The Journal of Marine Research, one of the oldest journals in American marine science, published important peer-reviewed original research on a broad array of topics in physical, biological, and chemical oceanography vital to the academic oceanographic community in the long and rich tradition of the Sears Foundation for Marine Research at Yale University.

An archive of all issues from 1937 to 2021 (Volume 1–79) are available through EliScholar, a digital platform for scholarly publishing provided by Yale University Library at https://elischolar.library.yale.edu/.

Requests for permission to clear rights for use of this content should be directed to the authors, their estates, or other representatives. The Journal of Marine Research has no contact information beyond the affiliations listed in the published articles. We ask that you provide attribution to the Journal of Marine Research.

Yale University provides access to these materials for educational and research purposes only. Copyright or other proprietary rights to content contained in this document may be held by individuals or entities other than, or in addition to, Yale University. You are solely responsible for determining the ownership of the copyright, and for obtaining permission for your intended use. Yale University makes no warranty that your distribution, reproduction, or other use of these materials will not infringe the rights of third parties.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. https://creativecommons.org/licenses/by-nc-sa/4.0/
ABSTRACT

The general theory of stochastic differential equations is presented in this chapter, including the theoretical background on how measured statistics from time series can be used to develop a stochastic parameterization. The general rules of stochastic calculus, including the important and often overlooked differences between Ito and Stratonovich calculus, are mentioned, and references are provided in which more detail may be found. We discuss how Stratonovich calculus is usually appropriate for fluid systems, whereas Ito calculus is often appropriate for data assimilation. We also discuss some common numerical pitfalls awaiting the unwary modeler, and warn against unsophisticated random number generators. Finally, we offer a selection of examples showing the importance of the variability of unresolved scales in an ocean model and, by citation, a variety of methods that have been employed.

Keywords. Stochastic modeling, Parameterization, Multiscale interactions, Air-sea interaction, Numerical modeling, Ensemble forecasting, Central Limit Theorem, Ocean Dynamics, Ocean Predictions, Predictability

1. Introduction

The improved prediction skill at long leads of ocean and atmospheric numerical models that comes with higher spatial and temporal resolution indicates that the importance of multiscale interactions cannot be ignored. Because skill is not yet perfect even when the dynamical interactions are well understood, we are faced with the choice of waiting for sufficient resolution to evaluate unresolved processes explicitly, or to parameterize them. Without knowing what constitutes “sufficient resolution,” parameterization presents a viable and relatively efficient course of action.

Deterministic parameterizations are a subset of stochastic parameterizations, and are valid when their variability is small enough that it does not significantly interact with the resolved dynamical scales over the course of a time step or spatial grid size. That is, the “method of averaging,” e.g., Arnold (2001), is valid. The method of averaging has been the traditional approach to parameterization in both oceanic and atmospheric modeling; a deterministic
forcing or correction to a parameter is applied to represent the average effects of subscale variability on the numerical integration over a time step. For example, viscosity in a liquid is a parameterization of average energy and momentum transfer between molecules due to intermolecular forces (Batchelor 1967).

The next step in complexity, and one that is increasingly popular, is the “stochastic parameterization.” Technically, this term is broader than what is usually meant when used by ocean modelers. After all, as stated above, deterministic systems are a special case, comprising those systems whose phase space distribution function (a.k.a., probability distribution function [PDF]) is a delta function having the equations of motion as its argument (Gibbs 1902). However, in what follows, we shall use this term to denote expressions involving nondeterministic random variables to represent unresolved processes. In particular, we shall concentrate on parameterizations involving stochastic differential equations (SDEs), which arise in ocean modeling when equations of motion contain terms with widely different dynamical temporal, or spatial scales. For ease of exposition, we will often describe systems with timescale separations; however, the ideas can also be applied to systems with space-scale separations. In any case, temporal and spatial scales are often related; adding temporal stochasticity can also improve spatial variability in a model (Berner et al. 2009; Berner, Jung, and Palmer 2012), and it is important to avoid “double counting,” i.e., introducing both spatial and temporal parameterizations when the phenomena have a single common cause.

The purpose of this chapter is to present a consistent development of data-based stochastic parameterizations. That is, one injects the statistics of high-resolution observations of rapidly-varying processes into expressions resulting from rigorously-derived limit theorems to represent these processes in a lower-resolution numerical model with a satisfactory degree of fidelity (cf. equation 6). There already is, of course, a large body of literature on stochastic parameterizations based on a variety of modeling considerations (e.g., Jansen and Held 2014 and references therein; see also the collection of articles in Adler, Müller, and Rozovskii 2012), and it is impossible to do justice to the entire discipline. We can only point the readers toward additional reading through which they may make their own journeys, while concentrating on one path that we believe to be relevant and useful to those wishing to parameterize subscale processes in ocean models.

This chapter is organized as follows: Hasselmann’s landmark paper (Hasselmann 1976) and the subsequent studies it influenced, which suggested treating the ocean as a stochastic system forced by a random atmosphere, serve as a non rigorous but enlightening exposition of the ideas involved. Thus, they are the subjects of Section 2. In Section 3, we introduce SDEs, discussing the central limit theorem in its dynamical context and presenting a procedure by which measured spectra of parameterized processes are used to estimate the amplitude of stochastic terms in a numerical model. Some issues that are introduced by the stochastic terms are discussed in Section 4. Section 5 discusses legitimate numerical methods for integrating SDEs, concentrating on the role of the time step and providing cautionary tales. Section 6 provides examples of the development of stochastic parameterization
Penland and Navarra: Issues in stochastic ocean modeling

in ocean models, and then the conclusions close the chapter. Because rigorous derivations of our subjects are published elsewhere, we cite them but do not duplicate them here. Rather, we emphasize concepts and methods of implementation.

