Uncertainty Propagation and Quantification in a Continuous Time Dynamical System

H.T. Banks and Shuhua Hu
Center for Research in Scientific Computation
Center for Quantitative Sciences in Biomedicine
North Carolina State University
Raleigh, NC 27695-8212 USA

September 6, 2012

Abstract

Uncertainty propagation and quantification has gained considerable research attention during recent years. In this paper we consider uncertainty propagation and quantification in a continuous-time dynamical system governed by ordinary differential equations with uncertain/stochastic components. Specifically, we focus on the time evolution of probability density functions of the resulting stochastic processes, and discuss their applications in different fields with particular focus on population dynamics. In addition, we compare the difference in the stochastic processes resulting from differential equations with different types of random inputs, and discuss the connections among them.

Key Words: stochastic differential equations, random differential equations, probability density functions, Fokker-Planck equation, Liouville equation, structured population model, growth rate distributed structured population model, Markov operators and semigroups, pointwise equivalence.

1 Introduction

Uncertainty is ubiquitous in almost all branches of science and their applications including physics, chemistry, bioscience, engineering, environmental science and social science. Following the seminal work of [58], uncertainty is often classified into two types. One is *epistemic* (or reducible) *uncertainty*, which can possibly be reduced by improved measurements or improvements in the modeling process. The other is *aleatory* (or irreducible) *uncertainty*, which is the result of intrinsic variability/stochasticity of the system.

Uncertainty propagation through a dynamic system has enjoyed considerable research attention during the past decade due to the wide applications of mathematical models in studying the dynamical behavior of systems. In this paper, we consider uncertainty propagation in continuous-time dynamical systems through two types of differential equations. One is stochastic differential equations (SDEs), a classification reserved for differential equations driven by white noise (The value of a white noise at any point is uncorrelated from its value at any other point; that is, its correlation function is some Dirac delta function. Hence, white noise is also referred to as an uncorrelated random field [1]). The other is random differential equations (RDEs), a classification reserved for differential equations driven by other type of random inputs such as colored noise (also referred to as a correlated random field as its value at any point is correlated from its value at any other point) and both colored noise and white noise. Readers can refer to [35, 41, 74] for some interesting discussions on white noise and colored noise as well as their applications in different fields.

Our own previous formulations based on these two types of differential equations are motivated by the growth variability/stochasticity in size-structured population models. Below we will give a brief introduction to these formulations. Unless otherwise indicated, a capital letter is used to denote a random variable, a bold capital letter is for a random vector, and their corresponding small letters are for their realizations throughout our presentation.

1.1 Probabilistic vs. Stochastic Formulations

Two conceptually different approaches (e.g., see [12, 14, 20]) have been investigated to incorporate growth uncertainty into size-structured population models. One entails imposing a probabilistic structure on all the possible growth rates across the entire population while the other involves formulating growth as a stochastic diffusion process. These were referred to as probabilistic formulation and stochastic formulation, respectively. We note that modeling growth uncertainty was the focus of these papers, and hence, both reproduction and mortality rates were not considered.

[1] In this paper, a random field is a family of random variables indexed by time t, or indexed by position $\mathbf{x}$, or indexed by both time t and position $\mathbf{x}$.

1.1.1 The Probabilistic Formulation

The probabilistic formulation is motivated by the fact that genetic differences or non-lethal infections of some chronic disease can have an effect on individual growth rates. For example, for many marine species such as mosquitofish, females grow faster than males, which means that individuals even with the same size may have different growth rates. In addition, it was reported in [30] that non-lethal infection of *Penaeus vannamei* postlarvae by IHHNV (infectious hypodermal and hematopoietic Necrosis virus) may reduce growth rates and increase size variability.

Motivated by these empirical observations, one may partition the entire population into (possibly a continuum of) subpopulations where individuals in each subpopulation have the same size-dependent growth rate. That is, the growth process for individuals in a subpopulation with growth rate g is described by

$$\dot{x}(t;g) = g(t, x(t;g)), \quad g \in \mathcal{G}, \tag{1.1}$$

where $\mathcal{G}$ is a collection of admissible growth rates. Then one assigns a probability distribution $\mathcal{P}$ to this partition of possible growth rates $\mathcal{G}$ in the population. Hence, we can see that for this formulation the growth uncertainty is introduced into the entire population by the variability of growth rates among subpopulations, and each individual grows according to a deterministic growth model, but different individuals (even of the same size) may have different size-dependent growth rates.

With this assumption of a family of admissible growth rates and an associated probability distribution, one thus obtains a generalization of classical size-structured population models called the *growth rate distributed size-structured* (GRDSS) model in this paper, which was first formulated and studied in [7, 16]. This model was studied in [20], where it is called the class rate distribution Sinko-Streifer (CRDSS) model, and also in [7, 10, 11, 12, 13, 16, 17] as the growth rate distribution (GRD) model. Let $v(t, x; g)$ be the population density of individuals with size x at time t having growth rate g (in $\mathcal{G}$). Then the equation for $v(t, x; g)$ with proper boundary and initial conditions is given by

$$\begin{aligned} &\frac{\partial}{\partial t}v(t, x; g) + \frac{\partial}{\partial x}(g(t, x)v(t, x; g)) = 0, \\ &g(t, \underline{x})v(t, \underline{x}; g) = 0, \\ &v(0, x; g) = v_0(x; g), \end{aligned} \tag{1.2}$$

where $\underline{x}$ is the minimum size. A more general boundary condition (with birth involved) is given by

$$g(t, \underline{x})v(t, \underline{x}; g) = \int_{\underline{x}}^{\bar{x}} \beta(t, x)v(t, x; g)dx,$$

where $\bar{x}$ denotes the maximum size individuals may achieve in their life span, and $\beta(t, x)$ is the birth rate of individuals with size x at time t. By "summing" (with respect to the

probability) the corresponding solutions over all $g \in \mathcal{G}$, we obtain the population density for the total population with size x at time t,

$$u(t,x) = \int_{g\in\mathcal{G}} v(t,x;g)d\mathcal{P}(g), \tag{1.3}$$

where $\mathcal{P}$ is the probability measure on $\mathcal{G}$ introduced above. Hence, the GRDSS model involves solving (1.2) and (1.3). In addition, we see that this formulation involves a *stationary probabilistic structure* on a *family of deterministic dynamical systems.*

In practice, the probabilistic structure $\mathcal{P}$ on $\mathcal{G}$ is the fundamental "parameter" to be determined from aggregate data for the population (e.g., observations $\hat{u}$ of the population density $u(t,x)$ for the total population). A delta function based method and a linear spline based method are two approaches that have been used in the literature (e.g., [10, 11, 16, 17] and the references therein) to estimate the growth rate distribution, with the estimation accuracy between these two methods being compared in [10]. It should be noted that a theoretic foundation for these computational methods involves the Prohorov metric [63]. Interested readers can refer to the earlier paper [5] or a later summary [4] for details.

Remark 1.1. We observe that (1.1) is a realization of the random differential equation

$$\dot{x}(t;G) = G(t, x(t;G)), \tag{1.4}$$

where G is a random function with g being its realization. Hence, the solution to (1.1) with all possible g (that is, our probabilistic formulation on the growth rate functions) is the same as the solution to the random differential equation (1.4). In some sense, implementation of this probabilistic formulation (on the growth rate) is equivalent to numerically solving a random differential equation through Monte Carlo methods (see Section 3 below for details).

1.1.2 The Stochastic Formulation

Our stochastic formulation is motivated by the assumption that environmental or emotional fluctuations can be the primary influencing factor on individual growth. For example, it is known that the growth rates of shrimp are affected by several environmental factors such as temperature, dissolved oxygen level and salinity. The stochastic formulation is then constructed under the assumption that movement from one size class to another can be described by a stochastic diffusion process.

Let $X(t)$ represent the size of an individual at time t. We assume that $X(t)$ is a Markov diffusion process, which is described by the Itô stochastic differential equation

$$dX(t) = g(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{1.5}$$

where $W(t)$ is the standard Wiener process [36, 59] (also referred to as *Brownian motion*). Here $g(t,x)$ is the average or mean growth rate of individuals with size x at time t, and is given by

$$\lim_{\Delta t\to 0+} \frac{1}{\Delta t}\mathbb{E}\{\Delta X(t)|X(t) = x\} = g(t,x), \tag{1.6}$$

where $\Delta X(t) = X(t+\Delta t) - X(t)$, and $\mathbb{E}\{\cdot \mid \cdot\}$ denotes the conditional expectation. The function $\sigma(t,x)$ represents variability in growth rate of individuals, and is given by

$$\lim_{\Delta t \to 0+} \frac{1}{\Delta t}\mathbb{E}\{[\Delta X(t)]^2 | X(t) = x\} = \sigma^2(t,x). \tag{1.7}$$

Thus, for this formulation the growth uncertainty is introduced into the population by the stochastic growth of each individual. In addition, individuals with the same size at the same time have the same uncertainty in growth, and individuals also have the possibility to reduce their size during their growth period.

This assumption on the growth process leads to the *Fokker-Planck* (FP) or *Kolmogorov's forward equation* for the population density u (see Section 2.1 for the derivation). The equation with appropriate boundary and initial conditions is given by

$$\begin{aligned}
&\frac{\partial}{\partial t}u(t,x) + \frac{\partial}{\partial x}(g(t,x)u(t,x)) = \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(t,x)u(t,x)),\\
&g(t,\underline{x})u(t,\underline{x}) - \frac{1}{2}\frac{\partial}{\partial x}(\sigma^2(t,x)u(t,x))|_{x=\underline{x}} = 0,\\
&g(t,\bar{x})u(t,\bar{x}) - \frac{1}{2}\frac{\partial}{\partial x}(\sigma^2(t,x)u(t,x))|_{x=\bar{x}} = 0,\\
&u(0,x) = u_0(x),
\end{aligned} \tag{1.8}$$

where $\underline{x}$ and $\bar{x}$ denote the minimum and maximum size individuals may achieve in their life span, and u_0 is the initial population density. A more general left boundary condition (with birth involved) is given by

$$g(t,\underline{x})u(t,\underline{x}) - \frac{1}{2}\frac{\partial}{\partial x}(\sigma^2(t,x)u(t,x))|_{x=\underline{x}} = \int_{\underline{x}}^{\bar{x}} \beta(t,x)u(t,x)dx, \tag{1.9}$$

where $\beta(t,x)$ is the birth rate of individuals with size x at time t.

One of our goals in this paper is to give extended and rigorous foundations for some of these ideas based on the known results from the theories of random differential equations and stochastic differential equations. We remark that the theoretical and computational analysis for SDEs and RDEs are presented in a wide range of journals spread throughout the physics, engineering, statistics, mathematics, etc. literature. Hence another goal of this paper is to give a short and introductory review on a number of theoretical results on SDEs and RDEs presented in different fields with the hope of making researchers in one field aware of the work in other fields and to explain the connections between them. It should be noted that in practice one may often find it is sufficient to obtain certain statistical properties such as the mean value function and covariance function of the stochastic process resulting from the underlying SDEs and RDEs through appropriate numerical methods or perhaps even directly obtain the probability density function of the resulting process. Here we focus on the equations describing the time evolution of the probability density functions of the associated stochastic processes. As we shall see below these equations have their own wide applications.

The remainder of this presentation is organized as follows. In Section 2, stochastic differential equations driven by Gaussian white noise are considered in both the Itô and Stratonovich senses. Random differential equations are then discussed in Section 3 with three particular cases considered: the first with only initial conditions being random, the second with both random initial conditions and random model parameters, and finally the last is driven by correlated stochastic processes. We then discuss the relationship between the SDEs and RDEs in Section 4. Specifically, the probability density functions of the solutions to SDEs and some RDEs can be associated with Markov operators. In addition, there is a class of SDEs with solutions having the same probability density functions at each time t as those for the associated corresponding RDEs. We conclude the paper in Section 5 with some summary remarks and suggestions for future research efforts.

2 Stochastic Differential Equations

White noise, analogous to white light, has a flat power spectrum (the Fourier transform of its correlation function). It is widely used in practice to account for the randomness in the inputs of systems governed by differential equations, especially in the case where one has little knowledge about the precise nature of the noise.

The (generalized) derivative of a Wiener process is white noise, and it is often referred to as *Gaussian white noise*. Note that Gaussian white noise provides a good approximation to many real world problems, and it is with this white noise that stochastic differential equations were originally studied and then subsequently investigated extensively in the literature. Hence, in this section our discussion is focused on the stochastic differential equations driven by Gaussian white noise in both the Itô and Stratonovich senses. Interested readers can refer to [39, 62, 64] and the references therein for information on stochastic differential equations driven by Poisson white noise (the derivative of a compound Poisson process, also known as white shot noise) as well as their applications in different fields.

2.1 Itô Stochastic Differential Equations

In this section we consider the Itô stochastic differential equations

$$d\mathbf{X}(t) = \mathbf{g}(t, \mathbf{X}(t))dt + \boldsymbol{\sigma}(t, \mathbf{X}(t))d\mathbf{W}(t), \quad \mathbf{X}(0) = \mathbf{X}_0, \tag{2.1}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$, $\mathbf{g} = (g_1, g_2, \ldots, g_n)^T$ is a nonrandom n-dimensional function of t and $\mathbf{x}$, $\boldsymbol{\sigma}$ is a nonrandom $n \times l$ matrix function of t and $\mathbf{x}$, $\mathbf{W}(t)$ is an l-dimensional standard Wiener process independent of the random initial vector $\mathbf{X}_0$. We assume that $\mathbf{g}$ and $\boldsymbol{\sigma}$ satisfy conditions guaranteeing the existence and uniqueness of solutions to initial value problem (2.1); that is, $\mathbf{g}$ and $\boldsymbol{\sigma}$ are Lipschitz continuous with respect to $\mathbf{x}$ with the same constant for all t and they satisfy a linear growth condition (e.g., see [45, Theorem 6.22] or [59, Theorem 5.2.1] for details). In addition, we further assume that the functions

involved in this section are sufficiently smooth so that all the differentiations carried out are valid.

Itô stochastic differential equations (introduced by Japanese mathematician K. Itô in the 1940s) are often found useful in financial mathematics as they only take into account the information about the past (for example, in modeling the stock price, the only information one has is about the past events). Itô stochastic differential equations have also been used in describing mechanism of climate variability. For example, they were used in [78] to describe the interactions between the atmosphere temperature and the ocean's surface temperature.

It is well known that the solution to the Itô stochastic differential equation (2.1) is a diffusion process with drift coefficient $\mathbf{g}$ and diffusion coefficient $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ (e.g., see [36, Theorem 3.10] or [70, Section 4.3]). Let $\rho(t, \cdot\,; s, \boldsymbol{\zeta})$ denote the transition probability density function of $\mathbf{X}(t)$ given $\mathbf{X}(s) = \boldsymbol{\zeta}$, where $0 \le s < t$ and $\boldsymbol{\zeta} \in \mathbb{R}^n$. Then we have

$$\begin{aligned}
&\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\zeta}) \rho(t + \Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) d\mathbf{x} = \mathbf{g}(t, \boldsymbol{\zeta}), \\
&\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\zeta})(\mathbf{x} - \boldsymbol{\zeta})^T \rho(t + \Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) d\mathbf{x} = \boldsymbol{\Sigma}(t, \boldsymbol{\zeta}), \\
&\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{\mathbb{R}^n} (x_{j_1} - \zeta_{j_1})(x_{j_2} - \zeta_{j_2}) \cdots (x_{j_\kappa} - \zeta_{j_\kappa}) \rho(t + \Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) d\mathbf{x} = 0, \ \kappa \ge 3,
\end{aligned} \tag{2.2}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$, and κ is some positive integer number. By the last equation in (2.2) we see that for a diffusion process any higher than second-order differential moments are zero.

If $\mathbf{g}$ and $\boldsymbol{\sigma}$ are further assumed to satisfy conditions guaranteeing existence of fundamental solutions to *Kolmogorov's backward equation*

$$\frac{\partial}{\partial s} w(s, \boldsymbol{\zeta}) + \sum_{k=1}^{n} g_k(s, \boldsymbol{\zeta}) \frac{\partial w(s, \boldsymbol{\zeta})}{\partial \zeta_k} + \frac{1}{2} \sum_{k,j=1}^{n} \Sigma_{kj}(s, \boldsymbol{\zeta}) \frac{\partial^2 w(s, \boldsymbol{\zeta})}{\partial \zeta_k \zeta_j} = 0, \tag{2.3}$$

with $\Sigma_{kj}(s, \boldsymbol{\zeta})$ being the (k, j)th element of matrix $\boldsymbol{\Sigma}(s, \boldsymbol{\zeta})$, then this fundamental solution is unique and is given by the transition probability density function $\rho(t, \mathbf{x}; s, \boldsymbol{\zeta})$ (e.g., see [46, Section 5.7]). This implies that for any fixed $(t, \mathbf{x})$, the transition probability density function $\rho(t, \mathbf{x}; s, \boldsymbol{\zeta})$ satisfies Kolmogorov's backward equation with terminal condition $\rho(t, \mathbf{x}; t, \boldsymbol{\zeta}) = \delta(\boldsymbol{\zeta} - \mathbf{x})$, where $0 \le s \le t$, and δ denotes the Dirac delta function. In addition, for any bounded continuous function $f : \mathbb{R}^n \to \mathbb{R}$, the function w given by

$$w(s, \boldsymbol{\zeta}) = \mathbb{E}\{f(\mathbf{X}(t)) \mid \mathbf{X}(s) = \boldsymbol{\zeta}\} = \int_{\mathbb{R}^n} f(\mathbf{x}) \rho(t, \mathbf{x}; s, \boldsymbol{\zeta}) d\mathbf{x}$$

satisfies Kolmogorov's backward equation (2.3) with terminal condition $w(t, \boldsymbol{\zeta}) = f(\boldsymbol{\zeta})$. Thus the solution of Kolmogorov's backward equation with a given terminal condition can be represented by the conditional expectation of some proper stochastic process. Conversely, the conditional expectation of a stochastic process can be obtained by solving Kolmogorov's backward equation with an appropriate terminal condition.

