, 2007.

[26] H. Choi. *Adaptive sampling and forecasting with mobile sensor networks.* PhD thesis, Massachusetts Institute of Technology, 2009.

[27] M.J. Choi, V. Chandrasekaran, D.M. Malioutov, J.K. Johnson, and A.S. Willsky. Multiscale stochastic modeling for tractable inference and data assimilation. *Computer Methods in Applied Mechanics and Engineering*, 197:3492–3515, 2008.

[28] M.J. Choi, V. Chandrasekaran, and A.S. Willsky. Gaussian multiresolution models: Exploiting sparse Markov and covariance structure. *IEEE Transactions on Signal Processing*, 58:1012–1024, 2010.

[29] N. Chopin and C.P. Robert. Properties of nested sampling. *Biometrika*, 97(3):741–755, 2010.

[30] National Research Council. *Statistics and Physical Oceanography.* The National Academies Press, 1993.

[31] D.P. Dee and A.M. Da Silva. The choice of variable for atmospheric moisture analysis. *Monthly Weather Review*, 131:155–171, 2003.

[32] S.C.R. Dennis and G.Z. Chang. Numerical solutions for steady flow past a circular cylinder at Reynolds numbers up to 100. *Journal of Fluid Mechanics*, 42(3):471–489, 1970.

[33] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte-Carlo Methods in Practice.* Springer-Verlag, 2001.

[34] D. Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B (Methodological)*, pages 45–97, 1995.

[35] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification.* Wiley-Interscience, 2001.

[36] E.S. Epstein. Stochastic dynamic prediction. *Tellus*, 21(6):739–759, 1969.

[37] P. Fearnhead. Computational methods for complex stochastic systems: A review of some alternatives to MCMC. *Statistics and Computing*, 18(2):151–171, 2008.

[38] W. Fennel and T. Neumann. *Introduction to the Modelling of Marine Ecosystems.* Elsevier Science, 2004.

[39] C.B. Field, M.J. Behrenfeld, J.T. Randerson, and P. Falkowski. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, 281(5374):237–240, 1998.

[40] I.K. Fodor. A survey of dimension reduction techniques. Lawrence Livermore National Laboratory technical report, 2002.

[41] B. Fornberg. A numerical study of steady viscous flow past a circular cylinder. *Journal of Fluid Mechanics*, 98(4):819–855, 1980.

[42] B. Fornberg. Steady viscous flow past a circular cylinder up to Reynolds number 600. *Journal of Computational Physics*, 61(2):297–320, 1985.

[43] B. Fornberg. Steady viscous flow past a sphere at high Reynolds numbers. *Journal of Fluid Mechanics*, 190:471–489, 1988.

[44] S. Frolov, A.M. Baptista, T.K. Leen, Z. Lu, and R. Merwe. Fast data assimilation using a nonlinear kalman filter and a model surrogate: An application to the Columbia River estuary. *Dynamics of Atmospheres and Oceans*, 48:16–45, 2009.

[45] A. Gelb. *Applied Optimal Estimation.* MIT Press, 1974.

[46] A. Gelman and X.L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, pages 163–185, 1998.

[47] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach.* Springer, 1991.

[48] M. Ghil and P. Malanotte-Rizzoli. Data assimilation in meteorology and oceanography. *Advances in Geophysics*, 33:141–266, 1991.

[49] B. Glaz, L. Liu, and P.P. Friedmann. Reduced-order nonlinear unsteady aerodynamic modeling using a surrogate-based recurrence framework. *AIAA Journal*, 48(10):2418, 2010.

[50] T. Gneiting and A.E. Raftery. Weather forecasting with ensemble methods. *Science*, 310(5746):248–249, 2005.

[51] T.E. Govindan, C. Ibarra-Valdez, and J. Ruiz de Chávez. A dynamical stochastic coupled model for financial markets. *Physica A: Statistical Mechanics and its Applications*, 381:317–328, 2007.

[52] C.M. Graves. Wearable robotics: An elegant alternative to muscles. Master's thesis, Massachusetts Institute of Technology, 2013.

[53] K.C. Hall, J.P. Thomas, and E.H. Dowell. Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows. *AIAA Journal*, 38(10):1853–1862, 2000.

[54] K.D. Heaney, G. Gawarkiewicz, T.F. Duda, and P.F.J. Lermusiaux. Nonlinear optimization of autonomous undersea vehicle sampling strategies for oceanographic data-assimilation. *Journal of Field Robotics*, 24(6):437–448, 2007.

[55] J.A. Hoeting, D. Madigan, A.E. Raftery, and C.T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, pages 382–401, 1999.

[56] E.E. Hofmann and M.A.M. Friedrichs. Predictive modeling for marine ecosystems. *The Sea*, 12:537–565, 2002.