2. Hasselmann’s vision

Hasselmann (1976), noting the timescale separation between atmospheric “weather” and oceanic-cryospheric-terrestrial “climate,” suggested a stochastic model of climate variability whereby the slow climate system was forced by rapidly-varying, essentially random weather noise. Denoting the climate as a vector $y$ and the weather as a vector $x$, Hasselmann postulated that, without dissipative feedback, $y$ might behave as Brownian motion if $x$ could be approximated as white noise. Hasselmann was aware that this approximation is valid as long as the spectrum of $x$ is flat in the range of angular frequencies $\omega$ to which $y$ is dynamically sensitive, and that this range is generally concentrated in the low frequencies (e.g., Taylor 1922). Figure 1, taken from Hasselmann (1976), describes these ideas schematically. In Figure 1, primes denote deviation from the average, $\tau_x$ is a characteristic timescale of the fast process $x$, $F_{ij}(\omega)$ is a component of the cross-spectrum of the fast dynamics for variables $i$ and $j$, and $G_{ij}(\omega)$ is the corresponding quantity for the slow response. The $\omega^{-2}$ dependence by $G_{ij}(\omega)$ in Figure 1b arises from Brownian motion-like behavior of the slow variables. As discussed in Frankignoul and Hasselmann (1977), the presence of a negative linear feedback in the equations for $y$ would flatten $G_{ij}(\omega)$ at low frequencies and retain the $G_{ij}(\omega)$ dependence at higher frequencies.

Applications of Hasselmann’s vision appeared the next year, with applications to a simple ocean model (Frankignoul and Hasselmann 1977) and to zonally-averaged energy balance models (Lemke 1977) of the Budyko–Sellers type (Budyko 1969; Sellers 1969). In both cases, the authors found reasonable quantitative agreement with observations and suggested that rapidly-varying forcing provided a plausible source of low-frequency variability. Frankignoul and Hasselmann (1977) also considered the spectral properties of candidate oceanic forcing variables such as latent and sensible heat flux, and showed that these spectra were essentially white at periods of weeks to years.

Modern oceanic general circulation models (GCMs), of course, have no trouble resolving explicitly the timescales considered in these 1977 articles. The beauty of Hasselmann’s vision, as regards stochastic parameterization, is that it is applicable to timescales much shorter than those he originally discussed. On purpose, there are no units on the graphs in Figure 1. Rapidly varying processes acting as stochastic forcing intrinsically excite slow climate variability. Thus, turning to the particular case of stochastic parameterizations, we might legitimately interpret $y$ as resolved dynamics in a GCM and $x$ as the unresolved dynamics. As long as $\tau_x$ is small enough compared with the time step $\Delta$ that $\frac{1}{\Delta}$ is in the flat part of the spectrum (cf. Fig. 1b), the dynamics of $x$ may be approximated as white noise. Even this condition is not strictly necessary; it is often preferable to simulate systems that are not really resolved by the model, but whose correlation times are of the same order of
The Sea: The Science of Ocean Prediction

Figure 1. Input and response functions of a stochastically forced climate model without feedback; (a) covariances, and (b) spectra, from Hasselmann (1976). Primes denote deviation from the average, $\tau_x$ is a characteristic timescale of the fast process $x$, $F_{ij}(\omega)$ is a component of the cross-spectrum of the fast dynamics, and $G_{ij}(\omega)$ is a component of the cross-spectrum of the slow response.

magnitude as the time step. In this case, one might parameterize $x$ as red noise. Because red noise is filtered white noise, this procedure technically increases the dimensionality of the system. Finally, Hasselmann (1976) article gave no practical guidance for treating multiplicative stochastic forcing. We address this issue below.

3. Technical development

In this section, we discuss the central limit theorem in its dynamical context and present a procedure by which measured spectra of parameterized processes are used to estimate the amplitude of stochastic terms in a numerical model.

The reader will note the linear increase with time by the covariance of the slow process in Section 1a. This is a property of a Wiener process, or Brownian motion, which is the integral of white noise with no feedback onto the slow processes. In GCMs, the feedback is large and complex, but the Wiener process is still important to the theory of stochastic parameterizations Wong and Zakai (1965). Before discussing why the Wiener process is
so important, it is necessary to know what it is. We therefore begin with a review of its definition and properties. For additional information, we refer the reader to Arnold (1974).

a. Gaussian white noise and the Wiener process

White noise, by definition, has equal amplitude at all frequencies and, thus, its variance would diverge if integrated over frequency. Although true white noise does not exist in nature, natural processes often behave as though they were generated by white noise. How, then, can one have a mathematically consistent description of these systems while maintaining physical consistency? The answer, nonrigorously speaking, is to consider a product of white noise with something of measure zero so that the size of that product is finite. The Wiener process $W(t)$ is therefore defined through its (generalized) derivative:

$$dW(t) = \xi dt$$

where $\xi$ is Gaussian white noise and $dt$ is the time differential. That is, $\xi$ has an autocovariance structure

$$\langle \xi(t + \tau)\xi(t) \rangle = \delta(\tau).$$

In this exposition, angle brackets denote ensemble average. The delta function in equation (2) has units of $(time)^{-1}$ and infinite amplitude. The Wiener process is a Gaussian random variable with the following properties:

$$W(0) = 0$$

$$\langle W(t) \rangle = \langle dW(t) \rangle = 0$$

$$\langle dW(t)dW(t') \rangle = 0 \quad (t \neq t')$$

$$\langle dW(t)^2 \rangle = dt$$

$$\langle W(t)W(t') \rangle = \min(t, t')$$

Please note that the variance of $dW(t)$ equals $dt$. The practical effect of this is that stochastic terms in a numerical model are updated using the square root of the time step. Numerically-generated examples of white noise and the Wiener process, as well as their spectra, are shown in Figure 2. In generating Figure 2, a time step of $10^{-4}$ temporal units (tu) was used and the time series was sampled every $10^{-2}$ tu. The variance of $dW = \xi dt$ was estimated as $9.6 \times 10^{-5}$ tu, to be compared with $10^{-4}$ tu. Spectra were normalized to the estimated variance of the time series and smoothed using a Bartlett filter with ten ensemble members. As noted by Hasselmann (1976), the spectrum of the Wiener process varies as $f^{-2}$; the singularity at zero frequency is consistent with the nonstationary statistics of $W(t)$.