We remark that (2.3) is called the *backward equation* because the differentiation is with respect to the backward variable $(s, \boldsymbol{\zeta})$. Kolmogorov's backward equation is found useful in numerous applications including option pricing theory (where it is common to hold the time of expiration constant) and genetics (where one is often interested in the fact that the system reaches a certain final state, e.g., see [42, Section 4.4]). It should be noted that there is a corresponding Kolmogorov's backward equation for the initial value problem through a change of time variable (as we shall see below; also see [57, Chapter 4] for further details).

Note that if the initial distribution of $\mathbf{X}_0$ is known, then the diffusion process $\{\mathbf{X}(t) : t \geq 0\}$ is uniquely defined by its transition probability density function $\rho(t, \mathbf{x}; s, \boldsymbol{\zeta})$ (actually this is true for any Markov process). Hence, the second order differential operator associated with Kolmogorov's backward equation (for convenience we write it in terms of variable $(t, \mathbf{x})$), that is,

$$\mathscr{L}_t \phi(\mathbf{x}) = \sum_{k=1}^{n} g_k(t, \mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_k} + \frac{1}{2} \sum_{k,j=1}^{n} \Sigma_{kj}(t, \mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_k x_j}, \tag{2.4}$$

is often called the *generator* of the diffusion process $\{\mathbf{X}(t) : t \geq 0\}$. This is also referred to as the *Kolmogorov backward operator*.

Remark 2.1. Before we move on to discussions of evolution of probability density functions as solutions to the Itô stochastic differential equation (2.1), we would like to comment on one important stochastic process, the *Ornstein-Uhlenbeck process*, which satisfies the Itô stochastic differential equation

$$dX(t) = -aX(t)dt + \sigma_0 dW(t), \quad X(0) = X_0. \tag{2.5}$$

Here a and σ_0 are both constants, and a is assumed to be positive. In addition, X_0 is assumed to be normally distributed with mean zero and variance $\dfrac{\sigma_0^2}{2a}$, i.e., $\mathcal{N}\left(0, \dfrac{\sigma_0^2}{2a}\right)$. Then $\{X(t) : t \geq 0\}$ is a Gaussian process with zero mean and exponential correlation function $\mathrm{Cor}\{X(t), X(s)\}$ given by (e.g., see [46, Section 5.6])

$$\mathrm{Cor}\{X(t), X(s)\} = \frac{\sigma_0^2}{2a} \exp\left(-a|t - s|\right), \tag{2.6}$$

where $1/a$ and $\sigma_0^2/(2a^2)$ are respectively called the correlation time (i.e., the bandwidth) and intensity (i.e., the spectral height) of the Ornstein-Uhlenbeck process. It should be noted that the Ornstein-Uhlenbeck process is the only process that is simultaneously Gaussian, Markov and stationary [2] (e.g., see [45, p. 172] or [64, Section 2.5]–this follows from a result due to Doob). In addition, the Ornstein-Uhlenbeck process is widely used as the driving colored noise because it is often the case that environmental fluctuations are the cumulative effect of weakly coupled environmental factors and the central limit theorem then implies that this fluctuation is Gaussian distributed (e.g., [41, 42, 64]). Hence, the Ornstein-Uhlenbeck process is also referred to as *Markov Gaussian colored noise*.

[2] Stationary means that finite dimensional distributions do not change with a shift of time. For a Gaussian process with zero mean, stationary is equivalent to the correlation function being a function of $|t - s|$ only.

2.1.1 Evolution of Probability Density Function of $\mathbf{X}(t)$

Theorem 2.2. *Assume that $\mathbf{X}_0$ has probability density function p_0, and $\mathbf{X}(t)$ satisfies (2.1). Then the probability density function of $\mathbf{X}(t)$ satisfies*

$$\frac{\partial}{\partial t}p(t,\mathbf{x}) + \sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(t,\mathbf{x})p(t,\mathbf{x})) = \frac{1}{2}\sum_{k,j=1}^{n}\frac{\partial^2}{\partial x_k \partial x_j}\left[\Sigma_{kj}(t,\mathbf{x})p(t,\mathbf{x})\right] \tag{2.7}$$

with initial condition $p(0,\mathbf{x}) = p_0(\mathbf{x})$.

By this theorem, we see that probability density function $p(t,\mathbf{x})$ of $\mathbf{X}(t)$ satisfies the *Fokker-Planck equation* (first used by Fokker and Planck to describe the Brownian motion of particles) or *Kolmogorov's forward equation* (rigorously developed by Kolmogorov in 1931). The Fokker-Planck equation is, of course, important in the fields of chemistry and physics. For example, the statistics of laser light and Brownian motion in potentials (important in solid-state physics, chemical physics and electric circuit theory) can both be very well treated with the Fokker-Planck equation (e.g., see [65]). In addition, the Fokker-Planck equation includes the well-known Klein-Kramers equation (or Kramers equation, used to describe particle movement in position and velocity space) and Smoluchowski's equation (describing the particle position distribution) as special cases.

There are several methods that can be used to derive the Fokker-Planck equation (2.7). One method is based on the principle of preservation of probability (see [31]) given by

$$\frac{d}{dt}\int_{\mathbb{S}_t} p(t,\mathbf{x})d\mathbf{x} = 0, \tag{2.8}$$

where $\mathbb{S}_t$ denotes the region of the state space at time t. We remark that the rate of the change of the integral of $p(t,\mathbf{x})$ (i.e., the left side of (2.8)) has two components: one corresponding to the rate of change of $p(t,\mathbf{x})$ in a given region $\mathbb{S}_t$, and the other corresponding to the convective and diffusive transfer through the surface of $\mathbb{S}_t$. Thus, the Fokker-Planck equation (2.7) can also be written in the form of a local conservation equation

$$\frac{\partial}{\partial t}p(t,\mathbf{x}) + \sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left[g_k(t,\mathbf{x})p(t,\mathbf{x}) - \frac{1}{2}\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left(\Sigma_{kj}(t,\mathbf{x})p(t,\mathbf{x})\right)\right] = 0.$$

Another method involves using conditions (2.2) and employing expansion arguments as those in Moyal [56, Section 8.1] or in [60, Section 5.3.2] (for the scalar case of (2.7)) to derive the Fokker-Planck equation (2.7). Since the arguments involved could be used to obtain more general results (which have far more widespread applications), we will present the derivation here.

Proof. Let Δt be a given positive number. Then we have

$$p(t+\Delta t,\mathbf{x}) = \int_{\mathbb{R}^n} p(t,\boldsymbol{\zeta})\rho(t+\Delta t,\mathbf{x};t,\boldsymbol{\zeta})d\boldsymbol{\zeta}. \tag{2.9}$$

The characteristic function $\Pi(t+\Delta t, \boldsymbol{\theta}; t, \boldsymbol{\zeta})$ of $\rho(t+\Delta t, \mathbf{x}; t, \boldsymbol{\zeta})$ is defined by

$$\Pi(t+\Delta t, \boldsymbol{\theta}; t, \boldsymbol{\zeta}) = \int_{\mathbb{R}^n} \rho(t+\Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) \exp(i\boldsymbol{\theta}^T(\mathbf{x}-\boldsymbol{\zeta}))d\mathbf{x} \tag{2.10}$$

$$\rho(t+\Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \Pi(t+\Delta t, \boldsymbol{\theta}; t, \boldsymbol{\zeta}) \exp(-i\boldsymbol{\theta}^T(\mathbf{x}-\boldsymbol{\zeta}))d\boldsymbol{\theta}, \tag{2.11}$$

where $\boldsymbol{\theta} \in \mathbb{R}^n$ and i is the imaginary unit. We see that the characteristic function $\Pi(t+\Delta t, \boldsymbol{\theta}; t, \boldsymbol{\zeta})$ is the Fourier transform of the transition probability density function $\rho(t+\Delta t, \mathbf{x}; t, \boldsymbol{\zeta})$. Expanding $\exp(i\boldsymbol{\theta}^T(\mathbf{x}-\boldsymbol{\zeta}))$ in terms of its argument we have

$$\exp(i\boldsymbol{\theta}^T(\mathbf{x}-\boldsymbol{\zeta})) = 1 + \sum_{k=1}^{\infty} \frac{i^k}{k!} \left[\sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \theta_{j_1} \cdots \theta_{j_k} (x_{j_1} - \zeta_{j_1}) \cdots (x_{j_k} - \zeta_{j_k}) \right]. \tag{2.12}$$

Let $e_{j_1,\cdots,j_k}(t, \boldsymbol{\zeta}, \Delta t) = \int_{\mathbb{R}^n} (x_{j_1} - \zeta_{j_1}) \cdots (x_{j_k} - \zeta_{j_k}) \rho(t+\Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) d\mathbf{x}$, where $j_1, j_2, \ldots, j_k$ are positive integers. Then substitution of (2.12) into (2.10) provides

$$\Pi(t+\Delta t, \boldsymbol{\theta}; t, \boldsymbol{\zeta}) = \quad 1 + \sum_{k=1}^{\infty} \frac{i^k}{k!} \left[\sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \theta_{j_1} \cdots \theta_{j_k} e_{j_1,\cdots,j_k}(t, \boldsymbol{\zeta}, \Delta t) \right]. \tag{2.13}$$

Note that

$$\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} i^k \theta_{j_1} \cdots \theta_{j_k} \exp(-i\boldsymbol{\theta}^T(\mathbf{x}-\boldsymbol{\zeta}))d\boldsymbol{\theta} = (-1)^k \frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}} \delta(\mathbf{x}-\boldsymbol{\zeta}),$$

where $\delta(\mathbf{x}-\boldsymbol{\zeta})$ is the Dirac delta function. Hence, by substituting (2.13) into equation (2.11), we obtain that

$$\rho(t+\Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) = \delta(\mathbf{x}-\boldsymbol{\zeta}) + \sum_{k=1}^{\infty} (-1)^k \frac{1}{k!} \left[\sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \frac{\partial^k \delta(\mathbf{x}-\boldsymbol{\zeta})}{\partial x_{j_1} \cdots \partial x_{j_k}} e_{j_1,\cdots,j_k}(t, \boldsymbol{\zeta}, \Delta t) \right].$$

Substituting the above equation into (2.9), we find

$$p(t+\Delta t, \mathbf{x}) = p(t, \mathbf{x}) + \sum_{k=1}^{\infty} (-1)^k \frac{1}{k!} \left[\sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}} \left(p(t, \mathbf{x}) e_{j_1,\cdots,j_k}(t, \mathbf{x}, \Delta t) \right) \right].$$

Dividing both sides by Δt and letting $\Delta t \to 0$, by (2.2) we obtain (2.7). □

Remark 2.3. We remark that if $\mathbf{g}$, $\boldsymbol{\sigma}$ and the transition probability density function are further assumed to be sufficiently smooth, then the transition probability density function $\rho(t, \mathbf{x}; s, \boldsymbol{\zeta})$ is also a fundamental solution to Fokker-Planck equation (e.g., see [46, Section 5.7]). Note that the probability density function $p(t, \mathbf{x})$ of $\mathbf{X}(t)$ can be written as

$$p(t, \mathbf{x}) = \int_{\mathbb{R}^n} p_0(\boldsymbol{\zeta}) \rho(t, \mathbf{x}; 0, \boldsymbol{\zeta}) d\boldsymbol{\zeta}.$$

Hence, p satisfies the Fokker-Planck equation with initial condition $p(0,\mathbf{x}) = p_0(\mathbf{x})$. This is the method used in [37, Section 5.2] to derive the Fokker-Planck equation (2.7). In addition, we observe that the differential operator associated with the Fokker-Planck equation

$$\mathscr{L}_t^*\psi(\mathbf{x}) = -\sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(t,\mathbf{x})\psi(\mathbf{x})) + \frac{1}{2}\sum_{k,j=1}^{n}\frac{\partial^2}{\partial x_k\partial x_j}\left[\Sigma_{kj}(t,\mathbf{x})\psi(\mathbf{x})\right] \tag{2.14}$$

is the formal adjoint of Kolmogorov's backward operator $\mathscr{L}_t$ defined in (2.4); this operator will be referred to as Kolmogorov's forward operator in this paper. Furthermore, we see that both Kolmogorov's backward operator and Kolmogorov's forward operator are associated with the transition probability density function of the diffusion process $\{\mathbf{X}(t) : t \geq 0\}$.

Remark 2.4. Let $\mathbf{X}(t)$ be an n-dimensional Markov process with transition probability density function $\rho(t,\mathbf{x};s,\boldsymbol{\zeta})$. Assume that for any positive integer k,

$$\lim_{\Delta t\to 0+}\frac{1}{\Delta t}\int_{\mathbb{R}^n}(x_{j_1}-\zeta_{j_1})\cdots(x_{j_k}-\zeta_{j_k})\rho(t+\Delta t,\mathbf{x};t,\boldsymbol{\zeta})d\mathbf{x}$$

exists, where $j_1, j_2, \ldots, j_k$ are positive integers. Then by using the same expansion arguments as those in the proof for Theorem 2.2 we have

$$\frac{\partial}{\partial t}p(t,\mathbf{x}) = \sum_{k=1}^{\infty}(-1)^k\frac{1}{k!}\left[\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\frac{\partial^k}{\partial x_{j_1}\cdots\partial x_{j_k}}\left(\alpha_{j_1,\cdots,j_k}(t,\mathbf{x})p(t,\mathbf{x})\right)\right], \tag{2.15}$$

where $p(t,\mathbf{x})$ denotes the probability density function of $\mathbf{X}(t)$ at time t, and $\alpha_{j_1,\cdots,j_k}$ is given by

$$\alpha_{j_1,\cdots,j_k}(t,\boldsymbol{\zeta}) = \lim_{\Delta t\to 0+}\frac{1}{\Delta t}\int_{\mathbb{R}^n}(x_{j_1}-\zeta_{j_1})\cdots(x_{j_k}-\zeta_{j_k})\rho(t+\Delta t,\mathbf{x};t,\boldsymbol{\zeta})d\mathbf{x}, \quad k = 1,2,3,\ldots.$$

The resulting equation (2.15) is often referred to as the *Kramer-Moyal expansion*, and is an important tool in statistical physics. The corresponding coefficients $\alpha_{j_1,\cdots,j_k}$ are called Kramer-Moyal coefficients or the kth order differential moment. It is worth noting that there is a corresponding Kramer-Moyal backward expansion (e.g., see [65, Section 4.2]). In addition, the transition probability density function $\rho(t,\mathbf{x};s,\boldsymbol{\zeta})$ satisfies the Kramer-Moyal expansion for fixed $(s,\boldsymbol{\zeta})$, and it also satisfies the Kramer-Moyal backward expansion for fixed $(t,\mathbf{x})$ (similar to the transition probability density function for the diffusion process). We observe that the Fokker-Planck equation is a special case of the Kramer-Moyal expansion with all the Kramer-Moyal coefficients vanishing except the ones with one index (i.e., $\alpha_1, \alpha_2, \ldots, \alpha_n$) and those with two indices (i.e., $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{nn}$). It is tempting to generalize the Fokker-Planck equation with some finite-order differential moments higher than two. However, based on the so-called Pawula's theorem (e.g., see [42, Section 4.5] or [65, Section 4.3]), this generalized Fokker-Planck equation leads to a contradiction to the positivity of the distribution function. In other words, the expansion term in the right-hand side of (2.15) may stop after $k > 1$ (which, as we shall see below, leads to Liouville's equation), or after $k > 2$ (which leads to the Fokker-Planck equation), or may never stop (in order to ensure the positivity of distribution function). We remark that the Kramer-Moyal expansion truncated at $k \geq 3$ has been found useful in certain applications (e.g., see [65, Section 4.6] for details) even though the distribution function has negative values.

2.1.2 Application of the Fokker-Plank Equation in Population Dynamics

The Fokker-Planck equation has been effectively used in the literature (e.g., see [26] and the references therein) to model the dispersal behavior of populations such as the female cabbage root fly movement in the presence of Brassica odors, and the movement of flea beetles in cultivated collard patches. In addition, as we saw in Section 1.1.2 the Fokker-Planck equation can be used to describe the population density in a size-structured population where the growth process is a diffusion process satisfying the Itô stochastic differential equation (1.5) in the case of absence of mortality. Next we consider the equation for a density $u(t,x)$ of a population in which mortality is present.