[57] E.E. Hofmann and C.M. Lascara. Overview of interdisciplinary modeling for marine ecosystems. *The Sea*, 10:507–540, 1998.

[58] F. Homann. Der einfluss grosser zähigkeit bei der strömung um den zylinder und um die kugel. *Zeitschrift für Angewandte Mathematik und Mechanik*, 16(3):153–164, 1936.

[59] J. Honerkamp. *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis.* Wiley-VCH, 1993.

[60] M.B. Hooten, D.R. Larsen, and C.K. Wikle. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, 18:487–502, 2003.

[61] M.B. Hooten, C.K. Wikle, R.M. Dorazio, and J.A. Royle. Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*, 63:558–567, 2007.

[62] C. Huang, D.W.C. Ho, J. Lu, and J. Kurths. Partial synchronization in stochastic dynamical networks with switching communication channels. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):023108, 2012.

[63] M.F. Huber, T. Bailey, H. Durrant-Whyte, and U.D. Hanebeck. On entropy approximation for Gaussian mixture random vectors. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 181–188. IEEE, 2008.

[64] D.B.P. Huynh, D.J. Knezevic, J.W. Peterson, and A.T. Patera. High-fidelity real-time simulation on deployed platforms. *Computers & Fluids*, 43(1):74–81, 2011.

[65] A.T. Ihler, S. Kirshner, M. Ghil, A.W. Robertson, and P. Smyth. Graphical models for statistical inference and data assimilation. *Physica D: Nonlinear Phenomena*, 230(1):72–87, 2007.

[66] E.L. Ionides, C. Bretó, and A.A. King. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443, 2006.

[67] C.P. Jackson. A finite-element study of the onset of vortex shedding in flow past variously shaped bodies. *Journal of Fluid Mechanics*, 182(1):23–45, 1987.

[68] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, 1970.

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

[70] R.E. Kalman and R.S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83:95–108, 1961.

[71] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 2003.

[72] R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, pages 773–795, 1995.

[73] M. Kopera. Non-dimensionalization of the Navier-Stokes equations. University of Warwick, Centre for Scientific Computing, January 24, 2008.

[74] B. Kouchmeshky, W. Aquino, J.C. Bongard, and H. Lipson. Co-evolutionary algorithm for structural damage identification using minimal physical testing. *International Journal for Numerical Methods in Engineering*, 69(5):1085–1107, 2007.

[75] J.R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* The MIT Press, 1992.

[76] T.N. Krishnamurti, C.M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran. Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate*, 13(23):4196–4216, 2000.

[77] D.M. Kubaczyk. Optimization of a 960 sf. spatial domain for bub hubitation. Master's thesis, Massachusetts Institute of Technology, 2013.

[78] P.K. Kundu and I.M. Cohen. *Fluid Mechanics.* Elsevier, 2004.

[79] W. Lahoz, B. Khattatov, and R. Menard. *Data Assimilation: Making Sense of Observations.* Springer, 2010.

[80] C.M. Lalli and T.R. Parsons. *Biological Oceanography: An Introduction.* Butterworth-Heinemann, 1997.

[81] N. Lartillot and H. Philippe. Computing Bayes factors using thermodynamic integration. *Systematic Biology*, 55(2):195–207, 2006.

[82] F.X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations. *Tellus*, 38A:97–110, 1986.

[83] P.F.J. Lermusiaux. Data assimilation via error subspace statistical estimation. Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation. *Monthly Weather Review*, 127(7):1408–1432, 1999.

[84] P.F.J. Lermusiaux. On the mapping of multivariate geophysical fields: Sensitivities to size, scales, and dynamics. *Journal of Atmospheric and Oceanic Technology*, 19(10):1602–1637, 2002.

[85] P.F.J. Lermusiaux. Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D: Nonlinear Phenomena*, 230(1-2):172–196, 2007.

[86] P.F.J. Lermusiaux, D.G.M. Anderson, and C.J. Lozano. On the mapping of multivariate geophysical fields: Error and variability subspace estimates. *The Quarterly Journal of the Royal Meteorological Society*, 126(565):1387–1429, 2000.

[87] P.F.J. Lermusiaux, C.S. Chiu, G.G. Gawarkiewicz, P. Abbot, A.R. Robinson, R.N. Miller, P.J. Haley, W.G. Leslie, S.J. Majumdar, A. Pang, and F. Lekien. Quantifying uncertainties in ocean predictions. *Oceanography*, 19(1):92–105, 2006.

[88] P.F.J. Lermusiaux, C.S. Chiu, and A.R. Robinson. Modeling uncertainties in the prediction of the acoustic wavefield in a shelfbreak environment. In E.C. Shang, Q. Li, and T.F. Gao, editors, *Proceedings of the 5th International Conference on Theoretical and Computational Acoustics, May 21-25, 2001, in Theoretical and Computational Acoustics*, pages 191–200. International Conference on Theoretical and Computational Acoustics, 2002.