It may be counterintuitive that stochastic parameterizations in a numerical model require more care than deterministic parameterizations, but such is the case. One cannot merely
“shake up” a model without the possibility of introducing unphysical effects. In the first place, the Wiener process is so variable that the rules of calculus are no longer unique when these processes are integrated. This is true for theoretical Wiener processes, because the evolution of a process found in nature may be described by one of two self-consistent but very different calculi. It is even more important for numerical systems, because naive implementation of stochastic numerical schemes may result in an unphysical system (G. Gottwald, personal communication). It is up to the modeler to ensure that the stochastic convergence properties are known to be appropriate. Avoiding Wiener processes and using “simpler” random variables simply compounds the problem. In practice, although care is necessary, it is not generally onerous (see Section 4). Extensive discussions of multiple calculi may be found in Kloeden and Platen (1992); simpler discussions are available in
Gardiner (1985). The two calculi found in nature are “Ito” and “Stratonovich” calculi. The Stratonovich calculus reproduces the deterministic rules of calculus and is appropriate for continuous systems, although they may be discretely sampled. Fluid systems generally obey Stratonovich calculus. The Ito calculus is appropriate for inherently discrete systems that are approximated as continuous. Financial models, describing individual transactions but on a continuous basis, often use Ito calculus. Because the subjects of stochastic parameterizations in ocean models (e.g., mesoscale eddies, internal gravity waves, variations in the equation of state) are usually continuous, we shall concentrate on Stratonovich calculus. Nevertheless, it is instructive to compare the results of a very simple calculus problem: the integral of a Wiener process with respect to itself.

We begin by writing the integral as a Riemann sum over $N$ finite intervals of width $\Delta t$, with $t = N\Delta t$:

$$\int_0^t W(t')dW(t') = \sum_{i=0}^{N-1} W(t_i + \alpha \Delta t)[W(t_{i+1}) - W(t_i)]$$  \hspace{1cm} (3)

In equation (3), $\alpha$ is a number in the interval $[0, 1]$. As shown in, for example, Horsthemke and Lefever (1984), the quadratic-mean limit of the sum for large $N$ and small $\Delta t$ is

$$\frac{1}{2}W^2(t) + \left(\alpha - \frac{1}{2}\right)t$$  \hspace{1cm} (4)

because $W(0) = 0$. In deterministic Riemannian calculus, the solution to equation (3) would be simply $W^2(t)/2$, independent of $\alpha$. In stochastic calculus, however, the value of $\alpha$ is crucial. The rules of deterministic calculus are reproduced for $\alpha = 1/2$, and this corresponds to Stratonovich calculus. Ito calculus results when $\alpha = 0$.

To illustrate this numerically, we generated a Wiener process (Fig. 2) for a total of 10 $tu$, but sampling every time step. The final value of $W(10tu)$ was 1.13. We divided this time series into equal intervals of 0.0002 $tu$ and evaluated the sum in equation (3) with $\alpha = 1/2$ (Stratonovich) and $\alpha = 0$ (Ito). The sum with $\alpha = 1/2$ was 0.62, compared with a theoretical value of 0.64, and the sum with $\alpha = 0$ was $-4.32$, compared with $-4.36$. Of course, numerical models do not use a Riemann sum as their numerical integration scheme, and we defer further discussion of stochastic numerical integration until Section 5.

b. The Wiener process and physical multiscale interactions

Consider a multivariate dynamical equation with different timescales, where $x$ is a vector in a suitable vector space and $F(x, t)$ is vector valued functions in the same space, whereas for simplicity we consider $G(x, t)$ as a scalar function,

$$\frac{dx}{dt} = F(x, t) + G(x, t)\eta(x, t)$$  \hspace{1cm} (5)

In equation (5), the time and space scales of $\nu F$ and $G$ are resolved in a numerical model we wish to use; the scales of $\eta$ are not resolved, but $\eta$ is too important to be ignored. Note
that both $F(x,t)$ and the product $G(x,t)\eta$ in equation (5) have units of $x/t$. In practice, $\eta$ represents the combined effects of many rapid, weakly-correlated effects on $x$; so, according to the central limit theorem (e.g., Bhattacharya and Waymire 1990), we expect it to be a Gaussian random vector. Although equation (5) is already a simplified version of the general multiscale problem, let us simplify it still further by ignoring explicit dependence on $x$; a more general treatment is derived in Papanicolaou and Kohler (1974). We will assume that all of the conditions required by their theorem apply here.

The stochastic parameterization of $\eta$ involves more than replacing it with a Gaussian random variable with variance equal to the variance of $\eta$. In fact, a dynamically consistent parameterization of $\eta$ involves the entire covariance structure of $\eta$, be it ever so unresolved. The procedure requires estimation of the matrix

$$C_{ij} = \int_{-\infty}^{\infty} \langle \eta_i(t+t')\eta_j(t) \rangle dt'$$

Because $C$ is symmetric it can be written as a product of matrices $\Phi\Phi^T$. The dynamically consistent stochastic parameterization of $\eta$ is as follows:

$$dx = F(x,t)dt + \frac{1}{2}G(x,t)\Phi \circ dW(t)$$

where $W(t)$ is a vector of independent Wiener processes, and where the symbol “$\circ$” denotes that the Wiener process is to be integrated in the sense of Stratonovich. If Ito’s circle is missing in an SDE, standard mathematical formalism would indicate that the equation is to be integrated in the sense of Ito. The point to be emphasized here is that if rapidly-varying continuous effects are to be modeled using white noise, the equations must be integrated in the sense of Stratonovich (Wong and Zakai 1965; Khaszminkii 1966; Papanicolaou and Kohler 1974). Integrating in the sense of Stratonovich ensures inclusion of the “noise-induced drift” (discussed in the next section).