We observe that the density of a population with size x at time $t+\Delta t$ can be calculated by multiplying the transition probability that individuals with size ζ at time t will move to x in a time increment Δt and summing over all the values of ζ. That is,

$$u(t+\Delta t,x)=\int_{\mathbb{R}} u(t,\zeta)(1-\Delta t d(t,\zeta))\rho(t+\Delta t,x;t,\zeta)d\zeta, \tag{2.16}$$

where $\rho(t+\Delta t,x;t,\zeta)$ denotes the transition probability density function given by

$$\rho(t+\Delta t,x;t,\zeta)=\frac{\partial}{\partial x}\text{Prob}\left\{X(t+\Delta t)\le x \mid X(t)=\zeta\right\},$$

and $d(t,x)$ is the mortality rate of individuals with size x at time t. Hence, if we assume that the mortality rate is uniformly bounded, then we can use similar expansion arguments to those in the proof for Theorem 2.2 to obtain

$$\frac{\partial u}{\partial t}(t,x)+\frac{\partial}{\partial x}(g(t,x)u(t,x))+d(t,x)u(t,x)=\frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(t,x)u(t,x)). \tag{2.17}$$

It is also of interest to derive (2.17) from a purely stochastic point of view. Let Z be an independent, exponential distributed random variable with mean 1. Then we define the *killing time* (also referred to as *lifetime*) by (e.g., see [46, Section 5.7])

$$\tau=\inf\left\{t\ge 0 \;\middle|\; \int_0^t d(s,X(s))ds\ge Z\right\},$$

and the killed diffusion process $\tilde{X}(t)$ by

$$\tilde{X}(t)=\begin{cases} X(t), & \text{if } t<\tau \\ \Delta, & \text{if } t\ge\tau, \end{cases} \tag{2.18}$$

where Δ is the "coffin" (or "cemetery") state which is not in $\mathbb{R}$. Then $\{\tilde{X}(t):\ t\ge 0\}$ is also a Markov process, and the associated second order differential operator is given by

$$\mathscr{L}_t\phi(x)=\frac{1}{2}\sigma^2(t,x)\frac{\partial^2\phi(x)}{\partial x^2}+g(t,x)\frac{\partial\phi(x)}{\partial x}-d(t,x)\phi(x).$$

We remark that the above differential operator is related to the famous Feynman-Kac formula (e.g., see [57, Section 4.1] for details), which is a stochastic representation to the solution of some partial differential equation (actually, a generalized Kolmogorov's backward equation with killing involved). Note that the formal adjoint $\mathscr{L}_t^*$ of $\mathscr{L}_t$ is given by

$$\mathscr{L}_t^*\varphi(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(t,x)\varphi(x)) - \frac{\partial}{\partial x}(g(t,x)\varphi(x)) - d(t,x)\varphi(x).$$

Hence, the population density function $u(t,x)$ satisfies

$$\frac{\partial u}{\partial t}(t,x) = \mathscr{L}_t^* u(t,x),$$

which is just (2.17). Thus, equation (2.17) results from an assumed killed diffusion process.

2.2 Stratonovich Stochastic Differential Equations

We next summarize some results for the *Stratonovich stochastic differential equation*

$$d\mathbf{X}(t) = \tilde{\mathbf{g}}(t,\mathbf{X}(t))dt + \boldsymbol{\sigma}(t,\mathbf{X}(t)) \circ d\mathbf{W}(t), \quad \mathbf{X}(0) = \mathbf{X}_0, \tag{2.19}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$, $\tilde{\mathbf{g}} = (\tilde{g}_1, \tilde{g}_2, \ldots, \tilde{g}_n)^T$ is a nonrandom n-dimensional vector function of t and $\mathbf{x}$, $\boldsymbol{\sigma}$ is a nonrandom $n \times l$ matrix function of t and $\mathbf{x}$, $\mathbf{W}(t)$ is an l-dimensional standard Wiener process independent of the random initial vector $\mathbf{X}_0$, and $\circ$ indicates that (2.19) is interpreted in the *Stratonovich* sense. We assume that $\tilde{\mathbf{g}}$ and $\boldsymbol{\sigma}$ satisfy conditions which guarantee the existence and uniqueness of the solution to (2.19), and that solutions are sufficiently smooth so that all the differentiations carried out in this section are valid.

Stratonovich stochastic differential equations [75] (introduced in 1966 by the Russian physicist Stratonovich) provide a natural model for many engineering problems as their solutions can be obtained as the limit of the solutions of differential equations driven by Gaussian colored noises of decreasing correlation time under appropriate scaling (e.g., see [59, Section 3.3], [76, 82, 83]). Even though the Stratonovich integral obeys the rules of classical calculus, it is less efficient than the Itô integral in calculations as the Stratonovich integral is not a martingale while the Itô integral is. Hence, in practice Stratonovich stochastic differential equations are usually converted to their equivalent Itô stochastic differential equations for subsequent mathematical analysis.

The condition for the Itô stochastic differential equation (2.1) to be equivalent to the Stratonovich stochastic differential equation (2.19) is given by (see [36, Section 3.4] or [75])

$$\tilde{g}_k(t,\mathbf{x}) = g_k(t,\mathbf{x}) - \frac{1}{2}\sum_{r=1}^{l}\sum_{j=1}^{n}\sigma_{jr}(t,\mathbf{x})\frac{\partial \sigma_{kr}(t,\mathbf{x})}{\partial x_j}. \tag{2.20}$$

This is often referred to as the *Wong-Zakai correction* (or the *Wong-Zakai theorem*). Thus the diffusion process obtained by the Stratonovich stochastic differential equation has drift

coefficient with kth component defined by $\tilde{g}_k(t,\mathbf{x}) + \frac{1}{2}\sum_{r=1}^{l}\sum_{j=1}^{n}\sigma_{jr}(t,\mathbf{x})\frac{\partial\sigma_{kr}(t,\mathbf{x})}{\partial x_j}$ instead of $\tilde{g}_k(t,\mathbf{x})$, but it still has diffusion coefficient $\boldsymbol{\Sigma}(t,\mathbf{x}) = \boldsymbol{\sigma}(t,\mathbf{x})\boldsymbol{\sigma}^T(t,\mathbf{x})$. By Theorem 2.2 we obtain the following results on the evolution of the probability density function of the solution $\mathbf{X}(t)$ to the Stratonovich stochastic differential equation (2.19).

Theorem 2.5. *Assume that $\mathbf{X}_0$ has probability density function p_0, and $\mathbf{X}(t)$ satisfies* (2.19). *Then the probability density function of $\mathbf{X}(t)$ satisfies*

$$\begin{aligned}\frac{\partial}{\partial t}p(t,\mathbf{x}) + \sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left[\left(\tilde{g}_k(t,\mathbf{x}) - \frac{1}{2}\sum_{r=1}^{l}\sum_{j=1}^{n}\sigma_{kr}(t,\mathbf{x})\frac{\partial\sigma_{jr}(t,\mathbf{x})}{\partial x_j}\right)p(t,\mathbf{x})\right] \\ = \frac{1}{2}\sum_{k=1}^{n}\frac{\partial}{\partial x_k}\left[\sum_{j=1}^{n}\Sigma_{kj}(t,\mathbf{x})\frac{\partial}{\partial x_j}p(t,\mathbf{x})\right]\end{aligned} \tag{2.21}$$

with initial condition $p(0,\mathbf{x}) = p_0(\mathbf{x})$.

3 Random Differential Equations

Colored noise has been observed in many contexts including astronomy, electronic devices, brain signals, heartbeat rhythms, psychological mental states and human auditory cognition (e.g., see [74]). In general, colored noise will not only affect the dynamical aspects of the system (e.g., Kubo showed in 1962 that a fluctuating magnetic field correlated over a long time scale (colored noise) can significantly modify the motion of spin while the one with very short correlation time (almost white noise) typically does not) but also the form of the stationary probability (e.g., studies of colored noise induced transitions). Interested readers can refer to [40] and the references therein for more information.

Due in part to their wide applicability, random differential equations have enjoyed considerable research attention in the past decade, especially efforts on computational methods. Widely used approaches include Monte Carlo methods, stochastic Galerkin methods and probabilistic collocation methods (also called stochastic collocation methods). Specifically, both Monte Carlo methods and probabilistic collocation methods seek to solve deterministic realizations of the given random differential equations. The difference between these two methods is the manner in which one chooses the "sampling" points. Monte Carlo methods are based on large sampling of the distribution of random input variables while probabilistic collocation methods are based on quadrature rules (or sparse quadrature rules in high-dimensional space). Stochastic Galerkin methods are based on (generalized) polynomial chaos expansions, which express the unknown random fields in terms of convergent series of (global) orthogonal polynomials in terms of random input parameters. Interested readers can refer to [84] and the references therein for details.

It should be noted that in practice one often decomposes and truncates the colored noise by a finite linear combination of uncorrelated random variables (e.g., using a Karhunen–

Loève expansion) for uncertainty quantification. Hence, in this section we consider systems governed by random differential equations in three cases. One is differential equations with random initial conditions (in Section 3.1). The second is differential equations with both random initial conditions and random model parameters characterized by a finite dimensional random vector (in Section 3.2). Finally we consider systems driven by a correlated stochastic process (in Section 3.3). As we shall see below, the first two cases have wide applications.

3.1 Differential Equations with Random Initial Conditions

We first consider the evolution of the probability density function for the solution to the system of ordinary differential equation with random initial conditions

$$\dot{\mathbf{x}} = \mathbf{g}(t, \mathbf{x}), \quad \mathbf{x}(0) = \mathbf{X}_0, \tag{3.1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, $\mathbf{g} = (g_1, g_2, \ldots, g_n)^T$ is an n-dimensional nonrandom vector function of t and $\mathbf{x}$, and $\mathbf{X}_0$ is an n-dimensional random vector.

Equation (3.1) is often referred to as a *crypto-deterministic* formulation (e.g., see [36, 56, 72]), and has been proven useful in a wide number of applications including classical statistical mechanics, statistical thermodynamics, kinetic theory and biosciences (e.g., see [72, Chapter 6] for more detailed discussions). Depending on the applications, the uncertainty in initial conditions could be classified as either epistemic uncertainty (e.g., the uncertainty is due to the measurement error) or aleatory uncertainty (e.g., individuals have different initial size, as will be discussed below).

3.1.1 Evolution of the Probability Density Function of $\mathbf{x}(t; \mathbf{X}_0)$

Theorem 3.1. *([72, Theorem 6.2.2]) Assume that $\mathbf{X}_0$ has probability density function p_0, and (3.1) has a mean square solution $\mathbf{x}(t; \mathbf{X}_0)$. Then the probability density function of the solution $\mathbf{x}(t; \mathbf{X}_0)$ satisfies*

$$\frac{\partial}{\partial t} p(t, \mathbf{x}) + \sum_{k=1}^{n} \frac{\partial}{\partial x_k} (g_k(t, \mathbf{x}) p(t, \mathbf{x})) = 0 \tag{3.2}$$

with initial condition $p(0, \mathbf{x}) = p_0(\mathbf{x})$.

The above result is often referred to as *Liouville's "theorem"*, which is a key result in statistical mechanics and statistical thermodynamics. The resulting equation (3.2) is sometimes called *Liouville's equation*, and as such it includes the equation for describing the evolution of the probability density function for a conservative Hamiltonian system (e.g., see [56, Section 1.3]) as a special case. Here we shall use the terminology "Liouville's theorem" and "Liouville's equation" interchangeably to refer to this *conservation law* even though "Liouville's

equation" (also attributed to Gibbs in 1902) often is used to refer to a different equation related to differential geometry and isothermal coordinates (compare "Liouville's equation" and "Liouville's theorem" in Wikipedia).

There are several methods that can be used to derive Liouville's equation. One is from a probabilistic point of view using the concept of characteristic functions (e.g., see [72, p. 147]) following the derivation of Kozin [51] from 1961, which we next give.

Proof. For any fixed t, the characteristic function $\Pi(t, \boldsymbol{\theta})$ of $\mathbf{X}(t) \equiv \mathbf{x}(t; \mathbf{X}_0)$ is given by

$$\Pi(t, \boldsymbol{\theta}) = \mathbb{E}\left[\exp\left(i\left(\boldsymbol{\theta}^T \mathbf{X}(t)\right)\right)\right], \tag{3.3}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^T \in \mathbb{R}^n$, and i is the imaginary unit. It is worth noting that the characteristic function $\Pi(t, \boldsymbol{\theta})$ is essentially the Fourier transform of the probability density function $p(t, \mathbf{x})$ of $\mathbf{X}(t)$. Differentiating both sides of (3.3) with respect to t we obtain

$$\frac{\partial \Pi}{\partial t}(t, \boldsymbol{\theta}) = \mathbb{E}\left[i\boldsymbol{\theta}^T \dot{\mathbf{X}}(t) \exp\left(i\boldsymbol{\theta}^T \mathbf{X}(t)\right)\right] = i \sum_{k=1}^{n} \theta_k \mathbb{E}\left[g_k(t, \mathbf{X}(t)) \exp\left(i\boldsymbol{\theta}^T \mathbf{X}(t)\right)\right].$$

Taking the inverse Fourier transform of both sides of the above equation we find

$$\frac{\partial}{\partial t} p(t, \mathbf{x}) = -\sum_{k=1}^{n} \frac{\partial}{\partial x_k}(g_k(t, \mathbf{x}) p(t, \mathbf{x})),$$

which yields the desired results. □

Another method is based on the principle of preservation of probability (e.g., see [69, pages 363-364]) which posits

$$\frac{d}{dt} \int_{\mathbb{S}_t} p(t, \mathbf{x}) d\mathbf{x} = 0, \tag{3.4}$$

where $\mathbb{S}_t$ denotes the region of the state space at time t. This method is called the *phase-space method* (or state-space method) in statistical physics. We remark that the rate of the change of the integral of $p(t, \mathbf{x})$ (i.e., the left side of (3.4)) has two components: one corresponding to the rate of change of $p(t, \mathbf{x})$ in a given region $\mathbb{S}_t$, and the other corresponding to the convective transfer through the surface of $\mathbb{S}_t$. In addition, we observe that Liouville's equation (3.2) with $n = 3$ is the *equation of continuity* in classical continuum mechanics, in which $p(t, \mathbf{x})$ denotes the density of the material at location $\mathbf{x}$ at time t, and $\mathbf{g}(t, \mathbf{x})$ denotes the velocity of a particle at location $\mathbf{x}$ at time t. The equation of continuity is derived based on the conservation of mass (e.g., see [21] and the references therein) which is given by

$$\frac{d}{dt} \int_{\Omega_t} p(t, x_1, x_2, x_3) dx_1 dx_2 dx_3 = 0. \tag{3.5}$$

Here Ω_t denotes the spacial domain occupied by a given set of material particles at time t.

In addition, we see that (3.1) is a special case of the Itô stochastic differential equation (2.1) with $\boldsymbol{\sigma} \equiv 0$. Hence, if we assume that $\mathbf{g}$ satisfies the conditions for the existence and uniqueness of solutions of (2.1), then we know that $\{\mathbf{X}(t) : t \geq 0\}$ with $\mathbf{X}(t) = \mathbf{x}(t; \mathbf{X}_0)$ is a diffusion process with drift coefficient $\mathbf{g}$ and diffusion coefficient being the zero matrix. Let $\rho(t + \Delta t, \mathbf{x}; t, \boldsymbol{\zeta})$ denote the transition probability density function. Then we have

$$\begin{aligned} &\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\zeta}) \rho(t + \Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) d\mathbf{x} = \mathbf{g}(t, \boldsymbol{\zeta}), \\ &\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{\mathbb{R}^n} (x_{j_1} - \zeta_{j_1}) \cdots (x_{j_\kappa} - \zeta_{j_\kappa}) \rho(t + \Delta t, \mathbf{x}; t, \boldsymbol{\zeta}) d\mathbf{x} = 0, \ \kappa \geq 2, \end{aligned} \tag{3.6}$$

where κ is a positive integer. By using the above conditions as well as the expansion arguments similar to those in the proof for Theorem 2.2 we can also derive Liouville's equation, which is a special case of the Kramer-Moyal expansion (2.15) with all the Kramer-Moyal coefficients vanishing except the ones with only one index (i.e., $\alpha_1, \alpha_2, \ldots, \alpha_n$).

3.1.2 Application of Liouville's Equation in Population Dynamics

We observe that Liouville's equation (3.2) with $g_1 \equiv 1$ is a special case of a model without mortality ($d \equiv 0$) given by Oster and Takahashi in 1974 [61] [3]

$$\frac{\partial}{\partial t} u(t, \mathbf{x}) + \frac{\partial}{\partial x_1} u(t, \mathbf{x}) + \sum_{j=2}^{n} \frac{\partial}{\partial x_j} (g_j(t, \mathbf{x}) u(t, \mathbf{x})) + d(t, \mathbf{x}) u(t, \mathbf{x}) = 0. \tag{3.7}$$

This model is used to describe the population density $u(t, \mathbf{x})$ in a spatially homogeneous population where individuals are characterized by states $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ with x_1 denoting the chronological age and $x_2, \cdots, x_n$ representing some physiological variables such as mass, volume, chemical composition, and other quantities having an influence on individual's growth and mortality rates, $(g_2(t, \mathbf{x}), g_3(t, \mathbf{x}), \ldots, g_n(t, \mathbf{x}))^T$ is the vector for the growth rates of individuals with states $\mathbf{x}$ at time t, and $d(t, \mathbf{x})$ is the mortality rate of individuals with states $\mathbf{x}$ at time t.

Equations (3.7) includes a number of well-known structured population models as special cases. Specifically, the model with only chronological age involved, i.e.,

$$\frac{\partial}{\partial t} u(t, x) + \frac{\partial}{\partial x} u(t, x) + d(t, x) u(t, x) = 0 \tag{3.8}$$

is the *age-structured population model* given by Mckendrick in 1926 [54] and Von Foerster in 1959 [79]. (Equation (3.8) is also referred to as the *McKendrick equation*, *the Lotka-Mckendrick equation*, *the Lotka-Von Foerster model*, *the von Forester equation*, or

[3] The connection between Liouville's equation and (3.7) was recognized by Oster and Takahashi in the case of zero mortality and constant growth rates g_i, $i = 2, \ldots, n$.