[89] P.F.J. Lermusiaux, P.J. Haley, W.G. Leslie, A. Agarwal, O.G. Logutov, and L.J. Burton. Multiscale physical and biological dynamics in the Philippine Archipelago. *Oceanography*, 24(1):70–89, 2011.

[90] P.F.J. Lermusiaux and AR Robinson. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Monthly Weather Review*, 127(7):1385–1407, 1999.

[91] P.F.J. Lermusiaux, M.P. Ueckermann, and T. Lolla. Linearization for the dynamically orthogonal evolution equations. MIT Multidisciplinary Simulation, Estimation, and Assimilation Systems Group Report, 2013.

[92] S.M. Lewis and A.E. Raftery. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, pages 648–655, 1997.

[93] H. Liao. *Uncertainty decoding for noise robust speech recognition.* PhD thesis, University of Cambridge, 2007.

[94] J.C. Lin, Y. Yang, and D. Rockwell. Flow past two cylinders in tandem: Instantaneous and averaged flow structure. *Journal of Fluids and Structures*, 16(8):1059–1071, 2002.

[95] J.L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations.* Springer Verlag, 1971.

[96] M. Loeve. *Probability Theory II.* Springer, 1978.

[97] K. Logemann, J.O. Backhaus, and I.H. Harms. SNAC: A statistical emulator of the north-east Atlantic circulation. *Ocean Modelling*, 7(1):97–110, 2004.

[98] O.G. Logutov and A.R. Robinson. Multi-model fusion and error parameter estimation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3397–3408, 2005.

[99] D.J. Lucia, P.S. Beran, and W.A. Silva. Reduced-order modeling: New approaches for computational physics. *Progress in Aerospace Sciences*, 40(1):51–117, 2004.

[100] P. Malanotte-Rizzoli. *Modern Approaches to Data Assimilation in Ocean Modeling.* Elsevier Oceanography Series, 1996.

[101] P.S. Maybeck. *Stochastic Models, Estimation and Control*, volume 1. Academic Press, 1979.

[102] X.L. Meng and W.H. Wong. Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6:831–860, 1996.

[103] R.F. Milliff, A. Bonazzi, C.K. Wikle, N. Pinardi, and L.M. Berliner. Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds for a Bayesian hierarchical model. *Quarterly Journal of the Royal Meteorological Society*, 2011.

[104] S. Mittal and B. Kumar. Flow past a rotating cylinder. *Journal of Fluid Mechanics*, 476:303–334, 2003.

[105] NASA's Goddard Space Flight Center. Landsat 7 image of phytoplankton around the Swedish island of Gotland in the Baltic Sea, July 13, 2005.

[106] NASA's Goddard Space Flight Center. Landsat 7 image of clouds off the Chilean coast near the Juan Fernandez Islands, September 15, 1999.

[107] R.M. Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.

[108] M.A. Newton and A.E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B (Methodological)*, pages 3–48, 1994.

[109] A. Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, 1965.

[110] C. Park, W. Bridewell, and P. Langley. Integrated systems for inducing spatio-temporal process models. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence.* Association for the Advancement of Artificial Intelligence, 2010.

[111] S.K. Park and L. Xu. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications.* Springer, 2009.

[112] A.E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5):1155–1174, 2005.

[113] J.D. Ramos. Experiments on sleep deprivation in engineering professionals. Master's thesis, Massachusetts Institute of Technology, 2013.

[114] K. Rimstad and H. Omre. Impact of rock-physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction. *Geophysics*, 75:93–108, 2010.

[115] M. Rixen and E. Ferreira-Coelho. Operational prediction of acoustic properties in the ocean using multi-model statistics. *Ocean Modelling*, 11(3):428–440, 2006.

[116] A.R. Robinson and P.F.J. Lermusiaux. Data assimilation for modeling and predicting coupled physical-biological interactions in the sea. In A.R. Robinson, J.R. McCarthy, and B.J. Rothschild, editors, *The Sea, Volume 12: Biological-Physical Interactions in the Ocean*, pages 475–536. Harvard University Press, 2002.

[117] A.R. Robinson, P.F.J. Lermusiaux, and N.Q. Sloan. Data assimilation. In K.H. Brink and A.R. Robinson, editors, *The Sea, Volume 10: Processes and Methods*, pages 541–594. Harvard University Press, 1998.

[118] R. Rodriguez and H.C. Tuckwell. Statistical properties of stochastic nonlinear dynamical models of single spiking neurons and neural networks. *Physical Review E*, 54(5):5585, 1996.