Examining the units of quantities in equations (6) and (7), it is clear that $\Phi$ does not have units of $\eta$, but rather has units of $\eta\sqrt{t}$. Thus, in the numerical generation of equation (7), the deterministic part of the equation is updated with the time step, and the stochastic part has to be updated with something proportional to the square root of the time step, i.e., the Wiener process. Note that $\Phi$ is not unique, as it may be replaced by $\Phi S$ where $S$ is any orthogonal matrix. Thus, what we present is only guaranteed to give an equation that faithfully produces the moments of $x$. However, it does give all of them, including the lagged moments, so this is not usually a serious issue.

The matrix $C$ may be a function of time, and so this dynamical form of the central limit theorem does not require $\eta$ to have stationary statistics. However, if $C$ does have stationary statistics, the interpretation of $C$ is particularly simple; the Wiener–Khintchine
Theorem (Wiener 1930; Khintchine 1934) interprets equation (6) as the cross-spectrum of \( \eta \) evaluated at zero frequency. Again, we refer to Figure 1b; the stochastic approximation is valid for time steps corresponding to the flat, low-frequency part of the spectrum.

In practice, the low-frequency part of a spectrum derived from time series can be highly uncertain. Thus, if a stochastic parameterization is to be based on time series of measured quantities, it may be desirable to estimate the lagged covariance function and integrate that over lag for comparison with the low-frequency intercept of the spectrum.

It must be emphasized that equation (8) is to be integrated with a numerical scheme yielding Stratonovich statistics. An Euler scheme, for example, cannot generate Stratonovich calculus, and a Heun scheme cannot generate Ito calculus (Rüemelin 1982; Kloeden and Platen 1992). We underscore the importance of this issue by considering the probability distribution function (PDF) of the stochastic system, equation (7), and what it says about the moments of \( x \).

c. Noise-induced drift

Intuitively, because \( \langle dW(t) \rangle = 0 \), one might think from equation (7) that \( \langle x \rangle \) would be evaluated from the zeroes of \( \langle F(x,t) \rangle \). Although this is true for Ito systems, it is not true for Stratonovich systems if either \( G \) or \( \Phi \) is a function of \( x \). For an SDE that may be Ito or Stratonovich,

\[
dx = F(x, t)dt + S(x, t) \circ dW
\]

the time-dependent PDF of \( x \), \( p(x, t) \), obeys a Fokker–Planck equation as follows (e.g., Risken 1984; Arnold 1974):

\[
\frac{\partial p(x, t)}{\partial t} = - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \left[ F_i(x, t) p(x, t) + \alpha \sum_{k=1}^{d} \sum_{j=1}^{m} \frac{\partial S_{ij}(x, t)}{\partial x_k} S_{kj}(x, t) p(x, t) \right] + \frac{1}{2} \sum_{i=1}^{d} \sum_{k=1}^{d} \sum_{j=1}^{m} \frac{\partial^2}{\partial x_i \partial x_k} \left( S_{ij}(x, t) S_{kj}(x, t) p(x, t) \right)
\]

In equation (9), \( d \) is the dimension of \( x \) and \( m \) is the dimension of \( dW \). As in equations (3) and (4), the quantity \( \alpha \) is equal to 0 for Ito calculus and to 1/2 for Stratonovich calculus. Note that this is a diffusion-type equation, and processes described by equation (8) are called “Markov diffusion processes” (Horsthemke and Lefever 1984; Penland 1996). Equation (9) is obeyed not only by the marginal PDF \( p(x, t) \), but also by the conditional probability density \( p(x, t|x_0, t_0) \). That is, \( p(x, t|x_0, t_0)dx \) is the probability that the state of the system is between \( x_0 \) and \( x + dx \) at time \( t \), given that the state of the system at time \( t_0 \) is \( x_0 \). The entire term involving square brackets is called the “drift,” and the part of that term involving \( \alpha \) is called the “noise-induced drift.”

“Noise-induced drift” is important and can have large consequences on the solution. The use of equations (9) and (6) to show a realistic effect of the noise-induced drift has
been investigated by (Sardeshmukh, Penland, and Newman 2001). In that study, stochastic perturbations to the friction term in the barotropic vorticity equation for the atmosphere were shown to amplify the mean response to a Rossby wave source by 20% to 30%. In contrast, stochastic perturbations to the mean flow were shown to add random phases to the Rossby-wave response so that the mean response was damped to about half of the deterministic response. In both cases, Equation (6) was employed, with the relevant timescales estimated from data.

Equation (9) shows clearly that the effect of stochasticity on the mean \( \langle x \rangle \), and any other averaged quantity, is generally not zero for Stratonovich calculus if the coefficient of the stochastic term is state-dependent. As an example, consider acoustic wave propagation through a stochastic oceanic internal gravity wave field (Penland 1985). As the wave propagates, random variations in the sound speed, modulated by the amplitude of the wave itself, alternately advance and retard the wave so that the average amplitude and phase speed of the wave are different from what the unperturbed values would be.