McKendrick-von Forester equation. The interested readers are referred to [47] for an interesting survey about the priority issues). Equation (3.7) with $n = 2$

$$\frac{\partial}{\partial t}u(t,\mathbf{x}) + \frac{\partial}{\partial x_1}u(t,\mathbf{x}) + \frac{\partial}{\partial x_2}(g_2(t,\mathbf{x})u(t,\mathbf{x})) + d(t,\mathbf{x})u(t,\mathbf{x}) = 0, \tag{3.9}$$

is the age-size structured population model developed in 1967 by Sinko and Streifer [71] (with x_2 being some physiological variable), and it is also the model given at the same time by Bell and Anderson [27] for cell populations (with x_2 denoting the volume of the cell). Equation (3.9) without the second term, that is,

$$\frac{\partial}{\partial t}u(t,x) + \frac{\partial}{\partial x}(g(t,x)u(t,x)) + d(t,x)u(t,x) = 0, \tag{3.10}$$

is often referred to as *Sinko-Streifer model* or *classical size-structured population model*. Here x is the structure variable, which may represent weight, length, volume, chronological age (in this case, equation (3.10) becomes the age-structured population model (3.8)), caloric content, maturity, etc, depending on the specific applications.

These structured population models with appropriate boundary and initial conditions along with their corresponding nonlinear versions have been widely studied in the literature for both computational and theoretical analysis including well-posedness studies, control studies, sensitivity analysis and parameter estimation techniques (e.g., [1, 3, 15, 22, 23, 28, 32, 44, 55, 80] and the references therein). In addition, they have been applied to many biological and medical applications (e.g., [6, 24, 67] and the references therein).

The derivation of the age-size structured population model (3.9) in [71] is essentially by conservation of total number of population with population sinks included, i.e.,

$$\frac{d}{dt}\int_{\Omega_t} u(t,\mathbf{x})d\mathbf{x} = -\int_{\Omega_t} d(t,\mathbf{x})u(t,\mathbf{x})d\mathbf{x}, \tag{3.11}$$

where $\mathbf{x} = (x_1, x_2)^T$ and Ω_t denotes all possible states occupied by individuals at time t. It is also interesting to derive this model from a stochastic point of view. For simplicity, we take the size-structured population model (3.10) as an example. Note that the growth of each individual in the size-structured population model (3.10) follows the same deterministic law

$$\dot{x} = g(t,x), \tag{3.12}$$

but individuals may start from different sizes at time $t = 0$. Hence, the size of an individual at time t can be viewed as a realization to the solution $x(t; X_0)$ of the following differential equation with random initial condition

$$\dot{x} = g(t,x), \quad x(0) = X_0. \tag{3.13}$$

Thus, the solution $X(t) = x(t; X_0)$ to (3.13) is a diffusion process. Denote the transition probability density function by

$$\rho(t+\Delta t, x; t, \zeta) = \frac{\partial}{\partial x}\text{Prob}\left\{X(t+\Delta t) \le x \mid X(t) = \zeta\right\},$$

then $\rho(t, x; s, \zeta)$ satisfies the conditions

$$\begin{aligned} &\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x - \zeta)\rho(t + \Delta t, x; t, \zeta)dx = g(t, \zeta), \\ &\lim_{\Delta t \to 0+} \frac{1}{\Delta t} \int_{-\infty}^{\infty} |x - \zeta|^{\kappa}\rho(t + \Delta t, x; t, \zeta)dx = 0, \text{ where } \kappa \geq 2. \end{aligned} \tag{3.14}$$

With conditions (3.14), we can use arguments similar to those in the proof for deriving equation (2.17) to obtain (3.10).

Based on the above discussions, we see that all the linear structured population models presented in this section can be associated with some stochastic processes, which results from the variability in the initial sizes of individuals in the population.

3.2 Differential Equations with Random Model Parameters and Random Initial Conditions

In this section we consider the system of random ordinary differential equations

$$\dot{\mathbf{x}} = \mathbf{g}(t, \mathbf{x}; \mathbf{Z}), \quad \mathbf{x}(0) = \mathbf{X}_0, \tag{3.15}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and $\mathbf{g} = (g_1, g_2, \ldots, g_n)^T$ is an n-dimensional nonrandom vector function of t and $\mathbf{x}$, $\mathbf{Z}$ is an m-dimensional random vector, and $\mathbf{X}_0$ is an n-dimensional random vector.

Equation (3.15) is often found useful in the biosciences to account for the variation between individuals. For example, a system of ordinary differential equations is often used in the literature to model the HIV dynamics for any specific individual, but the model parameters vary across the population as the clinical data reveals a great deal of variability among HIV patients (e.g., see [2, 9, 19, 66]). In addition, (3.15) has wide applications in physics to account for heterogeneity of complex materials. For example, oscillator models such as the Debye model (described by a first-order ordinary differential equation [8]) and the Lorentz model (described by a second-order ordinary differential equation) are often used to describe the polarization in a dielectric material, but there is now incontrovertible experimental arguments for distributions of relaxation times (parameters in the oscillator models) for complex materials (e.g., see [18] and the references therein). We remark that the uncertainty discussed in all the above cases is aleatory. It should be noted that equation (3.15) can also be used to describe cases where uncertainty is epistemic, for example, uncertainty arising from the parameter estimation error due to the noise in the measurements.

3.2.1 Evolution of the Joint Probability Density Function for $(\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z}),\mathbf{Z})^T$

We consider the time evolution of the joint probability density function of $(\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z}),\mathbf{Z})^T$. Let $\tilde{\mathbf{x}}=(\mathbf{x},\mathbf{z})^T$. Then (3.15) can be rewritten as

$$\dot{\tilde{\mathbf{x}}}=\tilde{\mathbf{g}}(t,\tilde{\mathbf{x}}),\quad \tilde{\mathbf{x}}(0)=(\mathbf{X}_0,\mathbf{Z})^T. \tag{3.16}$$

Here $\tilde{\mathbf{g}}(t,\tilde{\mathbf{x}})=\begin{bmatrix}\mathbf{g}(t,\mathbf{x};\mathbf{z})\\ \mathbf{0}_m\end{bmatrix}$, where $\mathbf{0}_m$ is m-dimensional column vector with all the elements being zeros. Now Theorem 3.1 can be applied to (3.16) to obtain the following result on the evolution of the joint probability density function for the solution to (3.16).

Theorem 3.2. *(see [73, Theorem 1]) Assume that* (3.15) *has a mean square solution* $\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z})$*. Then the joint probability density function* $\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})$ *of* $\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z})$ *and* $\mathbf{Z}$ *satisfies*

$$\frac{\partial}{\partial t}\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})+\sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(t,\mathbf{x};\mathbf{z})\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z}))=0 \tag{3.17}$$

with initial condition

$$\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(0,\mathbf{x},\mathbf{z})=\tilde{\varphi}^0_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}),$$

where $\tilde{\varphi}^0_{\mathbf{X},\mathbf{Z}}$ *is the joint probability density function of* $\mathbf{X}_0$ *and* $\mathbf{Z}$*.*

Equation (3.17) is sometimes referred to as the *Dostupov-Pugachev equation* (e.g., see [31]) as it seems to first appear in [34] by Dostupov and Pugachev in 1957. For derivation of (3.17), Dostupov and Pugachev first observed that for a given $\mathbf{z}$ of $\mathbf{Z}$ the system (3.15) is basically the crypto-deterministic system (3.1), and they then derived equation (3.17) through instantaneous transformation of random vectors.

It is of interest to note that we can employ arguments similar to those in the proof for Theorem 3.1 to (3.15) to obtain equation (3.17). This proof was originally given in 1965 by Kozin [51] and is repeated here for completeness.

Proof. For any fixed t, the joint characteristic function $\Pi(t,\boldsymbol{\theta},\boldsymbol{\vartheta})$ of $\mathbf{X}(t)\equiv\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z})$ and $\mathbf{Z}$ is given by

$$\Pi(t,\boldsymbol{\theta},\boldsymbol{\vartheta})=\mathbb{E}\left[\exp\left(i\left(\boldsymbol{\theta}^T\mathbf{X}(t)+\boldsymbol{\vartheta}^T\mathbf{Z}\right)\right)\right], \tag{3.18}$$

where $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_n)^T\in\mathbb{R}^n$ and $\boldsymbol{\vartheta}=(\vartheta_1,\vartheta_2,\ldots,\vartheta_m)^T\in\mathbb{R}^m$. We note again that the joint characteristic function $\Pi(t,\boldsymbol{\theta},\boldsymbol{\vartheta})$ is the Fourier transform of the joint probability density function $\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})$. Differentiating both sides of (3.18) with respect to t we find

$$\begin{aligned}\frac{\partial\Pi}{\partial t}(t,\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \mathbb{E}\left[i\boldsymbol{\theta}^T\dot{\mathbf{X}}(t)\exp\left(i\left(\boldsymbol{\theta}^T\mathbf{X}(t)+\boldsymbol{\vartheta}^T\mathbf{Z}\right)\right)\right]\\ &= i\sum_{k=1}^{n}\theta_k\mathbb{E}\left[g_k(t,\mathbf{X}(t);\mathbf{Z})\exp\left(i\left(\boldsymbol{\theta}^T\mathbf{X}(t)+\boldsymbol{\vartheta}^T\mathbf{Z}\right)\right)\right].\end{aligned}$$

Taking the inverse Fourier transform of both sides of the above equation we obtain

$$\frac{\partial}{\partial t}\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z}) = -\sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(t,\mathbf{x};\mathbf{z})\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})),$$

which yields the desired results. □

We also observe that (3.16) is a special case of a stochastic differential equation with all diffusion coefficients being zero. Let $\tilde{\mathbf{X}}(t) = (\mathbf{X}(t), \mathbf{Z})^T$, and $\rho(t,\cdot;s,\tilde{\boldsymbol{\zeta}})$ denote the transition probability density function of $\tilde{\mathbf{X}}(t)$ given $\tilde{\mathbf{X}}(s) = \tilde{\boldsymbol{\zeta}}$, where $0 \le s < t$ and $\tilde{\boldsymbol{\zeta}} \in \mathbb{R}^{n+m}$. Then we have

$$\begin{aligned}
&\lim_{\Delta t\to 0+}\frac{1}{\Delta t}\int_{\mathbb{R}^{n+m}}(\tilde{\mathbf{x}}-\tilde{\boldsymbol{\zeta}})\rho(t+\Delta t,\tilde{\mathbf{x}};t,\tilde{\boldsymbol{\zeta}})d\tilde{\mathbf{x}} = \tilde{\mathbf{g}}(t,\tilde{\boldsymbol{\zeta}}),\\
&\lim_{\Delta t\to 0+}\frac{1}{\Delta t}\int_{\mathbb{R}^{n+m}}(\tilde{x}_{j_1}-\tilde{\zeta}_{j_1})\cdots(\tilde{x}_{j_\kappa}-\tilde{\zeta}_{j_\kappa})\rho(t+\Delta t,\tilde{\mathbf{x}};t,\tilde{\boldsymbol{\zeta}})d\tilde{\mathbf{x}} = 0,\ \kappa \ge 2,
\end{aligned} \tag{3.19}$$

where κ is a positive integer. Thus by using the above conditions as well as the expansion arguments given above in deriving the Fokker-Planck equation (2.7), we can also obtain Dostupov-Pugachev equation (3.17), which, of course, is a special case of the Kramer-Moyal expansion (2.15) with all the Kramer-Moyal coefficients vanishing except $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Given the joint probability density function $\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})$, we can obtain the probability density function $p(t,\mathbf{x})$ for $\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z})$ given by

$$p(t,\mathbf{x}) = \int_{\mathbb{Z}}\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})d\mathbf{z}, \tag{3.20}$$

where $\mathbb{Z}$ denotes the set of all possible values for $\mathbf{z}$.

3.2.2 Evolution of Conditional Probability Density Function of $\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z})$ Given the Value z of Z

Finally we derive the conditional probability density function $\varphi(t,\mathbf{x};\mathbf{z})$ of $\mathbf{x}(t;\mathbf{X}_0,\mathbf{Z})$ given the value $\mathbf{z}$ of $\mathbf{Z}$ (as we shall see momentarily, this has applications in population dynamics). First we observe that the conditional probability density function $\varphi(t,\mathbf{x};\mathbf{z})$ is

$$\varphi(t,\mathbf{x};\mathbf{z}) = \frac{\tilde{\varphi}_{\mathbf{X},\mathbf{Z}}(t,\mathbf{x},\mathbf{z})}{\tilde{\varphi}_{\mathbf{Z}}(\mathbf{z})}, \tag{3.21}$$

where $\tilde{\varphi}_{\mathbf{Z}}(\mathbf{z})$ denotes the probability density function of $\mathbf{Z}$. Hence by (3.17) and (3.21) we find $\varphi(t,\mathbf{x};\mathbf{z})$ satisfies

$$\frac{\partial}{\partial t}\varphi(t,\mathbf{x};\mathbf{z}) + \sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(t,\mathbf{x};\mathbf{z})\varphi(t,\mathbf{x};\mathbf{z})) = 0, \tag{3.22}$$

with initial condition

$$\varphi(0, \mathbf{x}; \mathbf{z}) = \varphi_0(\mathbf{x}; \mathbf{z}),$$

where $\varphi_0(\mathbf{x}; \mathbf{z})$ is the probability density function of initial condition $\mathbf{X}_0$ given $\mathbf{Z} = \mathbf{z}$.

Observe that for any given value $\mathbf{z}$ of $\mathbf{Z}$ system (3.15) is crypto-deterministic. Hence, by using Liouville's equation (3.2) the probability density function $\varphi(t, \mathbf{x}; \mathbf{z})$ of solution $\mathbf{x}(t; \mathbf{X}_0, \mathbf{z})$ to (3.15) with given value $\mathbf{z}$ of $\mathbf{Z}$ satisfies (3.22). Thus, we can derive the equation for the time evolution of conditional probability density function $\varphi(t, \mathbf{x}; \mathbf{z})$ directly from Liouville's equation.

Given the conditional probability density function $\varphi(t, \mathbf{x}; \mathbf{z})$, the probability density function $p(t, \mathbf{x})$ of $\mathbf{x}(t; \mathbf{X}_0, \mathbf{Z})$ is given by

$$p(t, \mathbf{x}) = \int_{\mathbb{Z}} \varphi(t, \mathbf{x}; \mathbf{z}) \tilde{\varphi}_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}. \tag{3.23}$$

Remark 3.3. Observe that (3.22) is a *realization* of the random partial differential equation (RPDE)

$$\frac{\partial}{\partial t}\varphi(t, \mathbf{x}; \mathbf{Z}) + \frac{\partial}{\partial x}(g(t, \mathbf{x}; \mathbf{Z})\varphi(t, \mathbf{x}; \mathbf{Z})) = 0,$$

with initial condition

$$\varphi(0, \mathbf{x}; \mathbf{Z}) = \varphi_0(\mathbf{x}; \mathbf{Z}),$$

and (3.23) is the expected value of the solution to the above random partial differential equation. Hence, there is a nice correspondence between the probability density function of the solution of random ordinary differential equation (RODE) and the mean of the solution of an RPDE.

3.2.3 Applications in Population Dynamics

As we saw in Section 1.1.1, the scalar case of (3.22) with appropriate boundary and initial conditions along with (3.23) can be used to describe the population density in a size-structured population, where the total population is partitioned into subpopulations with each subpopulation having its own growth rate $g(t, x; \mathbf{z})$. This model has been successfully used to model mosquitofish populations in rice fields, where the data exhibits both bimodality and dispersion in size as time increases (e.g., see [7]) even though the population begins with a unimodal density (no reproduction involved). Moreover, the model was also used to model the early growth of shrimp populations, which exhibit a great deal of variability in size as time evolves even though the shrimp begin with approximately similar size [12, 13].