[119] V.S. Saba, M.A.M. Friedrichs, D. Antoine, R.A. Armstrong, I. Asanuma, M.J. Behrenfeld, A.M. Ciotti, M. Dowell, N. Hoepffner, K.J.W. Hyde, et al. An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe. *Biogeosciences*, 8(2):489–503, 2011.

[120] T.P. Sapsis. *Dynamically orthogonal field equations for stochastic fluid flows and particle dynamics.* PhD thesis, Massachusetts Institute of Technology, 2011.

[121] T.P. Sapsis and P.F.J. Lermusiaux. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D: Nonlinear Phenomena*, 238(23-24):2347–2360, 2009.

[122] T.P. Sapsis and P.F.J. Lermusiaux. Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty. *Physica D: Nonlinear Phenomena*, 241:60–76, 2012.

[123] M. Schmidt and H. Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[124] D. Simon. *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches.* Wiley-Interscience, 2006.

[125] J. Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833–860, 2006.

[126] J.M.L. Sloughter, T. Gneiting, and A.E. Raftery. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, 105(489):25–35, 2010.

[127] F.T. Smith. Laminar flow of an incompressible fluid past a bluff body: The separation, reattachment, eddy properties and drag. *Journal of Fluid Mechanics*, 92(1):171–206, 1979.

[128] K. Sobczyk. Information dynamics: Premises, challenges and results. *Mechanical Systems and Signal Processing*, 15(3):475–498, 2001.

[129] T. Sondergaard and P.F.J. Lermusiaux. Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme. *Monthly Weather Review*, submitted, 2012.

[130] T. Sondergaard and P.F.J. Lermusiaux. Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part II: Applications. *Monthly Weather Review*, submitted, 2012.

[131] Y. Song, C.K. Wikle, C.J. Anderson, and S.A. Lack. Bayesian estimation of stochastic parametrizations in a numerical weather forecasting model. *Monthly Weather Review*, 135:4045–4059, 2007.

[132] R. Srinivasan. *Importance Sampling: Applications in Communications and Detection.* Springer, 2002.

[133] P. Stoica and Y. Selén. Model-order selection: A review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004.

[134] K. Strounine, S. Kravtsov, D. Kondrashov, and M. Ghil. Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance. *Physica D: Nonlinear Phenomena*, 239(3):145–166, 2010.

[135] B.M. Sumer and J. Fredsøe. *Hydrodynamics Around Cylindrical Structures.* World Scientific Publishing Company, 2006.

[136] R.C. Tian, P.F.J. Lermusiaux, J.J. McCarthy, and A.R. Robinson. A generalized prognostic model of marine biogeochemical-ecosystem dynamics: Structure, parameterization and adaptive modeling. Harvard Reports in Physical/Interdisciplinary Ocean Science #67, 2004.

[137] D.J. Tritton. Experiments on the flow past a circular cylinder at low Reynolds numbers. *Journal of Fluid Mechanics*, 6(4):547–567, 1959.

[138] R. Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, 2008.

[139] M.P. Ueckermann and P.F.J. Lermusiaux. High-order schemes for 2D unsteady biogeochemical ocean models. *Ocean Dynamics*, 60(6):1415–1445, 2010.

[140] M.P. Ueckermann and P.F.J. Lermusiaux. Stochastic boundary conditions for the dynamically orthogonal Boussinesq equations, 2012.

[141] M.P. Ueckermann, T.P. Sapsis, and P.F.J. Lermusiaux. Numerical schemes for dynamically orthogonal equations of stochastic fluid and ocean flows. *Journal of Computational Physics*, submitted, 2011.

[142] R. van der Merwe, T.K. Leen, Z. Lu, S. Frolov, and A.M. Baptista. Fast neural network surrogates for very high dimensional physics-based models in computational oceanography. *Neural Networks*, 20(4):462–478, 2007.

[143] M. Van Dyke. *An Album of Fluid Motion.* Parabolic Press, 1982.

[144] A.H. Watson and I.C. Parmee. Identification of fluid systems using genetic programming. In *Proceedings of the Second Online Workshop on Evolutionary Computation (WEC2)*, pages 45–48, 1996.

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

[146] C.K. Wikle, L.M. Berliner, and R.F. Milliff. Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. *Monthly Weather Review*, 131:1051–1062, 2003.

[147] C.K. Wikle and M.B. Hooten. A general science-based framework for dynamical spatio-temporal models. *Test*, 19:417–451, 2010.

[148] A.S. Willsky. Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE*, 90(8):1396–1458, 2002.

[149] C. Wunsch. *The Ocean Circulation Inverse Problem.* Cambridge University Press, 1996.

[150] M.M. Zdravkovich. *Flow Around Circular Cylinders Volume 1: Fundamentals.* Cambridge University Press, 1997.

[151] M.M. Zdravkovich. *Flow Around Circular Cylinders Volume 2: Applications.* Cambridge University Press, 2002.