At this juncture, it is appropriate to mention that there is a transformation between Ito systems and Stratonovich systems. Note from equation (9) that the noise-induced drift might be combined with \( F(x,t) \), to form a term, say \( F' \). Replacing \( F \) with \( F' \) in equation (8) (and also removing Ito’s circle) gives an Ito SDE equivalent to the Stratonovich equation, equation (8). For this reason, the noise-induced drift is sometimes called the “Ito correction,” as in Hodyss et al. (2013). As emphasized in that article, the modeler should take extreme care to ensure in which calculus a theory is developed, and which calculus is to be used in the simulation, so that it is clear whether or not an Ito correction should be added or subtracted.

In principle, equations for the evolution of the average of any function \( f(x,t) \) can be derived by multiplying equation (9) by \( f(x,t) \) and then integrating over all \( x \). In practice, this is often prohibitively difficult and, even after the equation is derived, one still has to solve it. As an alternative, one may integrate equation (8) using an appropriate integration scheme to generate ensembles of \( x \) and \( f(x,t) \), and then average those results.

d. Uncertainty modeling and the Ito–Stratonovich quandary

As discussed in Section 3c, the noise-induced drift is important only when a state-dependent coefficient multiplies the white noise. If a noise with a finite decorrelation time is implemented in a model, the question is generally moot. For example, let us say that the noise term has a Gaussian distribution with decay time \( 1/\gamma \). In this case the dimensionality of the system is inherently increased, because such Gaussian noise is equivalent to an Ornstein–Uhlenbeck (OU) process (cf. equation 18), which involves only additive noise. The same is true for noise involving finite spatial correlations. Thus, the stochastic partial differential equations used by Lermusiaux and colleagues (Lermusiaux et al. 2006 and citations therein) to model errors due to finite resolution, limited knowledge of physical processes, and measurement errors of ocean data (needed for model initialization and parameters) need not incorporate Ito calculus in their integration.
Nevertheless, papers on data assimilation often assume that the stochastic description of uncertain processes obeys Ito calculus, e.g., Miller, Carter, and Blue (1999). There are two reasons for this. First of all, data assimilation attempts to estimate the PDF of the noisy system and, if the Fokker–Planck equation is used, the Ito formulation is much simpler than the Stratonovich formulation (cf. equation 9). Using the Fokker–Planck equation, however, is usually practical only for simple systems, in which the Ito correction can be added or subtracted by hand. The other reason is that, as discussed by Miller, Carter, and Blue (1999), the observation process itself can be modeled as a continuous process governed by an Ito SDE as the discrete observations become frequent enough to be estimated as continuous. When this is the case, the noise term in the equation for the observations is uncorrelated with the modeled system and, as in cases in which uncertainty is modeled as an OU-process, noise-induced drift issues are greatly simplified.

The case is more complicated for modeling location uncertainty of passive tracers and other cases where the convective derivative explicitly employs a multiplicative, stochastic velocity. True stochastic partial differential equations contain additional terms in the Stratonovich to Ito transformation (Kunita 1997), and there are disagreements in the literature (e.g., Holm 2015; Resseguier, Mémin, and Chapron 2017) regarding applications of stochastic partial differential equations in fluid dynamics. We do not discuss this issue in detail; however, modelers may take some comfort in that Stratonovich systems obey the same rules of calculus that deterministic systems do, and when the noise supplied to the model is either additive or red, the issue is moot. Nevertheless, it is important for modelers to know whether system equations they wish to implement are being derived in Ito or Stratonovich calculus, and whether or not there are terms belonging to an Ito correction in the resulting equations.

4. **Time symmetries in stochastic models**

The introduction of a stochastic term into a differential equation like equation (5), as is shown in equation (7), introduces another subtle mathematical point with important physical consequences. A deterministic differential equation is a prescription for the value of the derivative at every point of the phase space of the problem. Most problems that are relevant to physical systems involve not-pathological functions that behave nicely, providing the same value irrespective of the direction of the approach to a specific point in the phase space. In the case of time-dependent systems like equation (5), it means that the value of the instantaneous derivative in time is the same whether you are moving forward or backward in time. In other words, the right and left derivative with respect the time are identical.

However, when a stochastic term is introduced, as in equation (7), this symmetry is broken because the forward and backward time tendencies are no longer the same because the stochastic increments in the two directions are now different. This observation was first made by Nelson (1966) and Davidson (1979) in an attempt to derive a stochastic theory of quantum mechanics but, as Bacciagaluppi (2005, 2012) has discussed, it can have a more general meaning.
Dissipative deterministic systems are usually considered to be irreversible, because the forward and backward solutions are not both acceptable under physical hypothesis. One direction, for example, might lead to an unchecked growth of energy, breaking conservation laws. But it should be stressed that, from a mathematical point of view, those solutions have no special problems and they are generated by the same instantaneous prescription of the tendency as the deterministic equation prescribed. Stochastic terms break the mathematical symmetry between forward and backward solutions in a more fundamental way.

In many cases, including the ocean application we are discussing, the parameterization of $\eta$ will result in a Markov diffusion process with a noise intensity $G$ (it is taken here to be constant for simplicity, but the following argument will be the same for a state-dependent or time-dependent $G$)

$$d\mathbf{x} = \mathbf{F}(\mathbf{x}, t)dt + G\,d\mathbf{W}(t). \quad (10)$$

This is a generalized Langevin equation that combines deterministic dynamics (the $\mathbf{F}(\mathbf{x}, t)$ term) with a stochastic process $(d\mathbf{W}(t))$. It is expressed as an evolution law that gives a prescription for the time tendency of the physical process, depending on time and the physical attributes of the system under consideration. Under reasonable assumptions, excluding pathological situations, the evolution is completely determined by this prescription.