Here we derive this GRDSS model in the presence of mortality. Note that the size of an individual at time t in a subpopulation having growth rate $g(t, x; \mathbf{z})$ can be viewed as a realization to the solution $X_{\mathbf{z}}(t) = x(t; \mathbf{z}, X_{0\mathbf{z}})$ of the following differential equation with random initial condition

$$\dot{x}(t; \mathbf{z}) = g(t, x; \mathbf{z}), \quad x(0; \mathbf{z}) = X_{0\mathbf{z}}, \tag{3.24}$$

where $X_{0\mathbf{z}}$ denotes the initial size of individuals in this subpopulation having growth rate $g(t,x;\mathbf{z})$. Hence, $\{X_{\mathbf{z}}(t):\ t\geq 0\}$ is a diffusion process. Denote the transition probability density function by

$$\rho_{\mathbf{z}}(t+\Delta t,x;t,\zeta)=\frac{\partial}{\partial x}\text{Prob}\left\{X_{\mathbf{z}}(t+\Delta t)\leq x\mid X_{\mathbf{z}}(t)=\zeta\right\}.$$

Then $\rho_{\mathbf{z}}(t,x;s,\zeta)$ satisfies the following conditions

$$\begin{aligned}&\lim_{\Delta t\to 0+}\frac{1}{\Delta t}\int_{-\infty}^{\infty}(x-\zeta)\rho_{\mathbf{z}}(t+\Delta t,x;t,\zeta)dx=g(t,\zeta;\mathbf{z}),\\&\lim_{\Delta t\to 0+}\frac{1}{\Delta t}\int_{-\infty}^{\infty}|x-\zeta|^{\kappa}\rho_{\mathbf{z}}(t+\Delta t,x;t,\zeta)dx=0,\ \text{where }\kappa\geq 2.\end{aligned}\tag{3.25}$$

With conditions (3.25), we can use similar arguments as those in the proof for deriving equation (2.17) to obtain the population density $v(t,x;\mathbf{z})$ in a subpopulation with growth rate $g(t,x;\mathbf{z})$ (see, e.g., [25])

$$\frac{\partial}{\partial t}v(t,x;\mathbf{z})+\frac{\partial}{\partial x}(g(t,x;\mathbf{z})v(t,x;\mathbf{z}))+d(t,x)v(t,x;\mathbf{z})=0,$$

where $d(t,x)$ denotes the death rate of individuals with size x at time t. Given the subpopulation density $v(t,x;\mathbf{z})$, the population density $u(t,x)$ for the total population with size x at time t is given by

$$u(t,x)=\int_{\mathbb{Z}}v(t,x;\mathbf{z})\tilde{\varphi}_{\mathbf{Z}}(\mathbf{z})d\mathbf{z}.$$

We can generalize the above results to a population which is structured by multiple structure variable $\mathbf{x}\in\mathbb{R}^n$. Let $v(t,\mathbf{x};\mathbf{z})$ denote the population density of a subpopulation with growth rate $\mathbf{g}(t,\mathbf{x};\mathbf{z})$. Then v satisfies

$$\frac{\partial}{\partial t}v(t,\mathbf{x};\mathbf{z})+\sum_{j=1}^{n}\frac{\partial}{\partial x_j}(g_j(t,\mathbf{x};\mathbf{z})v(t,\mathbf{x};\mathbf{z}))+d(t,\mathbf{x})v(t,\mathbf{x};\mathbf{z})=0.$$

Given the subpopulation density $v(t,\mathbf{x};\mathbf{z})$, the population density $u(t,\mathbf{x})$ for the total population with structure variables $\mathbf{x}$ at time t is given by

$$u(t,\mathbf{x})=\int_{\mathbb{Z}}v(t,\mathbf{x};\mathbf{z})\tilde{\varphi}_{\mathbf{Z}}(\mathbf{z})d\mathbf{z}.$$

Based on these discussions, we see that growth rate distributed structured population model can be associated with certain stochastic processes, which are obtained due to variability in the individual's growth rate and also the variability in the initial size of individuals in the population.

3.3 Differential Equations Driven by Correlated Stochastic Processes

We turn next to random differential equations driven by stochastic processes and consider in this context the system of random ordinary differential equations

$$\dot{\mathbf{x}}(t;\mathbf{X}_0,\mathbf{Y}(t)) = \mathbf{g}(t,\mathbf{x}(t;\mathbf{X}_0,\mathbf{Y}(t));\mathbf{Y}(t)), \quad \mathbf{x}(0)=\mathbf{X}_0, \tag{3.26}$$

with states $\mathbf{x}=(x_1,x_2,\ldots,x_n)^T$, $\mathbf{g}=(g_1,g_2,\ldots,g_n)^T$ an n-dimensional non-random vector function of t and $\mathbf{x}$, and the driving process $\mathbf{Y}(t)=(Y_1(t),Y_2(t),\ldots,Y_\nu(t))^T$ given by a ν-dimensional correlated stochastic process independent of random initial vector $\mathbf{X}_0$. Moreover we allow that $\mathbf{Y}(t)$ may be dependent on the states, i.e., there exists a feedback between the states of the system (3.26) and the random driving process. Our formulation is partly motivated by applications in optimal control.

By Remark 2.1 we know that (3.26) with the driving process $\{\mathbf{Y}(t) : t \geq 0\}$ taken as a Markovian Gaussian colored noise is often found useful in a wide range of applications. In addition, equation (3.26) with the driving process chosen as a continuous time Markov chain (also referred to as Markov jump process) has been used in engineering for studying the dynamic reliability of a complex system. For example, it was used in [48] to study the reliability of the Ignalina nuclear power plant accident localization system (ALS), which was designed to protect the environment from a radioactive release after the piping rupture in the RBMK-1500 primary reactor cooling system. In this study, the states of the system consist of two components with one representing pressure and the other describing temperature, and the driving process $\mathbf{Y}(t)=(Y_1(t),Y_2(t))^T$ with $Y_1(t)$ and $Y_2(t)$ respectively denoting the number of failures of pumps and heat exchangers at time t. Equation (3.26) with a continuous time Markov chain driving process has also been found useful in a wide range of other fields including the environmental sciences (e.g., [64]) and physiology (e.g., in studies of electrically excitable nerve membranes in [42, Section 9.4]).

3.3.1 Joint Probability Density Function of $(\mathbf{x}(t;\mathbf{X}_0,\mathbf{Y}(t)),\mathbf{Y}(t))^T$

We first consider the evolution of the joint probability density function of $\tilde{\mathbf{X}}(t)$, where $\tilde{\mathbf{X}}(t)=(\mathbf{X}(t),\mathbf{Y}(t))^T$ with $\mathbf{X}(t)=\mathbf{x}(t;\mathbf{X}_0,\mathbf{Y}(t))$ being the solution to (3.26). Specifically, two cases are discussed. One is with driving process taken as a diffusion process, and the other is with driving process given by a continuous time Markov chain (CTMC) with finite state space.

A diffusion driving process: First we consider the case where $\{\mathbf{Y}(t) : t \geq 0\}$ is a diffusion process satisfying the Itô stochastic differential equations

$$d\mathbf{Y}(t) = \boldsymbol{\mu}_{\mathbf{Y}}(t,\mathbf{X}(t),\mathbf{Y}(t))dt + \boldsymbol{\sigma}_{\mathbf{Y}}(t,\mathbf{X}(t),\mathbf{Y}(t))d\mathbf{W}(t), \quad \mathbf{Y}(0)=\mathbf{Y}_0, \tag{3.27}$$

where $\boldsymbol{\mu}_{\mathbf{Y}}(t,\mathbf{x},\mathbf{y}) = (\mu_{\mathbf{Y},1}(t,\mathbf{x},\mathbf{y}),\mu_{\mathbf{Y},2}(t,\mathbf{x},\mathbf{y}),\ldots,\mu_{\mathbf{Y},\nu}(t,\mathbf{x},\mathbf{y}))^T$ is the drift coefficient, $\boldsymbol{\Sigma}_{\mathbf{Y}}(t,\mathbf{x},\mathbf{y}) = \boldsymbol{\sigma}_{\mathbf{Y}}(t,\mathbf{x},\mathbf{y})\,(\boldsymbol{\sigma}_{\mathbf{Y}}(t,\mathbf{x},\mathbf{y}))^T$ is the diffusion coefficient with $\boldsymbol{\sigma}_{\mathbf{Y}}$ being a non-

random $\nu \times l$ matrix function of t, $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{W}(t)$ is an l-dimensional standard Wiener process. Then by (3.26) and (3.27) we know that $\tilde{\mathbf{X}}(t)$ satisfies the Itô stochastic differential equation

$$d\tilde{\mathbf{X}}(t) = \boldsymbol{\mu}_{\tilde{\mathbf{X}}}(t, \tilde{\mathbf{X}}(t))dt + \boldsymbol{\sigma}_{\tilde{\mathbf{X}}}(t, \tilde{\mathbf{X}}(t))d\mathbf{W}(t), \quad \tilde{\mathbf{X}}(0) = \tilde{\mathbf{X}}_0, \tag{3.28}$$

where $\boldsymbol{\mu}_{\tilde{\mathbf{X}}}(t, \tilde{\mathbf{X}}(t)) = \begin{bmatrix} \mathbf{g}(t, \mathbf{x}(t; \mathbf{X}_0, \mathbf{Y}(t)); \mathbf{Y}(t)) \\ \boldsymbol{\mu}_{\mathbf{Y}}(t, \mathbf{X}(t), \mathbf{Y}(t)) \end{bmatrix}$, $\boldsymbol{\sigma}_{\tilde{\mathbf{X}}}(t, \tilde{\mathbf{X}}(t)) = \begin{bmatrix} \mathbf{O}_{n\times l} \\ \boldsymbol{\sigma}_{\mathbf{Y}}(t, \mathbf{X}(t), \mathbf{Y}(t)) \end{bmatrix}$ with $\mathbf{O}_{n\times l}$ being an $n \times l$ zero matrix, and $\tilde{\mathbf{X}}_0 = (\mathbf{X}_0, \mathbf{Y}_0)^T$. Thus, if we assume that $\tilde{\mathbf{X}}_0$ is independent of $\mathbf{W}(t)$, and $\boldsymbol{\mu}_{\tilde{\mathbf{X}}}$ and $\boldsymbol{\sigma}_{\tilde{\mathbf{X}}}$ satisfy conditions guaranteeing the existence and uniqueness of solutions to the initial value problem (3.28), then by Theorem 2.2 we obtain the following results on the evolution of the joint probability density function of $\tilde{\mathbf{X}}(t)$.

Theorem 3.4. *Assume that* $\{\mathbf{Y}(t) : t \geq 0\}$ *is a diffusion process satisfying the Itô stochastic differential equations* (3.27)*. Then the joint probability density function* $\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t, \mathbf{x}, \mathbf{y})$ *of* $\tilde{\mathbf{X}}(t)$ *satisfies*

$$\begin{aligned} &\frac{\partial}{\partial t}\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t, \mathbf{x}, \mathbf{y}) + \sum_{k=1}^{n} \frac{\partial}{\partial x_k}(g_k(t, \mathbf{x}; \mathbf{y})\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t, \mathbf{x}, \mathbf{y})) \\ &\quad + \sum_{k=1}^{\nu} \frac{\partial}{\partial y_k}(\mu_{\mathbf{Y},k}(t, \mathbf{x}, \mathbf{y})\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t, \mathbf{x}, \mathbf{y})) = \frac{1}{2}\sum_{k,j=1}^{\nu} \frac{\partial^2}{\partial y_k \partial y_j}\left[\Sigma_{\mathbf{Y},kj}(t, \mathbf{x}, \mathbf{y})\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t, \mathbf{x}, \mathbf{y})\right] \end{aligned} \tag{3.29}$$

with initial condition

$$\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(0, \mathbf{x}, \mathbf{y}) = \tilde{\varphi}^0_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}),$$

where $\Sigma_{\mathbf{Y},kj}(t, \mathbf{x}, \mathbf{y})$ *is the* (k, j)*th element of* $\boldsymbol{\Sigma}_{\mathbf{Y}}(t, \mathbf{x}, \mathbf{y})$*, and* $\tilde{\varphi}^0_{\mathbf{X},\mathbf{Y}}$ *is the joint probability density function of* $\mathbf{X}_0$ *and* $\mathbf{Y}_0$*.*

It should be noted that the results presented in the above theorem also apply to the case where $\mathbf{Y}(t)$ is not affected by the states of the system (that is, $\boldsymbol{\mu}_{\mathbf{Y}}$ and $\boldsymbol{\sigma}_{\mathbf{Y}}$ are independent of $\mathbf{X}$). In addition, we observe that the resulting equation (3.29) is a special form of the Fokker-Planck equation (2.7), and this equation with $n = 1$, $g_1 \equiv 1$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}(t, x, \mathbf{y})$ being a diagonal matrix is a special case of another model given in 1974 by Oster and Takahashi [61]

$$\begin{aligned} &\frac{\partial}{\partial t}u(t, x, \mathbf{y}) + \frac{\partial}{\partial x}u(t, x, \mathbf{y}) + \sum_{k=1}^{\nu} \frac{\partial}{\partial y_k}(\mu_{\mathbf{Y},k}(t, x, \mathbf{y})u(t, x, \mathbf{y})) \\ &\qquad + d(t, x, \mathbf{y})u(t, x, \mathbf{y}) = \frac{1}{2}\sum_{k=1}^{\nu} \frac{\partial^2}{\partial y_k^2}\left[\Sigma_{\mathbf{Y},kk}(t, x, \mathbf{y})u(t, x, \mathbf{y})\right] \end{aligned} \tag{3.30}$$

without mortality ($d \equiv 0$). Equation (3.30) is used to describe the population density $u(t, x, \mathbf{y})$ in a spatially homogeneous population, where individuals are characterized by the chronological age x, and the physiological variables $\mathbf{y}$ whose growth rates are affected by various stochastic factors. The function $\boldsymbol{\mu}_{\mathbf{Y}}(t, x, \mathbf{y})$, $\boldsymbol{\Sigma}_{\mathbf{Y}}(t, x, \mathbf{y})$ and $d(t, x, \mathbf{y})$ in this model denote the mean growth rates, diffusion coefficient for the growth rates, and mortality rate of individuals with chronological age x and physiological variables $\mathbf{y}$ at time t, respectively.

Observe that (3.30) includes the structured population model (3.7) as a special case (specifically, when $\mathbf{\Sigma_Y}$ is a zero matrix). In addition, (3.30) with $\nu = 1$

$$\begin{aligned} &\frac{\partial}{\partial t}u(t,x,y) + \frac{\partial}{\partial x}u(t,x,y) + \frac{\partial}{\partial y}(\mu_Y(t,x,y)u(t,x,y)) \\ &\qquad +d(t,x,y)u(t,x,y) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[\Sigma_Y(t,x,y)u(t,x,y)\right] \end{aligned} \tag{3.31}$$

is a model given in 1968 by Weiss [81] for growth of cell populations.

By the discussions in Section 2.1.2 we know that model (3.30) can be associated with a killed diffusion process, and the arguments presented there can be applied to obtain (3.30).

A continuous time Markov chain driving process: For simplicity, we consider the case where the driving process is a one-dimensional (i.e., $\nu = 1$) continuous time Markov chain (CTMC) with finite state space, and this chain is time-homogeneous and unaffected by the states of the system. In addition, we assume that $\mathbf{g}$ is time independent, that is, system (3.26) becomes

$$\dot{\mathbf{x}}(t;\mathbf{X}_0,Y(t)) = \mathbf{g}(\mathbf{x}(t;\mathbf{X}_0,Y(t));Y(t)), \quad \mathbf{x}(0) = \mathbf{X}_0. \tag{3.32}$$

Without loss of generality, the time-homogeneous CTMC $\{Y(t) : t \geq 0\}$ is assumed to have state space $\{1,2,3,\ldots,l\}$. Note that system (3.32) describes the movement of particles (or individuals) under the action of l dynamical systems corresponding to the equations $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x};j)$, $j = 1,2,\ldots,l$, and $Y(t)$ determines which dynamical system acts at time t. Hence, under these conditions the system (3.32) is referred to as a *randomly controlled dynamical system* (see [68]).

We remark that (3.32) is a special case of *random evolutions*, an abstract mathematical model for describing evolution systems in a general (either Markovian or non-Markovian) random medium, introduced in 1969 by Griego and Hersh [38], where certain abstract Cauchy problems were found to be related to the generators of Markov processes. There are a wide range of theoretic results on random evolutions including representation theory, control theory, stability analysis, limit theorems in averaging and diffusion approximations with respect to the properties (such as ergodicity and reducibility) of the random medium. The investigations of random evolutions are often carried out by one of two approaches, one using asymptotic perturbations theory, and the other using a martingale approach. Interested readers can refer to [43, 49, 50, 77] and the references therein for further details.

Let q_{jk} denote the transition rate of $Y(t)$ from state k to state j, that is,

$$\text{Prob}\{Y(t+\Delta t) = j \mid Y(t) = k\} = q_{jk}\Delta t + o(\Delta t).$$

Then we have

$$q_{jk} \geq 0 \text{ (for } j \neq k) \quad \text{and} \quad q_{jj} = -\sum_{k\neq j} q_{jk}.$$

Define the functions $\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},j)$ and $\tilde{\varphi}^0_{\mathbf{X},Y}(\mathbf{x},j)$ by

$$\text{Prob}\{\tilde{\mathbf{X}}(t)\in\mathbb{B}\times\{j\}\}=\int_{\mathbb{B}}\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},j)d\mathbf{x},\quad j=1,2,\ldots,l,$$

and

$$\text{Prob}\{\tilde{\mathbf{X}}(0)\in\mathbb{B}\times\{j\}\}=\int_{\mathbb{B}}\tilde{\varphi}^0_{\mathbf{X},Y}(\mathbf{x},j)d\mathbf{x},\quad j=1,2,\ldots,l,$$

respectively, where $\tilde{\mathbf{X}}(t)=(\mathbf{X}(t),Y(t))^T$. Let $\mathcal{Q}=(q_{jk})_{l\times l}$, and

$$\begin{aligned}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})&=(\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},1),\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},2),\ldots,\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},l))^T,\\ \tilde{\boldsymbol{\varphi}}^0_{\mathbf{X},Y}(\mathbf{x})&=\left(\tilde{\varphi}^0_{\mathbf{X},Y}(\mathbf{x},1),\tilde{\varphi}^0_{\mathbf{X},Y}(\mathbf{x},2),\ldots,\tilde{\varphi}^0_{\mathbf{X},Y}(\mathbf{x},l)\right)^T.\end{aligned}$$

Then the coupled stochastic process $\left\{\tilde{\mathbf{X}}(t):t\geq 0\right\}$ is a Markov process on $\mathbb{R}^n\times\{1,2,\ldots,l\}$, and $\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}$ satisfies the following system of partial differential equations (e.g., see [68])

$$\frac{\partial}{\partial t}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})=\mathscr{A}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})+\mathcal{Q}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}) \tag{3.33}$$

with initial condition $\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(0,\mathbf{x})=\tilde{\varphi}^0_{\mathbf{X},\mathbf{Y}}(\mathbf{x})$. Here

$$\mathscr{A}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})=\left(\mathscr{A}_1\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}),\mathscr{A}_2\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}),\ldots,\mathscr{A}_l\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})\right)^T,$$

where $\mathscr{A}_j\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})=-\sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(\mathbf{x};j)\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},j))$, $j=1,2,\ldots,l$.

Remark 3.5. We remark that the results presented in the above two cases can be easily applied to the scenario where differential equations are driven by both colored noise and white noise. We take the random differential equation

$$\begin{aligned}\dot{\mathbf{X}}(t)&=\mathbf{g}(\mathbf{X}(t);Y(t))+\boldsymbol{\sigma}(\mathbf{X}(t);Y(t))d\mathbf{W}(t)\\ \mathbf{X}(0)&=\mathbf{X}_0,\end{aligned} \tag{3.34}$$

as an example to illustrate this. Here $Y(t)$ is a one-dimensional time-homogeneous CTMC with state space $\{1,2,\ldots,l\}$ which is independent of the Wiener process $\mathbf{W}(t)$ and the initial condition $\mathbf{X}_0$. We see that (3.34) describes the movement of particles (or individuals) under the action of l stochastic systems corresponding to $\dot{\mathbf{X}}(t)=\mathbf{g}(\mathbf{X}(t);j)+\boldsymbol{\sigma}(\mathbf{X}(t);j)d\mathbf{W}(t)$, $j=1,2,\ldots,l$, and $Y(t)$ determines which stochastic system acts at time t. Hence, this type of differential equation is referred to as a *switching diffusion system*, *stochastic differential equations with Markov switching*, *hybrid switching diffusion*, or *randomly flashing diffusion*. Due to its applications in a wide range of fields such as financial mathematics, engineering, biological systems, and manufacturing systems (e.g., see [85] and the references therein for details), this type of differential equation has received considerable research attention in recent years. Interested readers can refer to the recent texts [53, 86] and the references therein for information on the recent advance of this subject. Let $\tilde{\mathbf{X}}(t)=(\mathbf{X}(t),Y(t))^T$ with

$\mathbf{X}(t)$ being the solution to (3.34). Then the coupled stochastic process $\left\{\tilde{\mathbf{X}}(t) : t \geq 0\right\}$ is still a Markov process on $\mathbb{R}^n \times \{1, 2, \ldots, l\}$, and $\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}$ satisfies the following system of partial differential equations (again see [68])

$$\frac{\partial}{\partial t}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}) = \mathscr{A}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}) + \mathcal{Q}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}). \tag{3.35}$$

Here $\mathscr{A}\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}) = \left(\mathscr{A}_1\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}), \mathscr{A}_2\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}), \ldots, \mathscr{A}_l\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x})\right)^T$ with the differential operators $\mathscr{A}_r$, $r = 1, 2, \ldots, l$ defined by

$$\mathscr{A}_r\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}(t,\mathbf{x}) = -\sum_{k=1}^{n}\frac{\partial}{\partial x_k}(g_k(\mathbf{x};r)\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},r)) + \frac{1}{2}\sum_{k,j=1}^{n}\frac{\partial^2}{\partial x_k \partial x_j}\left[\Sigma_{kj}(\mathbf{x};r)\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},r))\right],$$

where $\Sigma_{kj}(\mathbf{x};r)$ denotes the (k,j)th element of matrix $\boldsymbol{\sigma}(\mathbf{x};r)\boldsymbol{\sigma}(\mathbf{x};r)^T$.

Remark 3.6. It should be noted that for all the above cases the primary process $\{\mathbf{X}(t) : t \geq 0\}$ is not Markovian even though the coupled stochastic process $\left\{\tilde{\mathbf{X}}(t) : t \geq 0\right\}$ is. This is because the evolution of the states of the system depends on the past history due to the autocorrelation (in time) of the driving process.

3.3.2 The Probability Density Function of $\mathbf{X}(t)$

In this section, we consider the probability density function of $\mathbf{X}(t)$ based on the two cases that we discussed in Section 3.3.1, that is, the case with driving process being a diffusion process, and the case with driving process being a continuos time Markov chain.

A diffusion driving process: For the given joint probability density function $\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t,\mathbf{x},\mathbf{y})$, we can obtain the probability density function of $\mathbf{X}(t)$

$$p(t,\mathbf{x}) = \int_{\mathbb{Y}} \tilde{\varphi}_{\mathbf{X},\mathbf{Y}}(t,\mathbf{x},\mathbf{y})d\mathbf{y}, \tag{3.36}$$

where $\mathbb{Y}$ denotes the set of all possible values for $\mathbf{y}$. Equation (3.36) combined with equation (3.29) for $\tilde{\varphi}_{\mathbf{X},\mathbf{Y}}$ can be used to describe the population density in a population ecology system where the growth rates $\mathbf{g}$ of the species are affected by a driving diffusion process. This could be due to the climate variability, which is often described by Itô stochastic differential equations (based on the discussions in Section 2.1).

A continuous time Markov chain driving process: Based on the discussions in Section 3.3.1, we see that under this case the probability density function $p(t,\mathbf{x})$ of $\mathbf{X}(t)$ is given by

$$p(t,\mathbf{x}) = \sum_{j=1}^{l}\tilde{\varphi}_{\mathbf{X},Y}(t,\mathbf{x},j). \tag{3.37}$$

Equation (3.37) combined with equation (3.33) for $\tilde{\boldsymbol{\varphi}}_{\mathbf{X},Y}$ can be used to describe the population density in a population where the growth rates $\mathbf{g}$ of the species are affected by a continuous time Markov chain driving process. For example, for the case $l = 2$ it could be used to describe a population in an environment with random switching between two opposite conditions such as drought-nondrought and stressed-unstressed conditions.

Remarks on the equation for the probability density function of $\mathbf{X}(t)$ In general, one is unable to obtain a closed form for the equation of probability density function of the solution to (3.26) even in the one-dimensional case. Two scenarios of the following scalar random differential equation

$$\dot{x} = f(x) + \sigma(x)Y(t), \quad x(0) = X_0 \tag{3.38}$$

were considered in [41] where the closed form can be derived for the equation of probability density function of the resulting process. One case is when f is an affine function of x (i.e., $f(x) = a_0 - a_1 x$), $\sigma \equiv 1$, and $Y(t)$ is a general stationary Gaussian colored noise (being either Markovian or non-Markovian) with zero-mean and correlation function $h(|t-s|)$. The evolution of the probability density function of the resulting stochastic process is described by the Fokker-Planck-like equation

$$\frac{\partial}{\partial t}p(t,x) + \frac{\partial}{\partial x}\left[(a_0 - a_1 x)p(t,x)\right] = \left[\int_0^t h(t-s)\exp(-a_1(t-s))ds\right]\frac{\partial^2}{\partial x^2}p(t,x). \tag{3.39}$$

We observe that even with a time-independent function f and stationary Gaussian noise, the effective diffusion in (3.39) is time-dependent and may even become negative in the cases where the correlation function is an oscillatory-like function.

The other case that permits a closed form equation for the density function is when $Y(t) = a(-1)^{N(t)}$ with a being a constant and $N(t)$ being a Poisson process with parameter λ. In other words, $Y(t)$ is a continuous time Markov chain with two states, which is often referred to as a dichotomous Markovian noise and also known as the random telegraph signal. In this case, $Y(t)$ has a time-inhomogeneous mean function, but it has a time-homogeneous exponential correlation function given by

$$\mathrm{Cor}\{Y(t), Y(s)\} = a^2 \exp\left(-2\lambda|t-s|\right).$$

The evolution of the probability density function is described by the equation

$$\begin{aligned}&\frac{\partial}{\partial t}p(t,x) + \frac{\partial}{\partial x}(f(x)p(t,x)) \\ &\quad = a^2\frac{\partial}{\partial x}\left[\sigma(x)\int_0^t \exp\left(-\left(2\lambda + f'(x)\right)(t-s)\right)\left(\frac{\partial}{\partial x}(\sigma(x)p(s,x))\right)ds\right],\end{aligned} \tag{3.40}$$

where f' denotes the derivative of f with respect to x. We observe from (3.40) that the evolution of the probability density function $p(t,x)$ depends on its past history, which clearly

reveals that the stochastic process $\{X(t) : t \geq 0\}$ is non-Markovian. Moreover, if $f \equiv 0$ and $\sigma \equiv 1$, then by (3.40) it is easy to see that $p(t,x)$ satisfies telegrapher's equation

$$\frac{\partial^2}{\partial t^2}p(t,x) + 2\lambda\frac{\partial}{\partial t}p(t,x) = a^2\frac{\partial^2}{\partial x^2}p(t,x).$$

The method for derivation of (3.39) and (3.40) in [41] is based on the expression for $p(t,x)$ given by

$$p(t,x) = \mathbb{E}\{\delta(X(t) - x)\}.$$

One then differentiates the above equation with respect to t, and uses the statistics of the driving process $Y(t)$ as well as some functional derivative technique to obtain the desired equation.

4 Relationships Between Random and Stochastic Differential Equations

As noted in Section 2.2, the solution of Stratonovich stochastic differential equations can be obtained as the limit of the solutions of some random differential equations. In this section, we will given some other relationships between stochastic and random differential equations. Specifically, the probability density functions of the solutions to stochastic differential equations and crypto-deterministic differential equations can be associated with Markov operators as detailed in the next section. In addition, there is a class of Itô stochastic differential equations with solutions having the same probability density function at each time t as those for the solution of their corresponding random differential equations. These are summarized in Section 4.2.

4.1 Markov Operators and Markov Semigroups

Markov operators and Markov semigroups have been widely used to investigate dynamical systems and dynamical systems with stochastic perturbations as they have been proven to be suitable to study the evolution of probability density functions of the resulting processes. Hence, in this section we focus on Markov operators and Markov semigroups acting on the set of probability density functions.

Let (Ω, Σ, μ) be a σ-finite measure space, and $L^1(\Omega, \Sigma, \mu)$ be the Banach space of all possible real-valued measurable functions $f : \Omega \to \mathbb{R}$ satisfying $\displaystyle\int_\Omega |f(\mathbf{x})|\mu(d\mathbf{x}) < \infty$ with norm defined by

$$\|f\|_{L^1(\Omega,\Sigma,\mu)} = \int_\Omega |f(\mathbf{x})|\mu(d\mathbf{x}).$$

In addition, $L^\infty(\Omega, \Sigma, \mu)$ denotes the Banach space of all possible real-valued measurable functions that are bounded almost everywhere with the norm defined by

$$\|h\|_{L^\infty(\Omega,\Sigma,\mu)} = \operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} |h(\mathbf{x})|.$$

With the above notation, Markov operators are defined in the usual manner (e.g., see [52, Definition 3.1.1] or [33, 68]).

Definition 4.1. Let (Ω, Σ, μ) be a σ-finite measure space, and $\mathbb{D}$ be the subset of Banach space $L^1(\Omega, \Sigma, \mu)$ functions which contains all possible densities

$$\mathbb{D} = \left\{ f \in L^1(\Omega, \Sigma, \mu) \;\middle|\; f \geq 0, \|f\|_{L^1(\Omega,\Sigma,\mu)} = 1 \right\}.$$

A linear mapping $\mathscr{P} : L^1(\Omega, \Sigma, \mu) \to L^1(\Omega, \Sigma, \mu)$ is called a *Markov operator* if $\mathscr{P}(\mathbb{D}) \subset \mathbb{D}$.

By the above definition, we see that a Markov operator is a linear operator that maps any probability density function to a probability density function. Markov operators have a number of nice properties such as monotonicity and contractivity. In addition, if Ω is a Polish space (i.e., a complete separable metric space), Σ is the σ-algebra of Borel subsets of Ω, and μ is a probability Borel measure on Ω, then every Markov operator on $L^1(\Omega, \Sigma, \mu)$ is given by a transition probability function (e.g., see [68]). For more information on Markov operators and their applications, interested readers can refer to [52, 68] and the references therein.

Definition 4.2. (e.g., see [68]) A family $\{\mathscr{P}(t)\}_{t\geq 0}$ of Markov operators which satisfies conditions

- $\mathscr{P}(0) = \mathscr{I}$
- $\mathscr{P}(t+s) = \mathscr{P}(t)\mathscr{P}(s)$
- for each $f \in L^1(\Omega, \Sigma, \mu)$ the function $t \to \mathscr{P}(t)f$ is continuous

is called a *Markov semigroup*. Here $\mathscr{I}$ denotes the unity operator.

An Example of Markov operators: The Frobenius-Perron operator: Frobenius-Perron operators are a special case of Markov operators, and are very important in the ergodic theory of chaotic dynamical systems (see [33] for more information). Before we give the precise definition, we need the following definition on nonsingular transformations.

Definition 4.3. Let (Ω, Σ, μ) be a σ-finite measure space. A transformation $\boldsymbol{\pi} : \Omega \to \Omega$ is *measurable* if

$$\boldsymbol{\pi}^{-1}(\mathbb{O}) \in \Sigma, \text{ for all } \mathbb{O} \in \Sigma.$$

In addition, if $\mu(\boldsymbol{\pi}^{-1}(\mathbb{O})) = 0$ for all $\mathbb{O} \in \Sigma$ such that $\mu(\mathbb{O}) = 0$, then $\boldsymbol{\pi}$ is *nonsingular*.

With the above definition, the Frobenius-Perron operator is defined as follows (e.g., see [52, Definition 3.2.3]).

Definition 4.4. Let (Ω, Σ, μ) be a σ-finite measure space. If $\boldsymbol{\pi} : \Omega \to \Omega$ is a nonsingular transformation, then the unique operator $\mathscr{P} : L^1(\Omega, \Sigma, \mu) \to L^1(\Omega, \Sigma, \mu)$ defined by

$$\int_{\mathbb{O}} \mathscr{P} f(\mathbf{x}) \mu(d\mathbf{x}) = \int_{\boldsymbol{\pi}^{-1}(\mathbb{O})} f(\mathbf{x}) \mu(d\mathbf{x}), \quad \mathbb{O} \in \Sigma \tag{4.1}$$

is called the *Frobenius-Perron operator* corresponding to $\boldsymbol{\pi}$.

It is important to note that in some special cases where the transformation $\boldsymbol{\pi}$ is differentiable and invertible, then the explicit form for $\mathscr{P} f$ is available. This is given in the following lemma (e.g., see [52, Corollary 3.2.1]) .

Lemma 4.5. *Let (Ω, Σ, μ) be a σ-finite measure space, and $\boldsymbol{\pi} : \Omega \to \Omega$ be an invertible nonsingular transformation (i.e., $\boldsymbol{\pi}^{-1}$ nonsingular) and $\mathscr{P}$ be the associated Frobenius-Perron operator. Then for every $f \in L^1(\Omega, \Sigma, \mu)$ we have*

$$\mathscr{P}(f(\mathbf{x})) = f(\boldsymbol{\pi}^{-1}(\mathbf{x})) \mathcal{J}^{-1}(\mathbf{x}), \tag{4.2}$$

where $\mathcal{J}(\mathbf{x})$ is the determinant of the Jacobian matrix $\dfrac{\partial \pi(\mathbf{x})}{\partial \mathbf{x}}$, that is, $\mathcal{J}(\mathbf{x}) = \left| \dfrac{\partial \pi(\mathbf{x})}{\partial \mathbf{x}} \right|$.

Adjoint of the Frobenius-Perron operator: The operator adjoint to the Frobenius-Perron operator is called the *Koopman operator*; this is widely used in applications as it can be used to describe the evolution of the observables of a system on the phase space.

Definition 4.6. (See [52, Definition 3.3.1]) Let (Ω, Σ, μ) be a σ-finite measure space. If $\boldsymbol{\pi} : \Omega \to \Omega$ is a nonsingular transformation and $h \in L^\infty(\Omega, \Sigma, \mu)$. Then the operator $\mathscr{U} : L^\infty(\Omega, \Sigma, \mu) \to L^\infty(\Omega, \Sigma, \mu)$ defined by

$$\mathscr{U}(h(\mathbf{x})) = h(\boldsymbol{\pi}(\mathbf{x})) \tag{4.3}$$

is called the *Koopman operator* with respect to $\boldsymbol{\pi}$.

The Koopman operator has the following important properties:

- For any $h_1, h_2 \in L^\infty(\Omega, \Sigma, \mu)$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ we have

  $$\mathscr{U}(\lambda_1 h_1 + \lambda_2 h_2) = \lambda_1 \mathscr{U}(h_1) + \lambda_2 \mathscr{U}(h_2). \tag{4.4}$$

  This implies that the Koopman operator $\mathscr{U}$ is linear.

- For every $h \in L^\infty(\Omega, \Sigma, \mu)$, we have

  $$\|\mathscr{U}(h)\|_{L^\infty(\Omega,\Sigma,\mu)} \leq \|h\|_{L^\infty(\Omega,\Sigma,\mu)}, \tag{4.5}$$

  thus, the Koopman operator $\mathscr{U}$ is a contraction on $L^\infty(\Omega, \Sigma, \mu)$.

- For every $f \in L^1(\Omega, \Sigma, \mu)$ and $h \in L^\infty(\Omega, \Sigma, \mu)$ we have

$$\langle \mathscr{P} f, h \rangle = \langle f, \mathscr{U} h \rangle, \tag{4.6}$$

  which implies that the Koopman operator $\mathscr{U}$ is adjoint to the Frobenius-Perron operator $\mathscr{P}$.

Interested readers can refer to [52] for more information on Frobenius-Perron operators and Koopman operators.