However, the inclusion of the stochastic process changes the situation at a fundamental level. The stochastic force destroys the uniqueness of the prescription of the evolution based only on the specification of the derivative in time, mostly because, strictly speaking, the system does not have a derivative anymore. We have seen that we can still give meaning to the derivative and use the instruments of ordinary calculus with some care, but we cannot avoid the fact that the forward and backward time derivatives are different. Equation (10) specifies only one of those (just to fix the ideas we can call it the “forward” process) and in order to have a complete definition of the process we also need to specify the process in the other time direction (“backward”).

$$d\mathbf{x} = \mathbf{F}^*(\mathbf{x}, t)dt + G\,d\mathbf{W}^*(t) \quad (11)$$

The forward and backward processes will generate two companion Fokker–Planck equations of the same type as in equation (9),

$$\frac{\partial p_f(\mathbf{x}, t)}{\partial t} = -\nabla \cdot (\mathbf{F}(\mathbf{x}, t)p_f(\mathbf{x}, t)) + G\nabla^2 p_f(\mathbf{x}, t) \quad (12)$$

$$\frac{\partial p_b(\mathbf{x}, t)}{\partial t} = -\nabla \cdot (\mathbf{F}^*(\mathbf{x}, t)p_b(\mathbf{x}, t)) - G\nabla^2 p_b(\mathbf{x}, t) \quad (13)$$

We can define the following quantities with the dimensions of a “velocity” that Nelson (1966) called “osmotic” and “current” velocities, respectively.

$$\mathbf{u} = \frac{(\mathbf{F} - \mathbf{F}^*)}{2} \quad (14)$$
and
\[ \mathbf{v} = \frac{(\mathbf{F} + \mathbf{F}^*)}{2} \tag{15} \]

In order to have a physically consistent description, we have to require that the probability distribution be the same in the two descriptions, i.e., \( p_f(x, t) = p_b(x, t) = \rho(x, t) \). It is interesting to note, then, that the probability distribution obeys a current conservation equation
\[ \frac{\partial \rho}{\partial t} = -\nabla (\mathbf{v} \rho) \tag{16} \]
with \( \mathbf{v} \) as the current velocity.

The same symmetry requirement implies a relation between \( \mathbf{F} \) and \( \mathbf{F}^* \),
\[ \mathbf{F}^* = \mathbf{F} - 2G \frac{\nabla \rho}{\rho} \tag{17} \]
so, the osmotic velocity can be written as
\[ \mathbf{u} = G \frac{\nabla \rho}{\rho} \]

These considerations show that the addition of a stochastic term to an equation is more complex that it seems. It changes the fundamental nature of the physical law that we are trying to describe and, in principle, we should also take care of the time-symmetry properties of the system itself. The traditional approach to modeling is to find out the terms and forces that act on the system in such a way as to specify the time evolution. The addition of a stochastic component, which often can be done very simply in a numerical environment, nevertheless requires that some thought be given on the time-reversed process, leaving an incomplete description if it is left unspecified.

Although the time symmetries are most easily illustrated with additive noise, the more general concept is important, particularly when data assimilation is involved. In fact, as mentioned in Section 3, the Fokker–Planck equation, equation (9), is obeyed not only by the marginal distribution but by conditional probability. The formal adjoint of equation (9) describes the backward evolution of the system; this is why equation (9) is sometimes called the “forward Kolmogorov equation” and its formal adjoint is called the “backward Kolmogorov equation.” Again, the conditional probability \( p(x, t|x_0, t_0) \) obeys both equations. In terms of data assimilation, one tries to find the initial condition for which the present is the most probable prediction.

5. On numerical methods of stochastic integration

There are excellent books (e.g., Kloeden and Platen 1992) available on the numerical treatment of stochastic integration that offer an exhaustive treatment of how stochastic integration schemes are developed, how the order of convergence is derived, and how the
Ito and Stratonovich integrals are related to each other and give examples of the most commonly used stochastic integration schemes. Some of this treatment is summarized in Ewald and Penland (2009). One of the most useful references on stochastic numerical schemes is the easily-accessed article by Rüemelin (1982), which analyzes the stochastic analog of $n$th-order Runge–Kutta schemes.

A problem with implementing stochastic parameterizations in GCMs is that a GCM has generally been around in one form or other since long before the decision is made to add stochasticity to the mix of extant parameterizations. It is therefore often impractical to rewrite the GCM to accommodate stochastic calculus; however, because not accommodating the demands of stochastic calculus can result in severe errors, compromises are often made. As rigorous treatments of stochastic integration schemes are available elsewhere, we forego repeating them here and discuss some of these compromises and their effects. We emphasize that any nonrigorous compromises are solely the responsibility of the modeler, and the authors of this chapter accept no responsibility for the consequences of insufficient testing.

As a rule of thumb, a forward-stepping scheme results in Ito calculus and a predictor-corrector scheme results in Stratonovich calculus (but be sure to check this on a system where you know the answer!). Also, modelers sometimes simply perturb a parameter in a model with a random component (Buizza, Miller, and Palmer 1999). In that case, because dynamical consistency requires a legitimately-generated Wiener process, it is necessary to multiply that random component by the square root of the ratio of some characteristic time to the time step (cf. Section 3, particularly equations 6 and 7; see also Hansen and Penland 2006). It is possible to account for this step by tuning the amplitude of the random parameter to yield some desired result, but then it is necessary to tune the parameter whenever the time step is changed. This is particularly problematic in a model with variable time step.