4.1.1 Random Differential Equations

Assume that the system of random crypto-deterministic formulations given by the differential equations in (3.1)

$$\dot{\mathbf{x}} = \mathbf{g}(t, \mathbf{x}), \quad \mathbf{x}(0) = \mathbf{X}_0$$

has an explicit mean square solution $\mathbf{X}(t)$. Then the solution takes the general form

$$\mathbf{X}(t) = \boldsymbol{\pi}(t, \mathbf{X}_0), \tag{4.7}$$

which essentially represents an algebraic transformation of a set of random variables into another. In addition, we observe that in continuum mechanics the realization of (4.7) (i.e., $\mathbf{x}(t) = \boldsymbol{\pi}(t, \mathbf{x}_0)$) provides a description of particles moving in space as time progresses, that is, it provides a Lagrangian description of the motion of the particles.

The following theorem indicates that the probability density function $p(t, \mathbf{x})$ of $\mathbf{X}(t)$ at any given time t can be expressed in terms of the probability density function $p_0(\mathbf{x})$ of the initial condition $\mathbf{X}_0$.

Theorem 4.7. *(See [72, Theorem 6.2.1]) Suppose that the transformation defined by* (4.7) *is continuous in* $\mathbf{X}_0$ *and has continuous partial derivative with respect to* $\mathbf{X}_0$*, and defines a one-to-one mapping. Let the inverse transform be written by*

$$\mathbf{X}_0 = \boldsymbol{\pi}^{-1}(t, \mathbf{X}(t)). \tag{4.8}$$

Then the density function $p(t, \mathbf{x})$ *of* $\mathbf{X}(t)$ *is given by*

$$p(t, \mathbf{x}) = p_0(\boldsymbol{\pi}^{-1}(t, \mathbf{x}))|\mathcal{J}^{-1}|, \tag{4.9}$$

where $\mathcal{J}$ *is the determinant of Jacobian matrix, that is,* $\mathcal{J} = \left| \dfrac{\partial \mathbf{x}}{\partial \mathbf{x}_0} \right|$.

Observe that in continuum mechanics the realization of (4.8) (i.e., $\mathbf{x}_0 = \boldsymbol{\pi}^{-1}(t, \mathbf{x}(t))$) provides a tracing of the particle which now occupies the position $\mathbf{x}$ in the current configuration from its original position $\mathbf{x}_0$ in the initial configuration, and thus it provides an Eulerian

description of the motion of particles. Hence, (4.9) gives an explicit formula for the density of material $p(t, \mathbf{x})$ in the current configuration in terms of the density of material $p_0(\mathbf{x})$ in the original configuration, where $\mathcal{J}$ in this case denotes the determinant of the *configuration gradient* (often, in something of a misnomer, referred to as the *deformation gradient* in the literature-see [21]).

By Lemma 4.5 and Theorem 4.7 we know that for any given t the probability density function $p(t, \mathbf{x})$ can be expressed in terms of the Frobenius-Perron operator $\mathscr{P}(t)$ (corresponding to the map $\boldsymbol{\pi}(t, \cdot)$), that is,

$$p(t, \mathbf{x}) = \mathscr{P}(t)p_0(\mathbf{x}),$$

where the operator $\mathscr{P}(t)$ is given by

$$\mathscr{P}(t)f(\mathbf{x}) = f(\boldsymbol{\pi}^{-1}(t, \mathbf{x}))|\mathcal{J}^{-1}|.$$

Moreover, if $\mathbf{g}$ is time-independent, then $\{\mathscr{P}(t)\}_{t\geq 0}$ forms a Frobenius-Perron semigroup with infinitesimal generator given by (e.g., see [52, Chapter 7])

$$\mathscr{A}\psi(\mathbf{x}) = -\sum_{k=1}^{n} \frac{\partial}{\partial x_k} \left(g_k(\mathbf{x})\psi(\mathbf{x})\right), \tag{4.10}$$

which implies that $p(t, \mathbf{x})$ satisfies Liouville's equation

$$\frac{\partial}{\partial t}p(t, \mathbf{x}) = \mathscr{A}p(t, \mathbf{x}).$$

Remark 4.8. Consider the time-homogeneous case of (3.1) (i.e., $\mathbf{g}$ is time-independent) with mean-square solution $\pi(t, \mathbf{X}_0)$. Then by the definition of the Koopman operator we have

$$\mathscr{U}(t)\phi(\mathbf{x}) = \phi(\pi(t, \mathbf{x})).$$

Then $\{\mathscr{U}(t)\}_{t\geq 0}$ forms a semigroup (if we restrict ourself to continuous functions with compact support) with infinitesimal generator given by (e.g., see [52, Section 7.6])

$$\mathscr{L}\phi(\mathbf{x}) = \sum_{k=1}^{n} g_k(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_k}, \tag{4.11}$$

which is adjoint to the infinitesimal generator $\mathscr{A}$ of the Frobenius-Perron semigroup. Let

$$w(t, \mathbf{x}) = \mathscr{U}(t)w_0(\mathbf{x}).$$

Then w satisfies

$$\frac{\partial}{\partial t}w(t, \mathbf{x}) = \mathscr{L}w(t, \mathbf{x}) \tag{4.12}$$

with initial condition $w(0, \mathbf{x}) = w_0(\mathbf{x})$.

4.1.2 Stochastic Differential Equations

We recall that any Stratonovich stochastic differential equation can be converted to its corresponding Itô stochastic differential equation. Hence, in this section we focus on the Itô stochastic differential equation (2.1) given by

$$d\mathbf{X}(t) = \mathbf{g}(t, \mathbf{X}(t))dt + \boldsymbol{\sigma}(t, \mathbf{X}(t))d\mathbf{W}(t), \quad \mathbf{X}(0) = \mathbf{X}_0.$$

By Remark 2.3 we know that if $\mathbf{g}$ and $\boldsymbol{\sigma}$ are assumed to satisfy conditions guaranteeing the existence of the fundamental solution to the Fokker-Planck equation (2.7), then the solution of (2.7) can be written as

$$p(t, \mathbf{x}) = \mathscr{P}(t)p_0(\mathbf{x}), \tag{4.13}$$

where the operator $\mathscr{P}(t)$ is defined by

$$\mathscr{P}(t)\psi(\mathbf{x}) = \int_{\mathbb{R}^n} \rho(t, \mathbf{x}; 0, \boldsymbol{\zeta})\psi(\boldsymbol{\zeta})d\boldsymbol{\zeta}, \quad t \geq 0.$$

We see that the operators $\mathscr{P}(t), t \geq 0$, are Markov operators. In addition, if $\mathbf{g}$ and $\boldsymbol{\sigma}$ are time-independent, then the operators $\{\mathscr{P}(t)\}_{t\geq 0}$ form a Markov semigroup with infinitesimal generator (e.g., see [52, Section 11.8])

$$\mathscr{A}\psi(\mathbf{x}) = -\sum_{k=1}^{n} \frac{\partial}{\partial x_k} \left(g_k(\mathbf{x})\psi(\mathbf{x})\right) + \frac{1}{2}\sum_{k,j=1}^{n} \frac{\partial^2}{\partial x_k \partial x_j} \left[\Sigma_{kj}(\mathbf{x})\psi(\mathbf{x})\right],$$

which is just Kolmogorov's forward operator (for the time-homogeneous case). Hence, under this case we can write the Fokker-Planck equation concisely as

$$\frac{\partial}{\partial t}p(t, \mathbf{x}) = \mathscr{A}p(t, \mathbf{x}).$$

Remark 4.9. Note that for the time-homogeneous case of (2.1) (i.e., $\mathbf{g}$ and $\boldsymbol{\sigma}$ are time-independent), the transition probability density function $\rho(t, \boldsymbol{\varsigma}; s, \mathbf{x})$ of the resulting diffusion process $\{\mathbf{X}(t) : t \geq 0\}$ depends only on the difference of the arguments $t - s$. Hence, in this case we denote the transition probability density function by $\tilde{\rho}(t - s, \boldsymbol{\varsigma}, \mathbf{x})$, that is,

$$\tilde{\rho}(t - s, \boldsymbol{\varsigma}, \mathbf{x}) = \rho(t, \boldsymbol{\varsigma}; s, \mathbf{x}).$$

We define a family of operators $\{\mathscr{T}(t)\}_{t\geq 0}$ by

$$\mathscr{T}(t)\phi(\mathbf{x}) = \mathbb{E}\left\{\phi(\mathbf{X}(t)) \mid \mathbf{X}(0) = \mathbf{x}\right\} = \int_{\mathbb{R}^n} \tilde{\rho}(t, \boldsymbol{\varsigma}, \mathbf{x})\phi(\boldsymbol{\varsigma})d\boldsymbol{\varsigma}, \quad t \geq 0. \tag{4.14}$$

Then $\{\mathscr{T}(t)\}_{t\geq 0}$ forms a semigroup with the infinitesimal generator $\mathscr{L}$ given by (e.g., see [59, Theorem7.3.3])

$$\mathscr{L}\phi(\mathbf{x}) = \sum_{k=1}^{n} g_k(\mathbf{x})\frac{\partial\phi(\mathbf{x})}{\partial x_k} + \frac{1}{2}\sum_{k,j=1}^{n} \Sigma_{kj}(\mathbf{x})\frac{\partial^2\phi(\mathbf{x})}{\partial x_k x_j}, \tag{4.15}$$

which is just Kolmogorov's backward operator (for the time homogeneous case). The operator $\mathscr{L}$ is often termed the infinitesimal generator of the (time-homogeneous) diffusion process $\{\mathbf{X}(t) : t \geq 0\}$. Let w_0 be a twice continuously differentiable function with compact support, and define

$$w(t, \mathbf{x}) = \mathscr{T}(t) w_0(\mathbf{x}).$$

Then w satisfies the following equation (e.g., see [59, Theorem 8.1.1])

$$\frac{\partial}{\partial t} w(t, \mathbf{x}) = \mathscr{L} w(t, \mathbf{x}) \tag{4.16}$$

with initial condition $w(0, \mathbf{x}) = w_0(\mathbf{x})$. Equation (4.16) is also referred to as Kolmogorov's backward equation (for the initial value problem) as it can be converted to the form with respect to the backward variable (i.e., the form similar to (2.3)) by using a transformation on the time variable (e.g., see [57, Chapter 4] for details).

4.2 Pointwise Equivalence Results Between Stochastic Differential Equations and Random Differential Equations

It was shown in [20] that there are classes of Itô stochastic differential equations for which their solutions have the same probability density functions at each time t as those for the solutions of their corresponding random differential equations. This equivalence between SDEs and RDEs was referred to as *pointwise equivalence* in [20]. The original motivation for such research was to provide an alternative way to numerically solve the Fokker-Planck equation since it is well-known that the Fokker-Planck equation is difficult to solve when the drift dominates the diffusion (the case of primary interest in many situations).

In establishing the equivalence results in [20], the following relationship between normal distribution and log-normal distribution [29, page 109] was heavily used.

Lemma 4.10. *If $\ln Z$ is normally distributed with mean μ and standard deviation σ, that is, $\ln Z \sim \mathcal{N}(\mu, \sigma^2)$, then Z is log-normally distributed, with its probability density function $f_Z(z)$ is defined by*

$$f_Z(z) = \frac{1}{z\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln z - \mu)^2}{2\sigma^2}\right),$$

and its mean and variance given by

$$\mathbb{E}(Z) = \exp(\mu + \tfrac{1}{2}\sigma^2), \;\; Var(Z) = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2).$$

In addition, the following basic result on the process generated by Itô integrals for Wiener processes is another main tool used in establishing the equivalence results in [20].

Lemma 4.11. *(See [45, Sec 4.3, Thm 4.11]) Let T be any positive constant. Then for a non-random function $f \in L^2(0,T)$, the Itô integrals $Q(t) = \int_0^t f(s)dW(s)$ for $0 < t \leq T$ yield a Gaussian stochastic process with pointwise distributions $\mathcal{N}\left(0, \int_0^t f^2(s)ds\right)$. Moreover, $Cov(Q(t), Q(t+\tau)) = \int_0^t f^2(s)ds$ for all $\tau \geq 0$.*

We next summarize the pointwise equivalence results from [20], where it was shown how to transform from a given Itô SDE to the corresponding equivalent RDE, and from a given RDE to the corresponding pointwise equivalent Itô SDE.

Linear scalar differential equation (Case 1): It was shown in [20] that with properly chosen coefficient functions and/or parameters the random differential equation

$$\frac{dx(t;\mathbf{Z})}{dt} = \alpha(t)x(t;\mathbf{Z}) + \gamma(t) + \mathbf{Z}\cdot\boldsymbol{\varrho}(t), \quad x(0;\mathbf{Z}) = X_0, \tag{4.17}$$

and stochastic differential equation

$$dX(t) = [\alpha(t)X(t) + \xi(t)]dt + \eta(t)dW(t), \quad X(0) = X_0, \tag{4.18}$$

are pointwise equivalent. Here $\mathbf{Z} = (Z_0, Z_1, \ldots, Z_{m-1})^T$ with $Z_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 0, 1, 2, \ldots, m-1$, mutually independent, α, γ and $\boldsymbol{\varrho} = (\varrho_0, \varrho_1, \ldots, \varrho_{m-1})^T$ are non-random functions of t, and μ_j, σ_j, $j = 0, 1, 2, \ldots, m-1$ are properly chosen positive constants, where γ, ϱ_j, μ_j and σ_j, $j = 0, 1, 2, \ldots, m-1$, are the functions and parameters needed to be determined when we want to find the corresponding RDE for the SDE (4.18). In addition, ξ and η in (4.18) are non-random functions of t, and they are the functions which can readily be determined when one wishes to find the corresponding SDE for the RDE (4.17). See [20] for details.

Linear scalar differential equation (Case 2): With properly chosen coefficient functions and/or parameters it was shown that the random differential equation

$$\frac{dx(t;\mathbf{Z})}{dt} = (\mathbf{Z}\cdot\boldsymbol{\varrho}(t) + \gamma(t))(x(t;\mathbf{Z}) + c), \quad x(0;\mathbf{Z}) = X_0 \tag{4.19}$$

and the stochastic differential equation

$$dX(t) = \xi(t)(X(t) + c)dt + \eta(t)(X(t) + c)dW(t), \quad X(0) = X_0, \tag{4.20}$$

are pointwise equivalent. Here $\mathbf{Z} = (Z_0, Z_1, \ldots, Z_{m-1})^T$ with $Z_j \sim \mathcal{N}(\mu_j, \sigma_j^2), j = 0, 1, 2, \ldots, m-1$, mutually independent, and c is a given constant. In addition, $\boldsymbol{\varrho} = (\varrho_0, \varrho_1, \ldots, \varrho_{m-1})^T$ is a non-random vector function of t, γ is a non-random function of t, μ_j and σ_j are positive constants which can be chosen when one wishes to determine the corresponding RDE for the SDE (4.20). Moreover ξ and η in (4.20) are non-random functions of t, which can be readily determined when one wants to find the corresponding SDE for the RDE (4.19). Again, precise details are given in [20].

Affine systems: It was also shown in [20] that with properly chosen coefficient functions and/or parameters the random differential equation

$$\frac{d\mathbf{x}(t;\mathbf{Z})}{dt} = \mathcal{A}(t)\mathbf{x}(t;\mathbf{Z}) + \boldsymbol{\gamma}(t) + \mathcal{H}(t)\mathbf{Z}, \quad \mathbf{x}(0;\mathbf{Z}) = \mathbf{X}_0, \tag{4.21}$$

and the stochastic differential equation

$$d\mathbf{X}(t) = [\mathcal{A}(t)\mathbf{X}(t) + \boldsymbol{\xi}(t)]dt + \mathcal{F}(t)d\mathbf{W}(t), \quad \mathbf{X}(0) = \mathbf{X}_0, \tag{4.22}$$

are pointwise equivalent. Here $\mathbf{Z}$ is a m-dimensional random vector that is multivariate normal distributed with mean vector $\boldsymbol{\mu}_{\mathbf{Z}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$, and $\mathcal{A}$ is a non-random $n \times n$ matrix function of t. In addition, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ are non-random n-dimensional vector functions of t, $\mathcal{H}$ is a non-random $n \times m$ matrix function of t, $\mathcal{F}$ is a non-random $n \times l$ matrix function of t, and $\mathbf{W}(t) = (W_1(t), W_2(t), \ldots, W_l(t))^T$ is an l-vector standard Wiener process. If one wants to find the corresponding RDE for the SDE (4.22), then $\boldsymbol{\gamma}$, $\mathcal{H}$, $\boldsymbol{\mu}_{\mathbf{Z}}$ and $\boldsymbol{\Sigma}_{\mathbf{Z}}$ are the functions and parameters needed to chosen. Conversely, if one wants to find the corresponding SDE for the RDE (4.21), then $\boldsymbol{\xi}$ and $\mathcal{F}$ are the functions that need to be determined. Details are given in [20].

Nonlinear scalar differential equations: We first observe that if a nonlinear stochastic differential equation (or a system of nonlinear stochastic differential equations) can be reduced to one of the forms in (4.18), (4.20) or (4.22) by some invertible transformation, then one can find its corresponding random differential equation. A similar statement holds if one begins with a nonlinear random differential equation that one can reduce to one of the forms (4.17), (4.19), or (4.21). Using special results found in [36, Sec. 4.1], the authors of [20] give conditions and invertible transformations for some nonlinear stochastic differential equations that can be reduced to the desired linear ones.

In addition, based on these pointwise equivalence results it was demonstrated in [14] that the solution for the GRDSS model can be used to approximate the solution of the Fokker-Planck equation (which is computational intensive when the drift dominates the diffusion). However, it should be noted that the *stochastic processes* obtained by these random differential equations and their corresponding stochastic differential equations are different as their covariance functions are in general *not the same* (see [20] for more details).

Based on the equivalent condition (2.20) between Itô SDEs and Stratonovich SDEs one can easily apply these pointwise equivalence results to obtain the mapping between the Stratonovich SDEs and their corresponding RDEs.

5 Concluding Remarks and Future Research Efforts

In this paper we consider uncertainty propagation in continuous-time dynamical systems through two types of differential equations. One is stochastic differential equations (SDEs),

and the other is random differential equations (RDEs). This classification is based on whether the random input is correlated or not. We focused on the time evolution of the probability density function of the stochastic process resulting from these two types of differential equations, and discussed numerous applications of the resulting equations in different fields. Even though SDEs and RDEs are different in nature, we have found rather close relationships between SDEs and RDEs. Specifically, the probability density functions of the solutions to SDEs and some RDEs can be associated with Markov operators. In addition, there are classes of SDEs for which their solutions have the same probability density function at each time t as those for the solution of their corresponding RDEs.

Based on the discussions in Sections 2 and 3 we see that all the structured population models presented in this paper can be associated with stochastic processes, which can be obtained due to the variability/stochasticity in the growth rate and/or variability in the initial size. In reality, mortality and birth rates may also be random (characterized by finite dimensional random variables or colored noise or white noise). Thus, it would be interesting to investigate how to correctly incorporate these different types of randomness into the existing dynamical system (e.g., GRDSS or FP), and investigate the relationship between them. In addition, it would be interesting to further explore the stochastic processes with which those nonlinear structured population models (e.g., a size structured population model with growth rate depending on the total number of population) are associated.

The pointwise equivalence results presented in Section 4.2 were established without mortality and birth rates included in the models. It would also be interesting to investigate the equivalence between GRDSS and FP when these rates are present and have a significant effect on the dynamics (as is often the case). As we see from Section 2.1.2, equation (2.17) with mortality involved can be associated with a killed diffusion process. Hence, it is expected that we can also obtain pointwise equivalence results for the population density when the mortality rate is present.

Acknowledgements

This research was supported in part by grant number NIAID R01AI071915-09 from the National Institute of Allergy and Infectious Disease, and in part by the U.S. Air Force Office of Scientific Research under grant number FA9550-12-1-0188.

References

[1] A.S. Ackleh and K. Deng, A monotone approximation for the nonautonomous size-structured population model, *Quart. Appl. Math.*, **57** (1999), 261–267.

[2] B. M. Adams, H.T. Banks, M. Davidian, and E. S. Rosenberg, Model fitting and prediction with HIV treatment interruption data, CRSC-TR05-40, October, 2005; *Bulletin of Math. Biology*, **69** (2007), 563–584.

[3] O. Angulo and J.C. López-Marcos, Numerical integration of fully nonlinear size-structured population models, *Applied Numerical Mathematics*, **50** (2004), 291–327.

[4] H.T. Banks, *A Functional Analysis Framework for Modeling, Estimation and Control in Science and Engineering,* CRC Press/Taylor and Frances Publishing, June, 2012.

[5] H.T. Banks and K.L. Bihari, Modelling and estimating uncertainty in parameter estimation, *Inverse Problems*, **17** (2001), 95–111.

[6] H.T. Banks, V.A. Bokil, S. Hu, A.K. Dhar, R.A. Bullis, C.L. Browdy and F.C.T. Allnutt, Modeling shrimp biomass and viral infection for production of biological countermeasures, CRSC-TR05-45, NCSU, December, 2005; *Mathematical Biosciences and Engineering*, **3** (2006), 635–660.

[7] H.T. Banks, L.W. Botsford, F. Kappel and C. Wang, Modeling and estimation in size structured population models, LCDS-CCS Report 87-13, Brown University; Proceedings 2nd Course on Mathematical Ecology, (Trieste, December 8-12, 1986) World Press, Singapore, 1988, 521–541.

[8] H.T. Banks, M.W. Buksas and T. Lin, *Electromagnetic Material Interrogation Using Conductive Interfaces and Acoustic Wavefronts*, Frontiers in Applied Mathematics **21**, SIAM Publ., August, 2000.

[9] H.T. Banks, M. Davidian, S. Hu, G.M. Kepler and E.S. Rosenberg, Modeling HIV immune response and validation with clinical data, *Journal of Biological Dynamics*, **2** (2008), 357–385.

[10] H.T. Banks and J.L. Davis, A comparison of approximation methods for the estimation of probability distributions on parameters, CRSC-TR05-38, October, 2005; *Applied Numerical Mathematics*, **57** (2007), 753–777.

[11] H.T. Banks and J.L. Davis, Quantifying uncertainty in the estimation of probability distributions, CRSC-TR07-21, December, 2007; *Math. Biosci. Engr.*, **5** (2008), 647–667.

[12] H.T. Banks, J.L. Davis, S.L. Ernstberger, S. Hu, E. Artimovich, A.K. Dhar and C.L. Browdy, A comparison of probabilistic and stochastic formulations in modeling growth uncertainty and variability, CRSC-TR08-03, NCSU, February, 2008; *Journal of Biological Dynamics*, **3** (2009), 130–148.

[13] H.T. Banks, J.L. Davis, S.L. Ernstberger, S. Hu, E. Artimovich and A.K. Dhar, Experimental design and estimation of growth rate distributions in size-structured shrimp populations, CRSC-TR08-20, NCSU, November, 2008; *Inverse Problems*, **25** (2009), 095003(28pp).

[14] H.T. Banks, J.L. Davis and S. Hu, A computational comparison of alternatives to including uncertainty in structured population models, CRSC-TR09-14, June, 2009; in *Three Decades of Progress in Systems and Control*, X. Hu, U. Jonsson, B. Wahlberg, B. Ghosh (Eds.), Springer, 2010, 19–33.

[15] H.T. Banks, S.L. Ernstberger and S. Hu, Sensitivity equations for a size-structured population model, *Quarterly of Applied Mathematics*, **67** (2009) 627–660.

[16] H.T. Banks and B.G. Fitzpatrick, Estimation of growth rate distributions in size structured population models, *Quarterly of Applied Mathematics*, **49** (1991), 215–235.

[17] H.T. Banks, B.G. Fitzpatrick, L.K. Potter and Y. Zhang, Estimation of probability distributions for individual parameters using aggregate population data, CRSC-TR98-6, NCSU, January, 1998; in *Stochastic Analysis, Control, Optimization and Applications*, (Edited by W. McEneaney, G. Yin and Q. Zhang), Birkhauser, Boston, 1989, 353–371.

[18] H.T. Banks and N.L. Gibson, Inverse problems involving Maxwell's equations with a distribution of dielectric parameters, CRSC-TR05-29, Center for Research in Scientific Computation, North Carolina State University, July, 2005; *Quarterly of Applied Mathematics*, **64** (2006), 749–795.

[19] H.T. Banks, S.L. Grove, S. Hu and Y. Ma, A hierarchical Bayesian approach for parameter estimation in HIV models, *Inverse Problems*, **21** (2005) 1803–1822.

[20] H.T. Banks and S. Hu, Nonlinear stochastic Markov processes and modeling uncertainty in populations, CRSC-TR11-02, NCSU, January, 2011; *Mathematical Bioscience and Engineering*, **9** (2012), 1–25.

[21] H.T. Banks, S. Hu and Z.R. Kenz, A brief review of elasticity and viscoelasticity for solids, *Advances in Applied Mathematics and Mechanics*, **3** (2011), 1–51.

[22] H.T. Banks and F. Kappel, Transformation semigroups and L^1-approximation for size-structured population models, *Semigroup Forum*, **38** (1989), 141–155.

[23] H.T. Banks, F. Kappel and C. Wang, Weak solutions and differentiability for size-structured population models, *International Series of Numerical Mathematics*, **100** (1991), 35–50.

[24] H.T. Banks, K.L. Sutton, W.C. Thompson, G. Bocharov, D. Roose, T. Schenkel and A. Meyerhans, Estimation of cell proliferation dynamics using CFSE data, Center for Research in Scientific Computation Technical Report CRSC-TR09-17, NCSU, August, 2009; *Bull. Math. Biol.*, **73** (2011), 116–150.

[25] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton, FL, July, 2008, 308pp., published, January 2, 2009.

[26] H.T. Banks, H.T. Tran and D.E. Woodward, Estimation of variable coefficients in the Fokker-Planck equations using moving node finite elements, *SIAM J. Numer. Anal.*, **30** (1993), 1574–1602.

[27] G.I. Bell and E.C. Anderson, Cell growth and division I. a mathematical model with applications to cell volume distributions in mammalian suspension cultures, *Biophysical Journal*, **7** (1967), 329–351.

[28] A. Calsina and J. Saldana, A model of physiologically structured population dynamics with a nonlinear individual growth rate, *Journal of Mathematical Biology*, **33** (1995), 335–364.

[29] G. Casella and R.L. Berger, *Statistical Inference*, Duxbury, California, 2002.

[30] F.L. Castille, T.M. Samocha, A.L. Lawrence, H. He, P. Frelier and F. Jaenike, Variability in growth and survival of early postlarval shrimp (*Penaeus vannamei Boone 1931*) *Aquaculture*, **113** (1993), 65–81.

[31] J. Chen and J. Li, A note on the principle of preservation of probability and probability density evolution equation, *Probabilistic Engineering Mechanics*, **24** (2009), 51–59.

[32] J.M. Cushing, *An Introduction to Structured Population Dynamics*, in CMB-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA, 1998.

[33] J. Ding, T.Y. Li and A. Zhou, Finite approximations of Markov operators, *Journal of Computational and Applied Mathematics*, **147** (2002), 137–152.

[34] B.G. Dostupov and V.S. Pugachev, The equation for the integral of a system of ordinary differential equations containing random parameters, *Automation and Remote Control*, **18** (1957), 620–630.

[35] D. Ellis, About colored noise, `http://www.ee.columbia.edu/~dpwe/noise/`.

[36] T.C. Gard, *Introduction to Stochastic Differential Equations*, Marcel Dekker, New York, 1988.

[37] C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer-Verlag, Berlin, 1983.

[38] R.J. Griego and R. Hersh, Random evolutions, Markov chains, and systems of partial differential equations, *Proc. of National Academy of Sciences*, **62** (1969), 305–308.

[39] M. Grigoriu, *Stochastic calculus: applications in science and engineering*, Birkhauser, Boston, 2002.

[40] P. Hänggi, Colored noise in continuous dynamical systems: a functional calculus approach, in *Noise in Nonlinear Dynamical Systems*, vol. 1 (Edited by F. Moss and P. V. E. McClintock), Cambridge University Press, 1989, 307–328.

[41] P. Hänggi and P. Jung, Colored noise in dynamical system, in *Advances in Chemical Physics*, Volume LXXXIX, edited by I. Prigogine and S.A. Rice, John Wiley & Sons, Inc., 1995.

[42] W. Horsthemke and R. Lefever, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*, Springer-Verlag, New York, 1984.

[43] R. Hersh, Random evolutions: a survey of results and problems, *Rocky Mountain Journal of Mathematics*, **4** (1974), 443–477.

[44] N. Hritonenko and Y. Yatsenko, Optimization of harvesting age in an integral age-dependent model of population dynamics, *Mathematical Biosciences*, **195** (2005), 154–167.

[45] F. Klebaner, *Introduction to Stochastic Calculus with Applications*, 2nd ed., Imperial College Press, London, 2006.

[46] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, Second Edition, Springer, New York, 1991.

[47] B. Keyfitz and N. Keyfitz, The McKendrick partial differential equation and its uses in epidemiology and population study, *Math. Comput. Modell.*, **25** (1997), 1–9.

[48] V. Kopustinskas, J. Augutis, S. Rimkevicius, Dynamic reliability and risk assessment of the accident localization system of the Ignalina NPP RBMK-1500 reactor, *Reliability Engineering and System Safety*, **87** (2005), 77–87.

[49] V.S. Koroliuk and N. Limnios, Random evolutions toward applications, in *Encyclopedia of Statistics in Quality and Reliability*, Wiley 2008.

[50] V. Korolyuk and A. Swishchuk, *Evolution of System in Random Media*, CRC Press, Boca Ralton, Florida, 1995.

[51] F. Kozin, On the probability densities of the output of some random systems, *Trans. ASME Ser. E J. Appl. Mech.*, **28** (1961), 161–164.

[52] A. Lasota and M.C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, Applied Mathematical Science, **97**, Springer, New York, 1994.

[53] X. Mao and C. Yuan, *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, London, 2006.

[54] A.G. McKendrick, Applications of mathematics to medical problems, *Proceedings of the Edinburgh Mathematical Society*, **44** (1926), 98–130.

[55] J.A.J. Metz and E.O. Diekmann, *The Dynamics of Physiologically Structured Populations*, Lecture Notes in Biomathematics, Vol. 68, Springer, Heidelberg, 1986.

[56] J.E. Moyal, Stochastic processes and statistical physics, *Journal of the Royal Statistical Society. Series B (Methodological)*, **11** (1949), 150-210.

[57] J. Nolen, *Partial Differential Equations and Diffusion processes*, `http://math.stanford.edu`.

[58] W.L. Oberkampf, S.M. DeLand, B.M. Rutherford, K.V. Diegert and K.F. Alvin, Error and uncertainty in modeling and simulation, *Reliability Engineering and Systems Safety*, **75** (2002), 333–357.

[59] B. Oksendal, *Stochastic Differentail Equations*, 5th edition, Springer, Berlin, 2000.

[60] A. Okubo, *Diffusion and Ecological Problems: Mathematical Models*, Biomathematics, 10 (1980), Springer-Verlag, Berlin.

[61] G. Oster and Y. Takahashi, Models for age-specific interactions in a periodic environment, *Ecological Monographs*, **44** (1974), 483–501.

[62] A. Pirrotta, Multiplicative cases from additive cases: extension of Kolmogorov-Feller equation to parametric Poisson white noise processes, *Probabilistic Engineering Mechanics*, **22** (2007), 127–135.

[63] Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory, *Theor. Prob. Appl.*, **1** (1956), 157–214.

[64] L. Ridolfi, P. D'Odorico and F. Laio, *Noise-Induced Phenomena in the Environmental Sciences*, Cambridge University Press, 2011.

[65] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer, New York, 1996.

[66] E. S. Rosenberg, M. Davidian and H.T. Banks, Development of structured treatment interruption strategies for HIV infection, CRSC-TR06-12, April, 2006; *Drug and Alcohol Dependence*, **88S** (2007), S41–S51.

[67] R. Rudnicki, Models of population dynamics and genetics, in *From Genetics To Mathematics*, (edited by M. Lachowicz and J. Miekisz), World Scientific, Singapore, 2009, 103–148.

[68] R. Rudnicki, K. Pichór and M. Tyran-Kamińska, Markov semigroups and their applications, in *Dynamics of Dissipation*, P. Garbaczewski and R. Olkiewics (eds.), Lecture Notes in Physics, **597** (2002), Springer, Berlin, 215–238.

[69] T.L. Saaty, *Modern Nonlinear Equations*, Dover, Mineola, N.Y., 1981.

[70] Z. Schuss, *Theory and Applications of Stochastic Processes: An Analytical Approach*, Springer, New York, 2010.

[71] J. Sinko and W. Streifer, A new model for age-size structure of a population, *Ecology*, **48** (1967), 910–918.

[72] T.T. Soong, *Random Differential Equations in Science and Engineering*, Academic Press, New York and London, 1973.

[73] T.T. Soong and S.N. Chuang, Solutions of a class of random differential equations, *SIAM J. Appl. Math.*, **24** (1973), 449–459.

[74] M. Stoyanov, M. Gunzburger, and John Burkardt, Pink noise, $1/f^{\alpha}$ noise, and their effect on solutions of differential equations, *International Journal for Uncertainty Quantification*, **1** (2011), 257–278.

[75] R.L. Stratonovich, A new representation for stochastic integrals and equations, *J. SIAM Control*, **4** (1966), 362–371.

[76] H.J. Sussmann, On the gap between deterministic and stochastic ordinary differential equations, *The Annals of Prob.*, **6** (1978), 19–41.

[77] A. Swishchuk, *Random Evolutions and Their Applications: New Trends*, Kluwer Academic Pulishers, Netherlands, 2005.

[78] G.K. Vallis, Mechanisms of climate variability from years to decades, in *Stochastic Physics and Climate Modelling*, (edited by T. Palmer and P. Williams), Cambridge University Press, Cambridge, 2010, 1–34.

[79] H. Von Foerster, Some remarks on changing populations, in *The kinetics of cellular proliferation*, F. Stohlman, Jr. (ed.), Grune and Stratton, New York, 1959.

[80] G.F. Webb, *Theory of Nonlinear Age-dependent Population Dynamics*, Marcel Dekker, New York, 1985.

[81] G.H. Weiss, Equation for the age structure of growing populations, *Bull. Math. Biophy.*, **30** (1968), 427-435.

[82] E. Wong and M. Zakai, On the relation between ordinary and stochastic differential equations, *Int. J. Engng Sci.*, **3** (1965), 213–229.

[83] E. Wong and M. Zakai, Riemann-Stieltjes approximations of stochastic integrals, *Z. Wahr. verw. Geb.*, **12** (1969), 87–97.

[84] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton and Oxford, 2010.

[85] G. Yin, X. Mao, C. Yuan and D. Cao, Approximation methods for hybrid diffusion systems with state-dependent switching processes: numerical algorithms and existence and uniqueness of solutions, *SIAM J. Math. Anal.*, **41** (2010), 2335–2352.

[86] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*, Springer, New York, 2009.