The consequences of ignoring the stochastic term’s square-root dependence on the time step was illustrated by Ewald and Penland (2009) with a simple OU process (Uhlenbeck and Ornstein 1930):

\[ \frac{dx}{dt} = -\gamma x + \eta \xi \]  

in equation (18), $\gamma$ and $\eta$ are constants, and $\xi$ is Gaussian white noise. For the parameters chosen by Ewald and Penland (2009), the mean square value $\langle x^2 \rangle = 10$. The results of integrating equation (18) with a naive, deterministic Euler scheme

\[ x(t + \Delta t) = x(t) - \gamma x(t) \Delta t + \eta r \Delta t \]  

where $r$ is a Gaussian random variable with unit variance, compared with a stochastic Euler scheme

\[ x(t + \Delta t) = x(t) - \gamma x(t) \Delta t + \eta r \sqrt{\Delta t} \]  

are shown in Figure 3. The undesirable dependence on the time step by the deterministic Euler scheme is clear in Figure 3. Note also that the Gaussian random variable $r$ in
As stated above, a stochastic approximation places more conditions on the time step $\Delta t$ than the usual deterministic Courant–Friedrichs–Levy (CFL) (Courant, Friedrichs, and Levy 1928) criterion. Finding a time step $\Delta t$ in the flat part of the spectrum usually implies that $\Delta t$ needs to be smaller than required by the CFL criterion. Kloeden and Platen (1992) suggest choosing $\Delta t$ to be a factor of $2^9$ (about 500) smaller than the smallest physical timescale in the system; we have found that larger time steps are permissible, as long as they are one or two orders of magnitude smaller than that physical timescale. Larger time steps can seriously bias the desired statistics of the generated time series. The small-time steps, coupled with the large number of degrees of freedom generally found in GCMs, place a heavy burden on numerical pseudorandom number generators. The return period of many random number generators commonly found in languages such as C or Fortran are short enough to generate spurious periodicities in model output, so be sure to choose a random number generator carefully. The Mersenne Twister (Matsumoto and Nishimura 1998) has a very long period of $2^{19937} - 1$.

In the preceding discussion, we suggested that an integration scheme may not need to be rewritten as long as the stochastic term is suitably modified by the square root of the time step. There is one important exception to this compromise: implicit integration schemes, in
The Sea: The Science of Ocean Prediction

Figure 4. Probability density function estimated from output of a numerical integration of equation (21). The heavy solid line refers to truth. The light solid line refers to the deterministic implicit scheme. Symbols refer to the stochastically-modified, semi-implicit scheme. Figure from Ewald, Penland, and Temam (2004). ©American Meteorological Society. used with permission.

which a forcing term is often found in a denominator. Putting a Gaussian random number into a denominator is extremely dangerous because, with probability equal to one, the denominator will eventually vanish (Kloeden and Platen 1992). Even if the denominator does not vanish completely, errors large enough to compromise the entire integration may be introduced and, if the answer looks physically realistic, these errors may go completely unnoticed.

As an example (Ewald, Penland, and Temam 2004), consider the simple forced system

\[
\frac{dx}{dt} = (k^2 - r)x + F
\]  (21)

In equation (21), \( k \) and \( F \) are constant. The damping \( r \) has a stochastic component, \( r = r_0 + r_s \xi \), where \( \xi \) is white noise. Further details may be found in Ewald, Penland, and Temam (2004). The Stratonovich Fokker–Planck equation is easily solved for the PDF of this system (Fig. 4, heavy solid line). Equation (21) was integrated using a deterministic, implicit leapfrog scheme, similar to that used in models of the barotropic vorticity equation (Sardeshmukh and Hoskins 1988), and a stochastically modified version of that scheme derived in Ewald and Temam (2003), Ewald and Témam (2005), and discussed in Ewald, Penland, and Temam (2004). The PDF of the numerically-generated values of \( x \) are shown for the deterministic scheme (light solid line) and for the stochastically modified scheme (symbols). Not only does the deterministic scheme strongly underestimate the width of the PDF, it puts the peak in the wrong place.
We can summarize the important issues that arise when considering numerical treatment of equations with stochastic components. The inclusion of these terms can be straightforward in numerical models, but it should be kept in mind that they introduce significant conceptual and mathematical differences.

- Update the deterministic terms with the time step and stochastic terms with the square root of the time step. Otherwise, the statistics of your output will spuriously depend on the time step, and even a unit time step has units.
- Because the random term goes as the square root of the time step, time steps in a stochastic model often have to be smaller than in its deterministic counterpart or significant biases may result.
- Unless you’re using an implicit scheme, you can sometimes get away with dividing the random term by the square root of the time step and then using a deterministic integrator.
- The white-noise limit and the continuous limit do not commute (Arnold 1974). Use Stratonovich integration schemes for continuous systems with multiple timescales, in which the fast processes are to be approximated as white noise.
- There is a transformation between Itô systems and Stratonovich systems, but a measuring device will not make this transformation for you. If the model purports to mimic reality, you have to know which calculus the system you’re trying to model obeys.
- A random number generator with too short a period can contaminate the model output with unwanted periodicities.

6. A selection of stochastic parameterizations in ocean models

So far, we have addressed Hasselmann’s vision and how more rigorous developments of stochastic parameterizations corroborated that vision. We have also considered general mathematical and practical issues in the implementation of stochastic parameterization. This section is devoted to giving a sense of how these ideas have been applied in oceanic modeling research. We do not, and cannot, claim to give an exhaustive history; the following is intended to provide pointers to further reading of references within those we have provided here. Neither do we, in general, judge the rigor of these applications; the readers may form their own opinions.

The use of stochastic parameterization to represent wind-driven ocean forcing, in keeping with Hasselmann’s original idea, has been used to study low-frequency variability in idealized basin models (Frankignoul and Hasselmann 1977; Frankignoul, Müller, and Zorita 1997; Cessi and Louazel 2001; Sura, Fraedrich, and Lunkeit 2001; O’Reilly, Czaja, and LaCasce 2012). In intermediate to complex GCMs, stochastic wind forcing has been investigated as a driver of El Niño (Chang et al. 1996; Blanke, Neelin, and Gutzler 1997; Moore and Kleeman 1999; Zhang, Flügel, and Chang 2003; Jin et al. 2007). Other studies have identified the North Atlantic Oscillation (NAO) as a wide-band forcing of North
Atlantic wind–driven circulation (Chhak, Moore, and Milliff 2009; Penland and Hartten 2014; Bellucci et al. 2008).

Stochastic parameterization has been investigated as a means to model unresolved internal ocean processes as well as atmospheric effects. Developmental studies have followed several basic approaches. One such approach is to run a fine-scale model and use those statistics to force a coarse-grained system. Berloff (2005a, b) used this method to model “dynamically consistent” eddy fluxes, replacing the deterministic eddy-diffusion term usually employed in the contemporary ocean models with random forcing. (Later, Berloff modified the spatial characteristics of this forcing and imposed periodic [Berloff 2015] and deterministic transient [Berloff 2016] temporal dependence on this forcing instead of using a random model). The term “dynamical consistency” was used to indicate that the randomly-forced, low-resolution model could reproduce the coarse-grained statistics of the high-resolution model. Variations on this approach have been used to find dynamical scalings with a view toward implementing subgrid parameterizations without having to run the fine-scale model (Li and Storch 2013; Cooper and Zanna 2015; Mana and Zanna 2014; Williams et al. 2016). Alternatively, one may develop prognostic equations for the internal eddies based on equilibrium assumptions of the wavenumber spectrum (Grooms, Majda, and Smith 2015), conservation of energy (Jansen and Held 2014), or unresolved fluctuations in the equation of state (Brankart 2013).

A popular attempt to include specific stochastic corrections to the tendency equations in the ocean equation in coupled models, to reduce ensemble-forecast error and increase forecast spread, involves stochastic perturbations of physics parameters (SPPT) (Buizza, Miller, and Palmer 1999). Andrejczuk et al. (2016) found that including SPPT in a coupled model did not improve the seasonal forecast error in the ocean component, although a beneficial effect was found in the forecast spread. In contrast, Zheng and Zhu (2016), in an intermediate coupled model, found that such stochastic perturbations significantly improved the ensemble-mean forecast skill. SPPTs also provided significant improvements in the description and prediction of El Niño in a full GCM (Christensen et al. 2017).

In most cases, rigorous application of the central limit theorem, as described in Section 3, has not been implemented when developing stochastic parameterization in ocean models. In fact, rigorous implementation often requires extensive rewriting of a GCM, and it has been shown that it is sometimes possible to take advantage of stochastic parameterizations without doing so. However, we do advise that numerical architectures in new models be made as flexible as possible, and that the numerical needs of stochastic modeling be honored.

7. Conclusions

In this chapter, we have reviewed the general theory of SDEs, including the theoretical background on how measured statistics from time series can be used to develop a stochastic parameterization. The general rules of stochastic calculus have been mentioned, and references provided where more detail may be found. The differences between Ito and
Stratonovich calculus have been pointed out, and we have discussed how Stratonovich calculus is usually appropriate for fluid systems, whereas Ito calculus is often appropriate for data assimilation. We have illustrated some common numerical pitfalls awaiting the unwary modeler, and warned against unsophisticated random number generators. Finally, we have cited a selection of examples showing that the variability of unresolved scales in an ocean model is important and, by citation, a variety of methods that have been employed.

What is remarkable about some of these studies is that benefits of stochasticity in numerical forecast models can be found even when the application does not conform to the requirements of rigorous limit theorems (Buizza, Miller, and Palmer 1999; Christensen et al. 2017). Nevertheless, we strongly recommend flexibility in the design of new models, so that they may be rigorously modified while still in development.

Most basic fluid dynamical concepts were developed during 17th through 19th centuries by luminaries like Newton, Navier, and Stokes, so the application of SDEs to geophysical modeling is relatively new. Modern stochastic theory can be said to have begun with Kolmogorov (1933). The theory of Brownian motion had been developed before that by Einstein (1905), for which an excellent update was written by Chandrasekhar (1943). The calculus of Ito (1950) was actually developed for SDEs before the calculus that reproduces the classical rules obeyed by deterministic systems (Stratonovich 1966), and industrial uses of the noise-induced drift followed soon after (e.g., Personick 1971). There are several textbooks on SDEs accessible to the scientist, and we have cited some of them. Particularly useful texts are Arnold (1974), Gardiner (1985), Risken (1984), and Bhattacharya and Waymire (1990). In what has become a classic text, the numerical treatment of SDEs is covered extensively in Kloeden and Platen (1992).

As stated in the beginning of this chapter, and repeated throughout, this exposition is far from complete. We can only beg pardon of the many authors whose works might reasonably have been included here and were not. We have also neglected several other avenues of stochastic modeling that are currently the subjects of geophysical research. The most important of these, in our view, is the general theory of Levy processes, which are nicely presented in the textbook by Appelbaum (2004). These processes need not be subject to the continuity conditions required by classic SDEs, and we may speculate that they may be useful in conceptual models of convection.

This is an exciting time to be a scientist. With modern computers and the increasing ease of internet communication, cross-disciplinary influences are erasing the boundaries between theory and application. We hope that we have provided ocean modelers a place to start exploring the vast field of stochastic modeling, and look forward to the discoveries that are yet to come.

Acknowledgments. The authors gratefully acknowledge discussions with P.D. Sardeshmukh, I. Grooms, and the contributions of two thorough referees. In addition, we are grateful that the U.S. National Oceanic and Atmospheric Administration (NOAA) permitted C. Penland the time to contribute to this chapter.
REFERENCES


Received: 10 February 2017; revised: 10 August 2017.

Editor’s note: Contributions to The Sea: The Science of Ocean Prediction are being published separately in special issues of Journal of Marine Research and will be made available in a forthcoming supplement as Volume 17 of the series.